SPRINGER
REFERENCE

Joseph N. Pelton
Scott Madry
Sergio Camacho-Lara
Editors

Handbook of Satellite Applications

Second Edition

Springer

Handbook of Satellite Applications

Joseph N. Pelton • Scott Madry
Sergio Camacho-Lara
Editors

Handbook of Satellite Applications

Second Edition

With 567 Figures and 87 Tables

Springer

Editors
Joseph N. Pelton
International Space University
Arlington, VA, USA

Scott Madry
Global Space Institute
Chapel Hill
NC, USA

Sergio Camacho-Lara
Centro Regional de Enseñanza de Ciencia y
Tecnología del Espacio para América Latina
y el Caribe (CRECTEALC)
Santa María Tonantzintla
Puebla, México

ISBN 978-3-319-23385-7 ISBN 978-3-319-23386-4 (eBook)
ISBN 978-3-319-23387-1 (print and electronic bundle)
DOI 10.1007/978-3-319-23386-4

Library of Congress Control Number: 2012952160

1st edition: © Springer Science+Business Media New York 2013
© Springer International Publishing Switzerland 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Imagine if all satellite services would close down for even a few hours. The global consequences of such a catastrophic event – even for specialists in the field – would be hard to grasp. We can easily imagine huge disruptions in telecommunications traffic and banking operations occurring within minutes. In time, chaos would spread to stock markets, television broadcasts, weather forecasting, and storm alerts, as well as airline travel. By the second hour, the problems would have even spread to activities like education, health care, and many other basic services of industry and government. Some years back, a communications satellite failed and the satellite-based pager system for many doctors, surgeons, police, and firemen suddenly went down. For the first time people began to realize just how dependent they were on satellites in their daily lives.

At the 2012 International Space University (ISU) symposium on Space Sustainability, one expert referred to the "Day After" scenario for the possibility when all satellites might fail. Whereas one can survive without a utility like electricity for a short time, the longer-term consequences for a global society would be quite dramatic. The same would happen for a world without satellite services. A world stripped of its application satellites would be set back many decades in its progress.

We would suddenly inhabit a world where misinformation could reign again. It is not an overstatement to say that a world without satellites could actually plunge us into war.

In short, space applications today have become a utility, just as in the case of electricity or water. We basically do not wonder where our electricity or our water was produced when we use a power socket or turn on a water faucet. We just assume it will be available with good quality in a sustainable way. Today, in a world with rising population and climate change we are becoming more and more concerned about long-term availability of resources, and when we do so we also need to reflect on the availability of space resources. Just think about the consequences of a strong solar storm such as that occurred in 1859. This quite unusual Coronal Mass Ejection (CME) or solar eruption, now called the Carrington Event, managed to set telegraph offices on fire and brought the aurora borealis as far south as Cuba and Hawaii for many days. It is evident that a repetition of such an event nowadays would bring considerable damage to our application satellites and could interrupt global satellite services in a major way.

Knowledge of satellite applications is important, but it is equally important to understand the whole system starting from the underlying basics of the technology and how satellites are built and operated nationally, regionally, and globally. We need to know the potential of these satellites today and tomorrow as well as understand the threats that can influence their performance.

This handbook is exactly aimed to fulfill this purpose and provides an excellent and outstanding overview of satellite applications, at the same time emphasizing the regulatory, business, and policy aspects. The authors are among the best experts worldwide and it is a pleasure to note that many of them are regular lecturers at the International Space University, which at the same time guarantees the interdisciplinary character of this unique standard work.

The International Space University for these reasons is proud to fully endorse this important handbook in its now second edition with significant new updates and additional chapters on new satellite networks and applications.

Strasbourg, France Walter Peeters

Acknowledgments

The multiyear creation of the *Handbook of Satellite Applications* represents a mammoth effort by many dozens of authors. Some of the world's most outstanding experts in their field graciously supported this effort as volunteers. We thus wish, first and foremost, to thank these authors who have so generously contributed toward this attempt to creating a truly definitive reference work across the closely related fields of satellite communications, remote sensing, space navigation, and meteorological sensing from space.

We also wish to thank Dr. Michael Simpson, former President of the International Space University (ISU). It was he who first encouraged the creation of this reference work and had the vision to support a handbook that would comprehensively cover the entire range of major satellite applications. Also, it is important to recognize the current President of the International Space University (ISU) Walter Peeters, Dean Angelina Bukley, plus the current and past chairs of the ISU Academic Council, who were also involved with this project, namely, Drs. Edward Chester and Stefano Fiorilli.

Finally we would like to thank the many people at Springer Publishing who nurtured and supported this project during its 2-year gestation. We wish to thank Maury Solomon, who first conceived that such a project would be an important undertaking for the scientific literature in the field. Particular thanks also go to Barbara Wolf and Saskia Ellis, who carefully oversaw the final editing and kept the production schedule more or less on track. Then there was Banupriya Mohanraj and Jayanthi Vetriselvam, who, along with Ms. Ellis, very scrupulously oversaw the production of this extensive reference work on a day-after-day basis. We thank them both for their constant eye to detail and their tireless efforts.

Joseph N. Pelton
Sergio Camacho-lara
Scott Madry

Contents

Volume 1

Volume 2

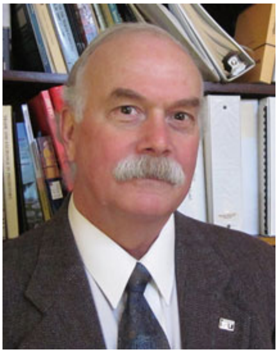
About the Editors

Dr. Joseph N. Pelton
International Space University
Arlington, VA, USA

Dr. Joseph N. Pelton is an award-winning author/editor of over 40 books and over 300 articles in the field of space systems. These include the four book series: e-Sphere, Future Talk, Future View, and Global Talk. For the latter book, he was nominated for a Pulitzer Prize. He served as Chairman of the Board (1992–1995) and Vice President of Academic Programs and Dean (1995–1996) of the International Space University of Strasbourg, France. He is currently a member of the ISU faculty and series editor for a number of books on behalf of the university. He is also the Director Emeritus of the Space and Advanced Communications Research Institute (SACRI) at George Washington University. This institute, which he headed from 2005 to 2009, conducted state-of-the-art research on advanced satellite system concepts and space systems. From 1988 to 1996, Dr. Pelton served as Director of the Interdisciplinary Telecommunications Program at the University of Colorado Boulder, which at that time was the world's largest graduate level telecommunications program. Prior to that, he held a number of positions at Intelsat and Comsat including serving as director of Strategic Policy and director of Project Share for Intelsat.

Dr. Pelton is a fellow of the International Association for the Advancement of Space Safety (IAASS), a member of its Executive Board, and chairman of its Academic Committee. He is also the Executive Editor of the IAASS publication series. He is the former president of the International Space Safety Foundation of the USA, the former president of the Global Legal Information Network, and the founding president of the Society of Satellite Professionals. Dr. Pelton was the founder of the Arthur C. Clarke Foundation and remains as the vice chairman of its Board of Directors. This Foundation honors Sir Arthur Clarke, who first conceived of the Communications Satellite (as of 1945). Dr. Pelton was elected to full membership in the International Academy of Astronautics in 1998. His other awards include: the H. Rex Lee Award by the Public Service Satellite Consortium (1986);

the outstanding educator award by the International Communication Association (1995); the Hall of Fame of the Society of Satellite Professionals International (2000), an honor only extended to less than 100 people since the beginning of this award in the 1980s; the Arthur Clarke Foundation Award for lifetime achievement in the field of satellite communications (2001); and the ISCe Award for Educational Excellence (2005). He was also awarded the Arthur C. Clarke Award for International Achievement by the British Interplanetary Society in 2013. He has also been elected an Associate Fellow of the American Institute of Aeronautics and Astronautics. His degrees in physics and international relations are from the University of Tulsa, New York University (NYU), and Georgetown University.



Dr. Scott Madry
Global Space Institute
Chapel Hill, NC, USA

Dr. Scott Madry is the Executive Director of the Global Space Institute. He is the Founder and President of Informatics International, Inc. and is a Research Associate Professor of Anthropology at the University of North Carolina at Chapel Hill, specializing in the applications of geomatics technologies for regional environmental and cultural research. He is a long-time member of the faculty of the International Space University and has served as the program director of the ISU Southern Hemisphere Summer Space Program for 4 years in Adelaide, Australia, as well as teaching in some 24 ISU programs around the world.

He received his Ph.D. from UNC in 1986 and then worked for 3 years at the Institute for Technology Development, Space Remote Sensing Center at the NASA Stennis Space Center. He then took the position of Senior Associate Director of the Center for Remote Sensing and Spatial Analysis at Rutgers University, where he taught for 9 years in the Anthropology, Geography, and Natural Resources departments. He is widely published and is the author of two books and over 75 articles and papers. He is a two-time Fulbright Scholar, having taught in Burgundy, France, and Johannesburg, South Africa.

He has conducted research in North America, Europe, and Africa, and has given over 150 short courses and seminars in over 30 countries around the world. He has consulted for numerous governments, corporations, and nonprofits, and has received over US$7.5 million in grants and contracts. Dr. Madry has acted as a consultant on satellite remote sensing for a major Hollywood film and has been a consultant for major museum exhibitions on the subject. Some of his research was featured in the *McGraw Hill Yearbook of Science and Technology* in 1991. In 1997 he was awarded the Russian Tsiolkovsky Gold Medal for his international research and teaching activities. He is currently conducting work in North Carolina, Florida, Rwanda, and France. He is very active in the American Red Cross Disaster Services and has won several awards for his activities on behalf of the Red Cross. In 2012 he was awarded, along with the other members of the GISCorps, the President's Volunteer Service Award by President Barack Obama for his work in applying geomatics technologies to disaster management.

Dr. Sergio Camacho-Lara
Centro Regional de Enseñanza de Ciencia y
Tecnología del Espacio para América
Latina y el Caribe (CRECTEALC)
Santa María Tonantzintla, Puebla, México

Dr. Sergio Camacho-Lara is the Secretary General of the Regional Centre for Space Science and Technology Education for Latin America and the Caribbean. He was Director of the United Nations Office for Outer Space Affairs and Chief of the Space Applications Section and the Committee Services and Research Section in the same office. He worked on the organization of the Third United Nations Conference on the Exploration and Peaceful Uses of Outer Space (UNISPACE III) and on implementing its recommendations, including the establishment of the International Committee on GNSS. Prior to joining the United Nations, he carried out research on the interaction of electromagnetic radiation with matter at the Instituto de Geofisica, Universidad Nacional Autónoma de México. He obtained his Ph.D. from the University of Michigan.

About the Authors

Audrey L. Allison
Frequency Management Services
The Boeing Company
Seattle, WA, USA

Allison is Director, Frequency Management Services for Boeing. Her organization provides enterprise-wide radiofrequency spectrum acquisition, and policy support for Boeing technology, products, services, and operations worldwide – and to Boeing's commercial and government customers. She is based in Boeing's Washington D.C. office. Her previous work history includes the Federal Communications Commission's International Bureau; Iridium LLC; and consulting to the Department of Defense on international spectrum and regulatory issues.

Allison is Boeing's representative to the International Telecommunication Union (ITU). She recently served as Chairman of Committee 6 of the 2015 World Radiocommunication Conference in Geneva, Switzerland, and was elected as Vice-Chairman of the ITU's Radiocommunication Advisory Group (2008–2015). She chaired the FCC's WRC-15 Advisory Committee Working Group on Aeronautical, Maritime, and Radar Issues (2012–2015); the United States ITU Association (2009, 2015); and cochaired the Satellite Industry Association's Regulatory Working Group (1991–2001, 2012–2015). Allison currently serves as a member of three Federal Advisory Committees on international and national spectrum management and telecommunications matters.

Allison is an attorney with a Master of Business Administration, *cum laude*, from the International Space University, a Master of Laws in international law from Georgetown University, and a *Juris Doctor* from Catholic University of America's Institute for Communications Law Studies. She is the author of *The ITU and Managing Satellite Orbital and Spectrum Resources in the 21st Century*, Springer, 2014. Allison is also adjunct faculty to the International Space University in Strasbourg, France, and a visiting lecturer at McGill University Institute of Air and Space Law in Montreal, where she lectures on international regulation of satellites.

Allison resides near Annapolis, Maryland, with her husband and daughter.

Jeremy E. Allnutt
Department of Electrical and Computer Engineering
George Mason University
Fairfax, VA, USA

Jeremy Allnutt earned his B.Sc. and Ph.D. in Electrical Engineering from the University of Salford, UK, in 1966 and 1970, respectively. From 1970 to 1977, he was at the Appleton Laboratory in Slough, England, where he ran propagation experiments with the US satellite ATS-6 and the European satellites SIRIO and OTS. In 1977, he moved to BNR, now Nortel, in Ottawa, Canada, and worked on satellite and rural communications projects before joining the International Telecommunications Satellite Organization (INTELSAT) in Washington, DC, in 1979. Jeremy Allnutt spent 15 years at INTELSAT in various departments. During this period, he ran experimental programs in Europe, Asia, Africa, North and South America, Australia, and New Zealand, finishing as chief of Communications Research Section. Jeremy Allnutt spent 1 year as Professor of telecommunications systems at the University of York, England, and then joined the Northern Virginia Center of Virginia Tech in 1986, where he later ran the master's program in ECE, as well as being on the team that designed and set up the masters in information technology program. In August of 2000, he moved to George Mason University with dual appointments: director of the new masters in telecommunications program and professor in the ECE department. In August 2009, he became director of the new MS in computer forensics program at George Mason University. Jeremy Allnutt has published over 100 papers in conferences and journals and written three books, most in his special field: radiowave propagation. He is a fellow of the UK IET (formerly the IEE) and a fellow of the US IEEE.

Michel Bousquet
Institut Supérieur de l'Aéronautique et de l'Espace (ISAE)
Toulouse, France

Professor Michel Bousquet manages the academic and research programs in satellite communications and navigation at ISAE, the French Aerospace Engineering Institute of Higher Education. He chairs the Scientific Board of TeSA (www.tesa.fr), a cooperative research lab on aerospace communications and navigation. With research interest covering many aspects of satellite systems, he participates in many national and international RR&DD programs (French DoD, CNES, ESA, COST, FPs, SatNex NoE). Professor Bousquet has authored many papers and books including the widely used Satellite Communication Systems and sits on several boards of international conferences and journals to promote RR&DD results.

Halil Cakir
Air Quality Analysis Group/AQAD/OAQPS
US Environmental Protection Agency
Research Triangle Park, NC, USA

Dr. Halil I. Cakir has master's degrees in Forestry from Clemson University and in Applied Economics from Georgia Southern University. He earned his doctoral degree from the College of Natural Resources at North Carolina State University (NCSU). Upon graduation, he remained on staff as a postdoctoral research associate, then as a research assistant professor at NCSU. Dr. Cakir is currently employed in the Office of Air and Radiation at the US Environmental Protection Agency.

Dr. Cakir's academic focus has been in the geospatial sciences and their application to natural resource issues. As a research associate at NCSU, he worked on and then led increasingly complex and multidisciplinary research projects for the Environmental Protection Agency, Water Resources Research Institute, Department of Defense, and various state and local governments. His research has advanced the geospatial sciences and has two provisional patents for two new image processing techniques and one new change detection technique. These techniques allow users to sharpen the spatial resolution of multispectral imagery to equal that of a more spatially precise geographically coincident panchromatic black and white image. The techniques retain the integrity of the multispectral characteristics of the imagery so that it can be used in most natural resource and environmental applications.

Dr. Cakir maintains an expertise in applied research as well. He has authored peer-reviewed publications, book chapters, and technical reports. He also served as a reviewer for landmark professional publications like PE&RS.

William E. Carter
NSF National Center for Airborne Laser Mapping
(NCALM)/Department of Civil and
Environmental Engineering
University of Houston
Houston, TX, USA

William E. (Bill) Carter is a Research Professor at the University of Houston and a co-PI for the National Science Foundation (NSF) National Center for Airborne Laser Mapping (NCALM). From 1996 to 2010, he was an Adjunct Professor at the University of Florida (UF), where he taught courses in geodesy and conducted research on advanced geodetic techniques. Prior to joining UF, Bill was Chief of the Geosciences Laboratory, NOAA, and led research programs in very long baseline interferometry (VLBI), absolute gravimetry, and GPS. Bill has coauthored two books and more than a hundred technical papers.

Denis Curtin
Communications Satellite Consultants Rockville
MD, USA

Denis Curtin retired from XTAR LLC in July 2010 and is now a communications satellite consultant. Before his retirement he was the COO of XTAR, the joint venture between Loral Space & Communications, Ltd. and HISDESAT Servicios Estrategicos, S.A., (HISDESAT) providing commercial X-band satellite services to the US and allied governments since October 2003. From July 2001 to October 2003 he was the Executive VP, Engineering and Operations of XTAR. Prior to that from January to July 2001, he led the Loral team that negotiated the joint venture with HISDESAT.

Previously, Dr. Curtin served as Senior VP, Engineering and Operations, for Loral ORION and as Executive VP, Loral Cyberstar Broadband Systems. He was responsible for the technical design of the ORION satellites and was instrumental in the formation of the ORION partnership. Before joining ORION, Dr. Curtin held a series of progressively senior engineering and management positions at COMSAT Laboratories, COMSAT General, and COMSAT, culminating in Senior Director, Satellite Engineering, responsible for all COMSAT's satellite engineering.

Dr. Curtin has an M.S. in Physics and a Ph.D. in Mechanical Engineering. He is a Fellow of the American Institute of Astronautics and Aeronautics (AIAA) and a senior member of the Institute of Electrical and Electronics Engineers (IEEE). In 2006, he was named the recipient of the AIAA Aerospace Communications Award, presented for outstanding contributions in the field of aerospace communications.

In March 2009, he was inducted into the Society of Satellite Professionals International (SSPI) Hall of Fame. In 2010, he was appointed to the Board of Directors of XTAR LLC and retired from that position in the Fall of 2014. He is on two nonprofit Boards of Directors and is known as a mentor for young engineers and engineering students. He is the coauthor of the article on "Communications Satellites" in the 10th edition of the McGraw Hill *Encyclopedia of Science and Technology* and has also published extensively on satellite technology and holds a patent on an infrared transparent solar cell.

Yves Durand
Thales Alenia Space
Cannes, France

Yves Durand began working on satellites 35 years ago at NASA Goddard Space Flight Center, in Washington, DC, where he designed digital filtering algorithms for the NASA/ESA Ulysses spacecraft. He then developed the digital flight control data handling for the Ariane 5 launcher at Aerospatiale in Les Mureaux, France.

He moved to Cannes to work at the satellite center of Aerospatiale, now Thales Alenia Space. He directed several telecom and scientific programs, as well as the Spacebus (GEO) and Proteus (LEO) platform product lines. He initiated the LEO constellation series of Thales Alenia Space with Globalstar 2. Yves Durand now coordinates technology developments for Thales Alenia Space.

Rogerio Enríquez-Caldera
Centre for Space Science and Technology
Education for Latin America and the Caribbean
Mexico Campus (CRECTEALC)
National Institute of Astrophysics
Optics and Electronics (INAOE)
Coordinación Electrónica
Tonantzintla, Puebla, Mexico

Rogerio Enriquez-Caldera is an engineering researcher at the National Institute of Astrophysics, Optics, and Electronics of Mexico since 2000. He received his Ph.D. from the University of New Brunswick, Canada. He has ample experience in the area of astronomical instrumentation including control engineering, high sensitive detectors, and optical and radio telescopes. His expertise is related to the areas of electrical engineering, including GPS receiver's software and low noise amplifiers and correlators for very high dynamics platforms in the presence of high levels of noise as well as information systems. He also works for the Mexico campus of the Regional Centre for Space Science and Technology Education for Latin America and the Caribbean teaching in the fields of GNSS, State Estimation for Nonlinear Dynamics Navigation and Tracking Systems, and developing an aerospace flight simulator for formation flying satellites.

Cynthia A. Evans
Code XI2, Astromaterials Acquisition and
Curation Office
NASA Johnson Space Center
Houston, TX, USA

Cynthia A. Evans received her Ph.D. in Earth Science from Scripps Institution of Oceanography in 1983. She joined the Space Shuttle Earth Observations Office at NASA Johnson Space Center (JSC) in 1989 and has participated in or led several Earth Observations experiments from the Shuttle, Shuttle-Mir, and International Space Station Programs. She also managed JSC's Image Science and Analysis Lab.

Today, Evans manages NASA's Astromaterials Acquisition and Curation Office – home to all of NASA's extraterrestrial samples. Her research interests include creating new image-based data products that combine visible imagery and three-dimensional information for Apollo samples from the moon and meteorites, as well as the use of remotely sensed data for investigation of coastal changes, including those related to human activities. In addition to training astronauts in Earth and planetary science, she participates in education and public outreach activities using astronaut photography.

Juan Carlos Fernandez Diaz
NSF National Center for Airborne
Laser Mapping (NCALM)/Department of Civil and
Environmental Engineering University of Houston
Houston, TX, USA

Juan Carlos was born in Tegucigalpa, Honduras, in 1976. His formal education includes a B.S. in Electrical Engineering and an M.B.A. obtained from universities in Honduras, M.S. and Ph.D. degrees in Geosensing Systems Engineering obtained from the University of Florida. Other interests include aviation, Earth/space science and exploration, and the application of space science and technology to bring progress to developing countries.

Craig L. Glennie
NSF National Center for Airborne
Laser Mapping (NCALM)/Department of Civil and
Environmental Engineering University of Houston
Houston, TX, USA

Craig is an Associate Professor at the University of Houston and co-PI of the NSF-funded National Center for Airborne Laser Mapping (NCALM). Dr. Glennie was formerly the vice president of engineering for Terrapoint, a LIDAR remote sensing company with offices in Canada and the USA. He has been active in the design, development, and operation of kinematic remote sensing systems for 17 years. Craig holds a B.Sc. and a Ph.D. in Geomatics Engineering from the University of Calgary and is a registered professional engineer in Alberta, Canada.

Daniel R. Glover
International Space University
Bay Village, OH, USA

Daniel R. Glover spent a 27-year career working as an electrical engineer at NASA's Glenn Research Center, Jet Propulsion Laboratory, and Johnson Space Center. He has worked on various projects including launch vehicles, shuttle experiments, satellite communications protocol research, image data compression, planetary spacecraft conceptual design, space suit design, software management, systems engineering, and strategic planning. He earned his MSEE and Ph.D. degrees from the University of Toledo and an MBA degree from Cleveland State University. Dr. Glover is a member of the faculty of the International Space University. He served as an engineering duty officer in the US Naval Reserve. He has also worked as an independent consultant.

Dr. Gérardine Goh Escolar
Faculty of Law
National University of Singapore
Singapore, Singapore

Dr. Gérardine Goh Escolar is a qualified attorney who specializes in public international law, investment and commercial arbitration, international humanitarian law, and the interaction of international law and new technologies. She is passionate about the crossroads of international law, public policy, and new technologies.

Geri is Legal Adviser to the President of the Iran-United States Claims Tribunal. Alongside Tribunal work, she acts as tribunal secretary and advisor on various state-to-state, investor-state, and commercial arbitrations under the ICSID, ICC, and UNCITRAL frameworks. She is Adjunct Associate Professor at the Faculty of Law, National University of Singapore. She was previously legal officer at the United Nations, counsel at the German Space Agency, and general counsel at a satellite geoinformation corporation.

Geri maintains an active academic portfolio, conducting research and teaching at various universities in Germany, the Netherlands, Singapore, and the United Kingdom. Her research publications have won various awards, including the 2010 Social Science Book Prize of the International Academy of Astronautics, and the 2003 Diederiks-Verschoor Medal. She is a contributor to the International Law Reports and the Annotated Leading Cases of International Criminal Tribunals, and is working on her fourth book, *International Law and Outer Space* (Oxford University Press: forthcoming 2016). Geri is a member of the New York bar.

Jesus A. Gonzalez
National Institute of Astrophysics
Optics, and Electronics (INAOE)/Regional Center for
Space Science and Technology Education for
Latin America and the Caribbean (CRECTEALC)
Campus Mexico
Tonantzintla, Puebla, Mexico

Jesus A. Gonzalez obtained his Ph.D. in Computer Science and Engineering from The University of Texas at Arlington, Texas, in 2001. He is currently a researcher and professor in the Computer Science Department at the National Institute of Astrophysics, Optics, and Electronics, Mexico. He also holds the academic coordinator position at the Regional Centre for Space Science and Technology Education for Latin-America and the Caribbean, Campus Mexico, affiliated to the United Nations.

Arthur N. Guest
International Space University
San Francisco, CA, USA

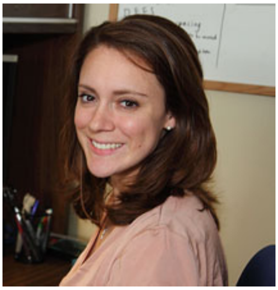
Arthur Guest is a space system engineering consultant who specializes in the principles of system architecting and is located in San Francisco, USA. Arthur graduated from the Massachusetts Institute of Technology with a master's in Aeronautics and Astronautics and is a graduate of the Masters of Space Studies Program at the International Space University. He is currently serving as the space system engineering department chair for the International Space University's 2011 Space Studies Program in Graz, Austria.

Ramesh Gupta
Ligado Networks
Reston, VA, USA

Dr. Ramesh K. Gupta is vice president of Satellite Net Engineering at Ligado Networks where he has supported the development, deployment, and operation of the next-generation satellite network. He has more than 30 years experience in the satellite and wireless communications industries. He has held senior management positions as vice president of advanced business and technology at AMCOM Communications and managing director at COMSAT Laboratories and Lockheed Martin Global Telecommunications. His work has included the integration of large satellite/wireless systems, business planning, and strategic management in a high-technology business environment. He holds a Ph.D. and an M.S. degree in Electrical Engineering from the University of Alberta, Canada; an M.B.A. degree from the Wharton Business School at the University of Pennsylvania; and a B.S. degree (with Honors) in Electronics and Communications Engineering from India. He has published extensively, and holds four US patents. He has received many honors and awards including Alberta Government Telephone's Centennial Fellowship for graduate research in Telecommunications and COMSAT Laboratories' Research Award. He was corecipient of the Best Paper Award at the 9th International Digital Satellites Communications Conference in Copenhagen, Denmark, and the 29th AIAA International Communications Satellite Systems Conference, 2011, in Nara, Japan. He has served as an adjunct associate professor of strategic management and technology planning at the University of Maryland, College Park, MD. Dr. Gupta has published more than 80 papers on satellite and wireless RF technology and systems in AIAA and IEEE conferences and technical journals. Dr. Gupta is a fellow of the IEEE. Dr. Gupta has served on the IEEE MTT-S Adcom since 2013 and as Chair of the Education Committee since 2014.

Henry R. Hertzfeld
Space Policy Institute
Elliott School of International Affairs
The George Washington University
Washington, DC, USA

Henry R. Hertzfeld is a research professor of space policy and international affairs at the Space Policy Institute, Center for International Science and Technology Policy, Elliott School of International Affairs, George Washington University. He is an expert in the economic, legal, and policy issues of space and advanced technological development. Dr. Hertzfeld has served as a senior economist and policy analyst at both NASA and the National Science Foundation, and is a consultant to both US and international agencies and organizations. He is also an Adjunct Professor of Law at The George Washington University Law School and has taught the space law course there for the past 12 years. Dr. Hertzfeld has a B.A. from the University of Pennsylvania, an M.A. from Washington University, and a Ph.D. in Economics from Temple University. He also holds a J.D. degree from the George Washington University and is a member of the Bar in Pennsylvania and the District of Columbia. He can be contacted at hhertzfeld@law.gwu.edu.



Melissa D. Higgins
Jacobs at Astromaterials Research and
Exploration Science Division
Exploration Integration and Science Directorate
NASA Johnson Space Center
Houston, TX, USA

Melissa D. Higgins received her Bachelors of Science degree in Meteorology with minors in Mathematics and Physics from the University of Oklahoma in 2011. She received an internship in 2009 with the Astromaterials Research and Exploration Science Division, and later accepted a position with the Crew Earth Observations (CEO) ISS payload group in July of 2011, both based at the NASA Johnson Space Center in Houston, Texas. She trains mission-assigned astronauts on Earth Science topics and creates a daily target list for the crew to take imagery from space, consisting of science and education-based Earth targets. She enjoys participating in educational outreach and frequently speaks in classroom webinars about astronaut photography.

Chris Hoeber
CFH Engineering
Los Altos, CA, USA

Chris Hoeber retired as the Chief Technology Officer of SSL in January, 2015, after nearly 40 years, including 25 years of Executive Management experience. Among many other things, during that time he was responsible for the management of about 70 satellite programs and the company's research and development and product development programs. He was the leader of the systems engineering team that developed the industry's longest lived satellite platform, the modular and adaptable 1300. He also served as the senior vice president of business development and the company's first chief engineer. In the first year of his "retirement" he has had a wide variety of consulting positions, including serving as the chief technology officer for several startup space companies that are both pushing the technology state-of-the-art and providing novel business approaches to practical problems.

Erwin Hudson
ViaSat, Inc. Carlsbad
CA, USA

Erwin Hudson is Vice President and Manager of ViaSat's commercial satellite communications business in Australia. ViaSat is prime contractor for the Long-Term Satellite Service segment of Australia's National Broadband Network.

Mr. Hudson was formerly the chief technical officer for the services segment of ViaSat including Denver, Colorado-based WildBlue, which was acquired by ViaSat in 2009.

Erwin joined WildBlue in 2000, when the company was an early stage start-up. The company built the first successful Ka-band satellite broadband network, including launch and deployment of two large multibeam satellites.

Prior to joining WildBlue, Erwin was Vice President of Mission Engineering at Space Systems/Loral, responsible for satellite and communications payload engineering, controls engineering, on-orbit test and on-orbit support, advanced systems, and research and development.

Previously, Mr. Hudson was satellite communications manager at TRW Space & Electronics Group. He also held engineering and management positions for satellite communication programs at TRW, including payload systems engineer for NASA's Tracking and Data Relay Satellite, and systems engineering manager for a number of military satellite programs, including 5 years as chief engineer for the Milstar I and Milstar II satellite communications payloads.

Mr. Hudson was a member of the Technical Advisory Counsel of the US Federal Communications Commission 2011–2012. He holds a Bachelor of Science from North Carolina State University, a Master of Science from Ohio State University, and an advanced degree in electrical engineering from the University of Southern California.

Takashi Iida
Tokyo Metropolitan University
Hachioji, Tokyo, Japan

Takashi Iida received B.E., M.E., and Dr. Engineering degrees from the University of Tokyo, Tokyo, Japan, in 1966, 1968, and 1971, all in Electronic Engineering. He joined Radio Research Laboratories (RRL), Ministry of Posts and Telecommunications (MPT), in 1971. He was with National Space Development Agency and involved in the CS-3 satellite development, 1987–1989. He was director of Space Communications Division, Communications Research Laboratory (CRL) (former RRL), MPT, 1989–1991. He was visiting professor of University of Colorado at Boulder, 1991–1992. He was director general of CRL, 1999–2001, and president of CRL, Incorporated Administrative

Agency, 2001–2004. He was executive director of JAXA (Japan Aerospace Exploration Agency), 2005–2007. He was invited advisor/distinguished researcher of National Institute of Information and Communications Technology (NICT, former CRL), Incorporated Administrative Agency, 2007–2009. He was research professor, Tokyo Metropolitan University, 2009–2010. Dr. Iida is now visiting professor of Tokyo Metropolitan University. He is editor and author of the books entitled *Satellite Communications – System and Its Design Technology*, Ohmsha/IOS Press, 2000, and *Satellite Communications in the 21st Century: Trends and Technologies*, AIAA, 2003. He received the AIAA Aerospace Communications Award, 2002, and Hall of Fame Award from SSPI, 2003. Dr. Iida is IEEE life fellow and AIAA fellow.

Bernard Jacqué
Thales Alenia Space
Cannes, France

Bernard Jacqué is Senior Manager, Innovative Space Solutions, Thales Alenia Space

He has assumed this post in the telecommunications sector only recently. From 2013 to 2015, he was Space System and Satellite Senior Sales Manager for North America. Prior to that he was Business Development and Proposal Manager from 2008 to 2012 for Thales Alenia Space. From 2005 to 2008, he was Technical Manager of Alphabus Platform Development Program, which represents largest European telecommunications platform. Prior to his role at Thales Alenia Space, Mr. Jacqué held several key engineering roles at Alcatel Space. He has been actively engaged in the space business since 1994 after earning his French Engineering Diploma (1993) from Aeronautical and Space Engineering (ISAE-SUP'AERO)

Ram S. Jakhu
Institute of Air and Space Law
McGill University
Montreal, QC, Canada

Prof. Ram S. Jakhu has over 30 years of experience in space-related fields. He is holding a tenured position of associate professor at the Institute of Air and Space Law, Faculty of Law, of McGill University in Montreal, Canada, where he teaches several courses covering numerous subjects including public international law, international and national space law and policy, space applications, space commercialization, telecommunications, etc. Currently, he heads a multimillion dollars research and outreach program for space law and policy. From January 1995 to December 1998, Dr. Jakhu served full-time the International Space University, Strasbourg, France, holding various titles, including a professor and the first director of the Master of Space Studies program.

Prof. Jakhu has coauthored three books, written over 80 articles and 20 research reports and edited 7 books. His edited book on National Regulation of Space Activities received the 2011 Book Award from the International Academy of Astronautics.

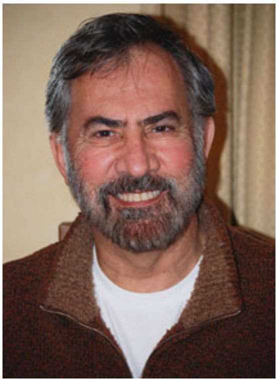
Dr. Jakhu has taught Space Law and Policy in several countries; made presentations to the United Nations Committee of Peaceful Uses of Outer Space (UNCOPUOS); participated in the drafting of Space Law Curriculum for the United Nations Office for Outer Space Affairs (UNOOSA); advised several countries in the preparation of national laws and policies, including Canada, India, and South Africa; and convened and participated in numerous international interdisciplinary space law and policy related conferences and workshops around the world.

Prof. Jakhu is a Member of the Global Agenda Council on Space of the World Economic Forum; Member of the Governance Group of the Space Security Index; and Fellow as well as the Chairman of the Legal and Regulatory Committee of the International Association for the Advancement of Space Safety. He is Managing Editor of the Space Regulatory Series, member of the Editorial Boards of Space and Evolution, Annals of Air and Space Law, Astropolitics, and German Journal of Air and Space Law. He was member of the Advisor Group of Legal Experts on Optional Rules for Arbitration of Disputes Relating to Outer Space within the Permanent Court of Arbitration, and member of the Board of Directors of International Institute of Space Law for 14 years. In 2007, he received a “Distinguished Service Award” from the International Institute of Space Law for significant contribution to the development of space law.

Dr. Jakhu holds a Doctor of Civil Law (Dean’s Honors List) degree in Space Law from McGill University; a Master of Law (LL.M.) degree in the field of Air and Space Law from McGill University. In addition, he has earned LL.M. (in Public and Private International Law), LL.B. (in Laws of India), and Bachelor of Arts (in Economics and Political Science) degrees from Punjab University, Chandigarh, India.

Tarik Kaya
Mechanical and Aerospace Engineering Department
Carleton University
Ottawa, ON, Canada

Tarik Kaya has been working as a Professor at the Mechanical and Aerospace Engineering Department of Carleton University since 2002. He received his Ph.D. from ENSAE, Toulouse, France, in 1993. Before joining Carleton University, he worked as a research associate at NASA Goddard Space Flight Center. Dr. Kaya’s current research interests include mathematical modeling of two-phase heat transfer systems (heat pipes and loop heat pipes) for spacecraft thermal control and miniaturization of the heat pipe technology for electronics packaging.



Siamak Khorram
Department of Environmental Science, Policy, and Management University of California
Berkeley, CA, USA

Center for Geospatial Analytics North Carolina State University Raleigh, NC, USA

Siamak Khorram is a Professor of Remote Sensing and Image Processing. Dr. Khorram has joint faculty appointments at both the University of California at Berkeley and North Carolina State University. He is also the Founding Director of the Center for Geospatial Analytics and a Professor of Electrical and Computer Engineering at North Carolina State University and a member of the Board of Trustees at the International Space University (ISU) in Strasbourg, France. Dr. Khorram is a former Vice President for Academic Affairs and the first Dean of ISU as well as a former Chair of the ISU's Academic Council, and the Principal Advisor to the President. In these capacities, he played major roles in establishing academic relationships between ISU and major space organizations such as NASA, European, French, Japanese, Russian, German, Austrian, and Indian space agencies. He has also served as an ASEE fellow at Stanford University and NASA Ames Research Center. Dr. Khorram has extensive research and teaching experience in remote sensing, image processing, and geospatial technologies and has authored well over 200 publications. He has served as the guiding professor for numerous Ph. D. and Master graduate students. He is a member of several professional and scientific societies and delivered keynote addresses in international scientific conferences. His graduate degrees are awarded by the University of California at Davis and Berkeley.

Heiner Klinkrad
Institute of Space Systems
Braunschweig University of Technology
(TU Braunschweig)
Braunschweig, Germany

Heiner Klinkrad graduated from the Braunschweig University of Technology (TUBS) in aeronautical engineering in 1980, and he received his Ph.D. from the same university in 1984. In 1980 he joined the European Space Agency, from which he retired in 2015 as Head of the Space Debris Office at the European Space Operations Centre ESOC, in Darmstadt/Germany. In that position he was ESA's focal point and senior advisor for space debris matters, and he represented ESA, for instance in the multinational Inter-Agency Space Debris Coordination Committee (IADC).

Heiner Klinkrad is a full Member of the International Academy of Astronautics (IAA), a lifetime Member and Fellow of AIAA, and a member of several other professional associations. He has served as a member or chairperson of working groups and panels of AIAA, COSPAR, ECSS, IAA, IAASS, IADC, IAG, ISO, and UNCOPUOS. Since 2009 he has been a professor at the Institute of Space Systems of the Braunschweig University of Technology. He published a textbook on "Space Debris – Models and Risk Analysis" in 2006, contributed as chapter author to four more books, and was author/coauthor of some 250 publications. He received the "Engineering Sciences Book Award" from IAA in 2013, the "Joseph P. Loftus Space Sustainability Award" from IAASS in 2015, the "IAF Distinguished Service Award" from IAF in 2013, and two "ESA Team Awards" in 2013 and 2014.

Martine Lutz
Thales Alenia Space
Cannes, France

Martine Lutz is an expert in advanced materials (in particular nanotechnologies, advanced composites, and ceramics). She received her Ph.D. from ENSC, Montpellier, France, in 1990 and received an expert nomination from Thales Alenia Space in 2013.

Since 2014 she has been in charge of R&D coordination for the Platform Competence Center of Thales Alenia Space.

Peter Marshall
Royal Television Society
England, UK

After working for the BBC as a journalist and editor, Peter Marshall was a pioneer in developing the use of international satellites for worldwide TV news coverage and distribution. In 1986, he joined INTELSAT in Washington, DC, as the first Director of Broadcast Services. Then as competition in the global satellite industry began to emerge, he moved on to the private sector in 1989 as president of the US-based satellite services company, Keystone Communications, which became a leader in worldwide satellite distribution services. Keystone was acquired by France Telecom in 1996 after which Peter Marshall served as a member of the Board of the France Telecom subsidiary, GlobeCast. He is a past president of the Society of Satellite Professionals International (SSPI) and was elected to the Society's "Satellite Hall of Fame" in 2002 for his pioneering work on the development of global satellite services for broadcasting. He was chairman of the Britain's Royal Television Society (RTS) in 1985, and he continues to serve as a director of the Arthur C. Clarke Foundation.

Florence Montredon
Thales Alenia Space
Cannes, France

Florence Montredon has an Engineering and master's degree in Chemical and Physical Engineering. She has worked at Thales Alenia Space since 2000 and has more than 15 years of experience in technology development and improvement for satellite applications. She started to work on additive manufacturing technologies in 2009 and has worked exclusively on this subject since 2015. She is in charge of material and processes, and coordinates all Thales Alenia Space AM development including partnerships and R&D management.

Stacy A. C. Nelson
Center for Geospatial Analytics
North Carolina State University
Raleigh, NC, USA

Dr. Stacy Nelson is currently an Associate Professor within the Department of Forestry and Environmental Resources, and the Center for Geospatial Analytics at North Carolina State University. Dr. Nelson received a B.S. from Jackson State University, M.A. from College of William and Mary's Virginia Institute of Marine Sciences, and Ph.D. from Michigan State University. His research centers around the use of remote sensing and GIS technologies to address questions of land use/cover change on aquatic systems at both regional and local scales. He has worked with several federal and state agencies including the Earth Systems Science Office at the Stennis Space Center in Mississippi, the NASA-Regional Earth Science Applications Center (RESAC), and the MI and NC Department of Environmental Quality. Dr. Nelson currently teaches as part of the spatial analyses curricula related to GIS science and technologies at NC State University and is active in several professional societies.

Bruno Perrot
SES
Betzdorf, Luxembourg

Senior Manager System Architect in SES, has a broad background, encompassing the engineering, services, and business sides of a worldwide satellites operator, leader in its domain. With 25 years in the field of space telecommunications industry, he has experienced first hand the dynamics of the rapidly evolving and changing global telecommunications marketplace.

The early years in Mr. Perrot's career were spent with Aerospatiale, France, a period during which he became a key person of the communication engineering staff. He joined Alenia Aerospazio, Italy, in 1991, as a payload manager, and then Telespazio in 1999. From 2000, he worked for SES ASTRA and moved to SES Global in 2005. In 2008, he took the responsibility of the SES European fleet. Since 2016, he is in charge of the System Architecture in SES Engineering.

Bruno graduated at ENSEA of Paris where he earned an honors degree in Telecommunication Engineering application and holds a management degree at INSEAD of Fontainebleau.

As a member of the Technical Committees on Communication Satellites of the AIAA and on the Ka-band and Broadband Communication Conference, he has chaired and served on numerous industry panels, seminars, and roundtable discussions across the globe and is fluent in English, French, Italian, and Spanish.

Susan K. Runco
Code KX, Astromaterials Research and
Exploration Science Directorate
NASA Johnson Space Center
Houston, TX, USA

Susan K. Runco completed her M.S. in Oceanography and Meteorology at the Naval Postgraduate School in 1986. She is currently the principal investigator for the Crew Earth Observation Payload on the International Space Station at NASA Johnson Space Center in Houston, Texas, USA. She has provided Earth science training to astronauts and participated in Crew Earth Observations since 1988. As PI, her interests include developing imaging techniques and imagery collections for broadening utilization of astronaut photography in the areas of Earth science, education, and public awareness of Earth and space.

Michael J. Rycroft
Cambridge Atmospheric, Environmental and Space Activities and Research (CAESAR) Consultancy Cambridge, UK

Prof. Michael Rycroft is visiting senior fellow at Bath University, UK, and emeritus faculty member at the International Space University, Strasbourg, France. Previously he was Professor of Aerospace at Cranfield University; Head, Atmospheric Sciences Division, British Antarctic Survey, Cambridge; Lecturer in Physics Department at Southampton University; and Postdoctoral NAS/NRC associate at NASA Ames Research Center, California. He obtained his honorary D.Sc. from De Montfort University, Ph.D. from Cambridge, and B.Sc. from Imperial College London. Prof. Rycroft has carried out research on solar-terrestrial physics, ionospheric and magnetospheric physics, and atmospheric studies. He is editor of the *Cambridge Encyclopedia of Space* (1990), *Journal of Atmospheric and Solar-Terrestrial Physics* (1989–1999), and *Surveys in Geophysics* (2002–present). He is author of more than 240 scientific publications and editor of more than 40 books and special issues of journals.

Haruhisa Shimoda
Research and Information Center
Tokai University Shibuya-ku
Tokyo, Japan

Haruhisa Shimoda graduated from the University of Tokyo in 1967 and the graduate school in 1972 where he got the Ph.D. He joined Tokai University in 1972 as a lecturer of the Faculty of Engineering. He became an associate professor in 1974. In the same year, he also joined Tokai University Research and Information Center as a senior researcher. In 1985, he became a professor. In 1994, he also joined National Space Development Agency (now JAXA) as an invited scientist and has been working as the program scientist of ADEOS, and later also for GCOM. In 2000, he became the director of Tokai University Space Information Center. He has been engaged in the field of remote sensing from 1974. His main achievements are development of remote sensing image analysis sytem including both hardware and software, development of IMG on ADEOS, etc. He is the chair of General Counsel of GCOM and the Chair of RA committee of GOSAT. He retired from Tokai University as well as from JAXA on March 2015. He is now a guest professor in Tokai University Research and Information Center.

Ramesh L. Shrestha
NSF National Center for Airborne
Laser Mapping (NCALM)/Department of Civil and
Environmental Engineering University of Houston
Houston, TX, USA

Ramesh L. Shrestha Ph.D., is a Hugh Roy and Lillie Cranz Cullen University Professor at the University of Houston (UH) and leads the GSE graduate research and academic programs. He is also PI and the director for the NSF-funded National Center for Airborne Laser Mapping (NCALM), which is jointly operated by UH and the University of California-Berkeley. Dr. Shrestha's main research activities are associated with the application of advanced geodetic and remote sensing techniques, particularly airborne laser swath mapping (ALSM, aka LiDAR) and digital mapping.

Vern Singhroy
Canada Centre for Remote Sensing
Natural Resources Canada
Ottawa, Canada

Dr. Vern Singhroy is the Chief Scientist of the Canadian Space Agency RADARSAT Constellation Mission scheduled to be launched in 2018. As a senior research scientist at the Canada Centre for Remote Sensing, he is internationally recognized as a world leader on the uses of remote sensing images for geohazard, geological, and geoengineering applications. Dr. Singhroy received his Ph.D. in Environmental and Resource Engineering from the State University of New York, Syracuse, and is a professional engineer in Ontario, Canada. He has published over 300 papers in scientific journals, proceedings, and books. He is also the coeditor of four books including of the *Encyclopedia of Remote Sensing* and was the editor-in-chief of the *Canadian Journal of Remote Sensing*. He is a Professor of Earth Observation at the International Space University in Strasbourg, France, since 1998 and an Adjunct Professor in Planetary and Space Sciences at the University of New Brunswick in Canada. Dr. Singhroy received the prestigious Gold Medal Award from the Canadian Remote Sensing Society and the (2012) Queen Elizabeth Diamond Jubilee Medal for his contribution to the Canadian and International Remote Sensing.

Andrew Stanniland
Inmarsat Ltd
London, UK

Andrew has over 20 years' experience in the satellite communications industry, with a background in systems and aeronautical engineering. He was responsible for military satcom programs in the UK, Australia, and South Korea before rejoining the team responsible for developing the UK's Skynet 5 Milsatcom program in 1997; initially with responsibility for the communications analysis before leading the Service Design that eventually evolved into the company known as Paradigm Secure Communications (now part of Airbus Defence and Space). During the Skynet 5 proposal and contract negotiation phase, Andrew was responsible for liaising with investors and satellite insurers to ensure that the complex and groundbreaking PFI deal was bankable.

Between the formal formation of Paradigm in October 2003 and January 2012, Andrew led the overseas business development activities and since 2007 was the senior VP responsible for all business development, sales, and marketing. This resulted in the first sales of commercially provided, military grade, satellite communications services into the defence customers of NATO, Canada, Europe, and, the US

DoD. In 2010, he became responsible for product development across all Airbus Defence and Space government communications customers.

Since 2012, Andrew has been the VP in charge of business development for Inmarsat's Global Government Business unit and has led the expansion of Inmarsat's business into new geographical markets as well as diversifying from defence customers into other government sectors.

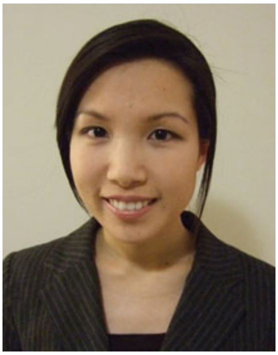
He is also an alumnus of the International Space University, having attended the Summer Session Program in 1997 in Houston, TX.

William L. Stefanov
Earth Science and Remote Sensing Unit
Astromaterials Research and Exploration
Science Division, Exploration Integration and
Science Directorate
NASA Johnson Space Center
Houston, TX, USA

William L. Stefanov completed his Ph.D. in Geology at Arizona State University in 2000. He is the technical lead and manager of the Earth Science and Remote Sensing Unit, and serves as the International Space Station Program Scientist for Earth Observations, at NASA Johnson Space Center in Houston, TX. In association with the International Space Station (ISS) Program Science Office, he works with instrument science/operation teams to coordinate collection, distribution, and analysis of remotely sensed data from the ISS in response to catastrophic events such as volcanic eruptions, earthquakes, and flooding. Dr. Stefanov's research interests include using remotely sensed data for investigation of geohazards, geomorphology, and surface material characterization; studying climate change with application to regional climate adaptation planning; and assessing the role of humans as geological agents on the landscape. He is an active proponent of geoscience education and public outreach using remotely sensed data.

Dan Swearingen
Arlington, VA, USA

Daniel Swearingen is an engineering consultant who worked at Communications Satellite Corporation (COMSAT) for 30 years. After earning degrees at Georgia Tech and Stanford and completing his military service, he worked for 3 years at ITT Telecommunications before joining COMSAT in 1970. After working in the spectrum utilization and special studies departments, he joined the COMSAT's mobile satellite systems group in 1973. There he served as a system design group manager for the first maritime mobile communications satellite system, MARISAT, launched in 1976. The MARISAT standards and protocols he and his group developed became the basis for the first International Maritime Satellite (INMARSAT) system. In 1980–1981, shortly after the INMARSAT was established, he worked with the small startup staff in London to plan the initial space and ground components of the multinational cooperative's system prior to its operational cutover in 1982. In the subsequent years from 1981 to 1996, he served as a member of the Inmarsat technical advisory committee and proposed key system architecture features that were adopted for the new INMARSAT systems. After leaving COMSAT, Mr. Swearingen served for several years as an adjunct lecturer at George Washington University and has been a consultant specializing in communications satellite systems.



Dr. Su-Yin Tan
Senior Lecturer in Geomatics
Director, Applied Geomatics Research Laboratory
Department of Geography and Environmental Management
University of Waterloo (Canada)
Ph.D. University of Cambridge (UK)
M.Sc. University of Oxford (UK)
M.A. Boston University (USA)
B.Sc. (Env) University of Guelph (Canada)

Dr. Su-Yin Tan is a Senior Lecturer in the Geomatics Program and Director of the Applied Geomatics Research Laboratory at the University of Waterloo, Canada. She is a Faculty Member and chairs the Academic Council of the International Space University (ISU). At ISU, she serves as Core Co-Chair and Space Applications Department Chair of the Space Studies Program (SSP) and lectures at the Masters of Space Studies (MSS) program. Dr. Tan's specialization is in geographic information systems, remote sensing, and spatial data analysis. She is a distinguished Gates Scholar and received her Ph.D. degree from the University of Cambridge (UK), two master's degrees from Oxford University (UK) and Boston University (USA), and B.Sc. (Env) from the University of Guelph (Canada). She was previously a Visiting Fellow at the University of Cambridge; Visiting Researcher at the Environmental Change Institute, University of Oxford (UK); and member of Christ Church College. Dr. Tan has an interdisciplinary background in the environmental sciences and spatial analysis methodologies in a range of application areas, such as climatology, ecosystem modelling, and remote sensing. In the past decade, Dr. Tan has received over 16 awards and scholarships, including a prestigious Presidential University Graduate Fellowship and Overseas Research Scholarship. Dr. Tan received the prestigious 2014 Outstanding Performance Award for exceptional contributions to teaching and scholarship, which recognizes the top professors at the University of Waterloo. Originally from Papua New Guinea, she has a diverse international background and built a record of teaching and research excellence in North America, Australia, Asia, South America, and Europe.

Paul T. Thompson
University of Surrey
Guildford, Surrey, UK

Dr. Thompson is a senior researcher in the Centre for Communications and Systems Research (CCSR) at the University of Surrey. He has an extensive background in satellite communications and spent 30 years with British Telecom where he led the Technology Development Division covering a wide range of international communications disciplines. During part of this time he was seconded to the SHAPE Technical Centre in The Hague where he was involved in the development and inorbit testing of NATO satellites.

Subsequent to his time at BT, Dr. Thompson has worked with ERA Technology, developing a range of radio communications products and was also the Director of Teledesic, UK.

In addition to research and teaching roles, he currently participates in the standards areas of DVB and the European Telecommunications Standards Institute (ETSI).

Dr. Thompson was the first UK delegate to become chairman of the INTELSAT Board of Governors Technical Committee (BG/T). He is a fellow of the British Interplanetary Society where he played several roles (member of BIS Council 1990–2002, president 1994–1998).

He was a visiting professor of engineering design at the Engineering Science Faculty of the University of Oxford, a role supported by the Academy of Engineering (1993–2001).

He has been a member of the editorial panel of the *International Journal of Satellite Communications* since 1982.

Dr. Thompson is also a senior member of AIAA since 1991.

Cynthia F. van der Wiele
US Environmental Protection Agency
Region IV, NEPA Program Office – NC Field Office
Research Triangle Park, NC, USA

Dr. Cynthia F. van der Wiele is a senior physical scientist with the US Environmental Protection Agency (USEPA), Region 4, NEPA Program Office. Previously, she was a research associate and adjunct faculty at North Carolina State University. Her research interests include the development of high accuracy land use/land cover classifications for analysis and improved land use and conservation planning and policies. Dr. van der Wiele received her B. S. in Engineering and Masters of Landscape Architecture from North Carolina State University, a Masters in Forestry, and a Masters in Environmental Economics and Policy from Duke University, and her Ph.D. in Community and Environmental Design from North Carolina State University. She is active in several national and international professional societies.

John L. Walker
Lockheed Martin Space Systems Company
Denver, CO, USA

John Walker has over 30 years of technical and managerial achievements in the design and development of advanced communication systems operating from ELF to EHF. He is currently a Senior Principal at Lockheed Martin RF Payload Center of Excellence pioneering the next generation digital flexible payloads and leading new business pursuits. He has also held positions at Space Systems Loral (SS/L), Lockheed Electronics, and Hughes Aircraft Company responsible for both space and terrestrial developments.

At Space Systems/Loral (SS/L), he held positions of the director of RF Electronics responsible for the development, design, manufacture, and test of space-borne RF electronics equipment and the director of advanced development responsible for the end-to-end communication systems activities within SS/L. At SS/L, he led the development and execution of the first two-way ground-based beam forming system from concept through onorbit integration, verification, and deployment.

M. Justin Wilkinson
Code XI/ESCG, Texas State University,
San Marcos, on contract to the
Astromaterials Research and Exploration Science
Directorate, NASA Johnson Space Center
Houston, TX, USA

M. Justin Wilkinson was born in South Africa and holds a Ph.D. in Geomorphology from the University of Chicago. Since 1988, he has held the position of astronaut trainer in geography, geology, and geomorphology with the Crew Earth Observations payload at the NASA Johnson Space Center in Houston, Texas, USA. His research interests include fluvial and desert geomorphology, especially the interface between geomorphology and sedimentology, the role of landscapes in species evolution, and geomorphic analogs for planetary geology. His teaching interests have resulted in the publication of books of astronaut imagery with National Geographic and a bilingual atlas of Costa Rica, for which he was awarded NASA's Public Service Medal. He has a patent in the area of automated identification of fluvial landscapes and is chief editor of a new multiauthor book on large continental fluvial systems known as megafans.

Kimberly Willis
Code XI2/, Astromaterials Research and Exploration
JACOBS-JETS Contract, Oceaneering Space Systems
NASA Johnson Space Center
Houston, TX, USA

Kimberly J. Willis completed her M.S. in Physical Science, with a concentration in geology, at the University of Houston–Clear Lake. She began her career at the NASA Johnson Space Center in Houston, TX, over 30 years ago where she first worked in the Lunar Laboratory with samples returned from the Apollo missions. Kim transitioned into Earth observations where she held progressively more responsible positions until returning to Astromaterials Curation in 2013. She is currently the manager for Astromaterials Curation and Science Engagement and Outreach contractor personnel. She also holds an adjunct faculty position at the University of Houston–Clear Lake in the School of Science and Computer Engineering, where she teaches the fundamentals of earth science.

Part I

Satellite Communications

Satellite Applications Handbook: The Complete Guide to Satellite Communications, Remote Sensing, Navigation, and Meteorology

Joseph N. Pelton, Scott Madry, and Sergio Camacho-Lara

Contents

Abstract

This chapter introduces what is meant by the term "applications satellite" and addresses why it makes sense to address the four main space applications in a consolidated reference work. This handbook employs a multidisciplinary approach and thus includes technical, operational, economic, regulatory, and market perspectives. These are all key areas wherein applications satellites share a great deal in common. This commonality can be seen in terms of

---

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

S. Madry
Global Space Institute, Chapel Hill, NC, USA
e-mail: Scottmadry@mindspring.com

S. Camacho-Lara
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC) Santa María Tonantzintla, Puebla, México
e-mail: sergio.camacho@inaoep.mx

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_91

spacecraft systems engineering, in terms of launch services, in terms of systems economics, and even in terms of past, present, and future market development.

This is not to suggest that there are not important technical and operational differences with regard to the four prime areas of satellite applications, namely, communications satellites, remote sensing satellites, global navigation satellites, and meteorological satellites. Such differences are addressed in separate sections of the handbook.

Yet in many ways there are strong similarities. Technological advances that come from one type of applications satellite can and often are applied to other services as well. The evolution of three-axis body-stabilized spacecraft, the development of improved designs for solar arrays and battery power systems, improved launch capabilities, and the development of user terminal equipment that employs application-specific integrated circuits (ASIC) are just some of the ways the applications satellites involve common technologies and often in a quite parallel manner.

All four types of applications satellites provide key and ever-important services to humankind. Around the world, people's lives, their livelihood, and sometimes their very well-being and survival are now closely tied to applications satellites. Clearly the design and engineering of the spacecraft buses for these various applications satellite services as well as the launch vehicles that boost these satellites into orbit are very closely akin. It is hoped that this integrated reference document can serve as an important source of information that addresses all aspects of application satellites from A to Z. This handbook thus seeks to address all aspects of the field in a totally comprehensive basis.

This Handbook of Satellie Applications thus covers spacecraft and payload design and engineering, satellite operations, the history of the various types of satellites, the markets, and their development – past, present, and future, as well as the economics and regulation of applications satellites, and key future trends.

Keywords

Applications satellite • Committee on the Peaceful Uses of Outer Space (COPUOS) • Earth observation-Global Navigation Satellite Systems • Launch services • Markets for satellite applications • Military satellite communications • Precision Navigation and Timing • Satellite broadcasting • Satellite communications • Satellite meteorology Satellite navigation and positioning • Satellite remote sensing • Scientific satellites • United Nations

Introduction

Artificial satellites have now been around for more than a half century. The launch of Sputnik in October 1957 ushered in the space age and confirmed Sir Isaac Newton's theoretical explanation of how an artificial satellite could be launched into Earth orbit. Today the world of satellites can be divided into two broad areas – scientific satellites and applications satellites.

Scientific satellites explore and help humanity acquire new information about our world, our solar system, our galaxy, and the great cosmos within which we exist. The scientific satellites explore radiation from the Van Allen Belts to cosmic radiation, from "black holes" to even the formation of stars and galaxies. Scientific missions engage in Geodesy to measure our Earth and tectonic movements. They study the workings of the Sun and the characteristics of the our Solar System, including the planets, their moons, asteroids, comets, and the Oort cloud well beyond the orbit of Pluto. Astronomical observatory satellites explore the stars and galaxies and give us a view of the Universe near its beginning as well as help us to discover exoplanets in other star systems.

This handbook, however, is about the applications satellites that provide practical and business services to people here on Earth. These are the communications satellites, the remote sensing satellites, the space navigation and positioning satellites, and the meteorological satellites that truly serve humankind.

Thousands of applications satellites have now been launched over the past half century. These practical satellites now represent a huge global industry. These satellites are a part of our everyday lives whether we know it or not. Every time you hear a weather report or every time you use a GPS or Glonass device to navigate your car, you are relying on an applications satellite. Services such as worldwide news, satellite entertainment channels, coverage of sporting events, communications to ships at sea or aircraft in the skies, and many more vital services frequently depend on satellites. Farmers now rely on satellites to irrigate their crops, add the right amount of fertilizer, or detect a crop disease. Fishing fleets use satellites to know where to fish. Energy and resource companies employ satellite imaging to know where to dig or drill. Efforts to combat global warming, preserve the Ozone layer that is essential to life on Earth, and other activities to sustain the biodiversity of plant and animal life on our planet all depend on applications satellites. Responding to major disasters routinely involves analysis of satellite imagery and mobile satellite communications.

This book is thus a comprehensive reference work about the practical use of satellites to serve humanity and make our lives better. Application satellites are thus machines in orbit offering vital services to provide vital capabilities to humankind.

The multibillion-dollar (US) world of commercial satellite applications and services continues to expand each year. This means that the technology is becoming more sophisticated and reliable and the practical applications ever broader. Commercial satellite applications and satellite technology are both becoming more sophisticated and efficient. This is particularly true in terms of finding more and more applications in different fields and in the expanded use of automation and the application of expert systems and artificial intelligence to allow more autonomous operation of satellites in outer space. Precise navigation, positioning, and timing satellites are key to the global synchronization of the Internet. The sizes of satellite applications markets are now measured in the hundreds of billions of dollars (US). Virtually every country and territory in the world relies on applications satellites for multiple space-based services.

The diversity of the submarkets within the field we have defined as "satellite applications" continues to expand and becomes more complex. Indeed, in view of this growing dependency on space applications and the expanding number of satellites in near-, middle-, and Geostationary Earth orbits, the international community, through the United Nations and other forums, is working to ensure the long-term sustainability of activities in outer space.

Today the field of "satellite applications" includes at least (1) satellite communications; (2) satellite broadcasting; (3) satellite precision navigation, positioning, and timing; (4) Geostationary and lower earth orbit satellite meteorology; (5) remote sensing and Earth observation; and (6) space-based information systems. And this is just the beginning. The above-cited satellite applications activities generate other major space-related activities and industries which are themselves of significant size.

For instance, the field of commercial satellite applications creates a substantial part of a multibillion-dollar (US) launch vehicle industry around the world. It also creates yet another multibillion-dollar market for earth station antennas, very small-aperture antennas, microterminals, and handheld satellite transceivers. Finally there are also important ancillary markets that also feed off of commercial "satellite applications." The supporting industries include

1. *Space-related insurance and risk management industries* (such coverage requires expenditures equivalent to 10–20 % of the value of the satellite and its launcher).
2. *Engineering, design, reliability testing, and regulatory support activities.* Key technical support is required to design new systems and carry out research related to new space and ground systems. (These engineering companies and research organizations prepare detailed technical specifications for satellite systems and work with specialized law firms to prepare requests for regulatory approvals and frequency assignments and allocations at the national, regional, and even global levels.)
3. *Financial institutions, investment banks, and underwriting corporations*. These institutions help to raise the billions of dollars in capital needed to build and launch the satellites and deploy hundreds of millions of earth station antennas, receive only terminals and two-way satellite telephones around the globe.
4. *Marketing and sales organizations.* These companies help with the sales of satellite applications services around the world to literally billions of people who depend on these satellites for severe weather warnings, for radio and television services, for Internet connection and synchronization, for navigational and routing information, and for vital information for farming, fishing, mining, or urban planning.

Today the overall field of "satellite applications" thus represents not only the primary sectors that build or operate commercial space satellite systems and launch them but a huge supporting workforce as well. These include hundreds of service, engineering, manufacturing, specialized banking, and insurance companies. These supporting service industries represent an important set of commercial enterprises representing billions of dollars (US) in revenues (see Fig. 1).

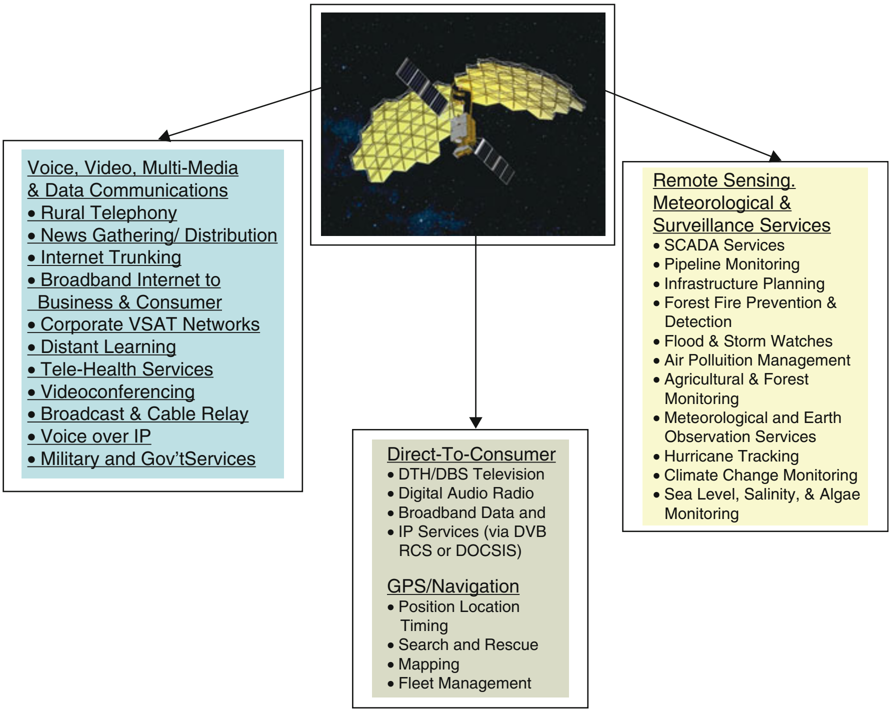


Fig. 1 The wide range of satellite applications services provided from space today (Graphic courtesy of the author)

Without all of these diverse satellite applications providing meteorological and weather information, communications, broadcasting, navigational, remote sensing, and supporting space-based information services, the world we live in would be greatly different. Without these systems, for instance, many more lives would be lost to hurricanes, tornadoes, tsunamis, earthquakes, volcanoes, and other violent acts of nature. Without these systems there would be less effective global communications systems. Satellite telecommunications systems support telephone, Internet, and other IP-based information services across virtually all of the world's 200 plus countries and territories. These satellite systems also provide communications to ships and offshore platforms and buoys in the seas as well as to the Polar region and to aircraft in the skies. Applications satellites are an important part of the world's search and rescue (SAR) infrastructure for downed pilots, stranded passengers and crews of shipwrecked vessels, or people lost in wilderness areas or subject to natural disasters.

Other space navigational, positioning, and precision timing satellite systems provide key real-time information to all parts of the world whether on land, the seas, or in the air, including tracking of goods in our global transportation network. These space systems, with increasing accuracy, can tell us where people or vehicles or buildings or a myriad of other things are located for a wide range of applications.

Without broadcasting satellites there would be limited television, radio, and other broadcasting services around the world.

Well over a billion people receive television, radio, or communications live via satellite to their homes, offices, or vehicles. Satellite services have become so pervasive that they have almost disappeared from the public consciousness as something unique and special. The use of outer space has become almost commonplace in a span of a half century. Much like electric motors or batteries, the vast and extensive use of satellites in our everyday lives has thus often become "invisible." These key machines in the skies help us to predict the weather, receive an entertainment or news broadcast, or connect to the Internet across the globe in a synchronized way. Satellites also let us know where we are in our cars and how to get to our destinations, take off and land safely, protect us from fires, or help us to have access to a wide range of resources from apples to zirconium.

The growth of the satellite applications market will continue to be rapid and diverse – and for many years to come. The first graph below shows the evolutionary nature of the markets from over a decade ago. Even then, in the late 1990s, the satellite communications industry in terms of satellite and earth station manufacturing, launch services, plus communications services when combined represented total annual revenues of about $80 billion. This compares to well over $200 billion today. The global positioning system (GPS) and other precision navigation and timing systems have exploded as a key market especially after GPS receivers, and services related to GPS, were combined together in smartphones. Commercial space transportation to support these industries is also a significant part of the launch market. Although the remote sensing market remains quite small, on a relative scale, its impact on the global economy is actually huge (Fig. 2).

The combined revenues for 2014 – just for global commercial satellite communications services – now totals $123 billion (U.S.). If one then adds in revenues for satellite launches, the manufacturing of satellites, manufacturing of earth stations, and various types of user terminal equipment, the total climbs another $80 billion to over $200 billion (U.S.) (Satellite Industry Association, The Tauri Group 2014).

If on top of this one then also adds technical consulting support, licensing fees, and insurance and risk management costs, the total revenues associated with the commercial communications satellite industry – for service revenues plus all other costs and expenditures – i.e., the net satellite telecommunications industry annual revenues for 2014, climb to about $230 billion per annum. These revenue figures have grown on average about 11 % per annum since 2004 but only at about 4 % in the most recent year (Fig. 3).

If one were then to add in the additional revenues associated with governmental and defense-related communications systems and the cost of governmentally operated Geosynchronous meteorological satellite services, plus all of the revenues associated with meteorological satellite networks, precision navigation, positioning and timing services, and remote sensing, then the annual financial turnover for the entire satellite applications industry would exceed $350 billion or over a third of a trillion dollars. It is thus safe to say that overall the combined field of "commercial satellite applications" represents quite a large global industry. If one takes into

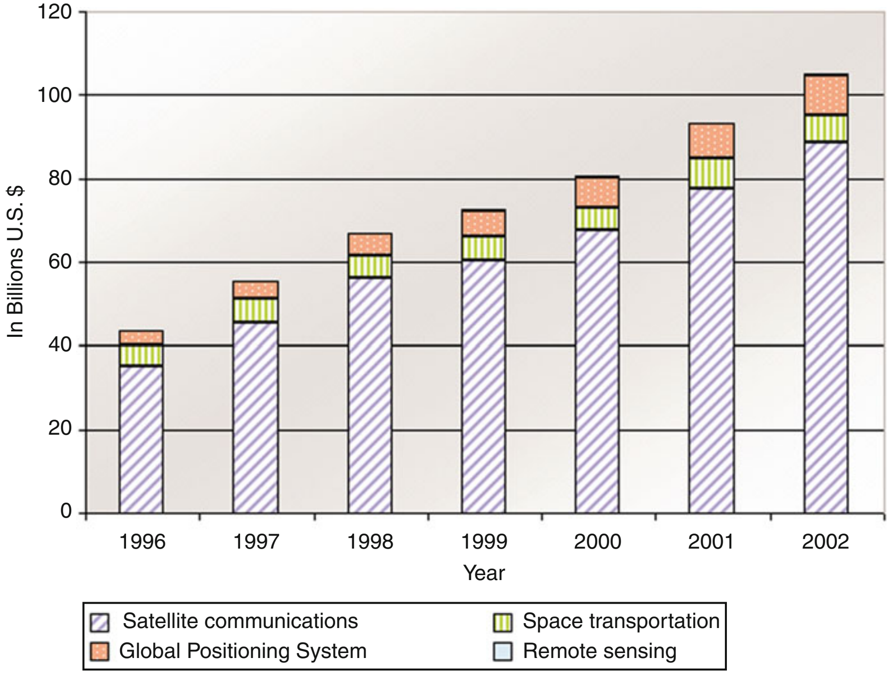


Fig. 2 Early stages of growth of the commercial satellite applications market (U.S. Department of Commerce, *Trends in Space Commerce, Office of Space Commercialization* (U.S. Department of Commerce, Washington, DC, 2002))

account the industries that application satellites support such as broadcasting, banking and insurance, airline, train, and bus travel, shipping, farming, fishing, mining, etc., there is virtually no part of the global economy that is not affected.

Further this is an industry that has shown consistent growth for several decades and has continued to grow even in times of global recession.

The Evolution of Commercial Satellite Applications

With the launch of the Sputnik satellite in October 1957 people began to think of outer space not as something in science fiction novels but as a real and increasingly important activity. Everett Edward Hale as early as 1867, when he wrote *The Brick Moon*, speculated on the use of artificial satellites for communications, navigation, and remote sensing. But as of the late 1950s, scientists and engineers began to conceive of practical ways to utilize artificial satellites for needed services. The first application was satellite telecommunications. In the late 1950s and early 1960s, international communications capacity for overseas links was very limited in scope and the per-minute rate of a telephone call was quite high. (Submarine cables for

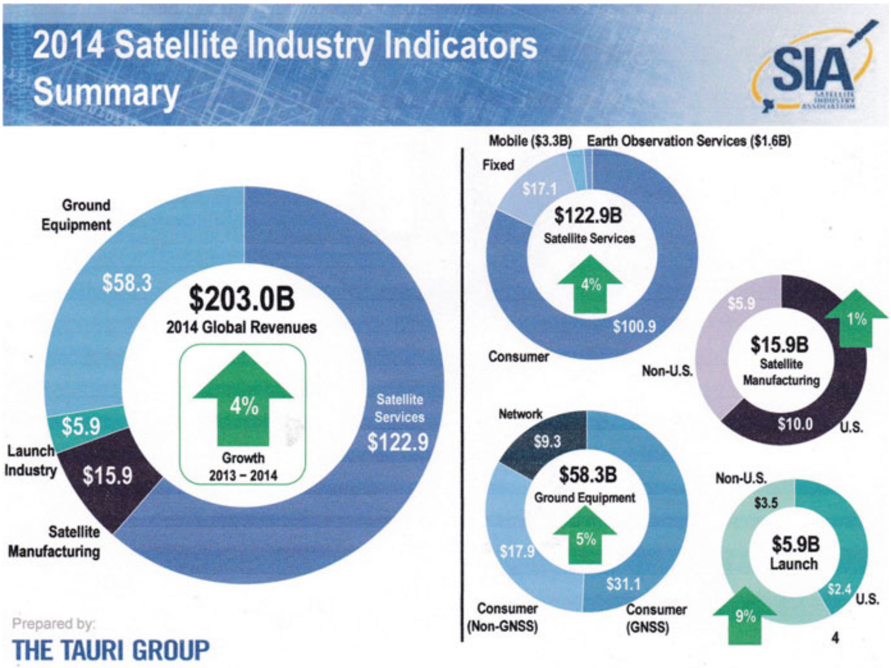


Fig. 3 Annual revenues for communications satellite market for 2014

voice communications could only connect 36–72 voice circuits at a time and could not handle even low-quality black and white television transmissions on a live basis. One might have to pay $20–50 a minute for an overseas telephone connection.)

In the various sections that follow, the history of satellite applications is provided in detail, but the following provides a brief overview. The first practical application of satellites was in the field of telecommunications. A series of experimental satellites were launched in the early 1960s to test the feasibility of communications satellites for commercial purposes. These satellites, known variously as SCORE, Courier 1B, Echo, Telstar, Relay, and Syncom, proved vital to the design of the operational systems that were to follow. These early experimental satellites helped space system designers to discard the idea of using passive satellites for telecommunications. Echo was a metallic coated balloon launched for meteorological experiments, but scientists also bounced electronic signals off its reflective surface without amplification. These experiments confirmed that this type of "passive satellite" represented much too low a capacity for commercial needs. These experiments and many others – particularly the Syncom satellites – showed that deploying satellites into Geosynchronous orbit and providing telecommunications services from orbiting spacecraft was technically feasible (Pelton 1974).

This special Geo orbit (sometimes called the Clarke orbit in honor of Sir Arthur Clarke who first suggested this orbit for communications satellites) allowed virtually complete global coverage with only three satellites and eliminated the need for Earth

Station antennas to track rapidly across the sky. This is because the Geo orbit allows the satellite to “seem to hover constantly” above the same location over the equator. (Note: A more formal definition of the Geostationary orbit is an Earth orbit having zero inclination and zero eccentricity, whose orbital period is equal to the Earth’s sidereal period. The altitude of this unique circular orbit is very close to 35,786 km.)

These various experiments led to the deployment of operational telecommunications satellites in 1965. The three initial satellite telecommunications systems, all launched in 1965, were the Intelsat system that deployed the “Early Bird,” the low-orbit Initial Defense Communications Satellite System deployed by the US defense department, and the Molniya satellite system for the USSR. There were three Molniya satellites deployed into highly elliptical orbits that were suited to northern latitude coverage over Russia and to the satellite countries known as Soviet Socialist Republics.

The rest is history. The Intelsat satellite system grew into a truly global network that now serves nearly 200 countries and territories around the world. A number of national satellite systems were launched to meet domestic communications needs (particularly to meet television and radio broadcasting needs and services to rural and remote areas). In time regional satellite systems evolved and yet other systems were deployed to meet maritime, aeronautical, and land mobile communications. Military-, security-, and defense-related satellite systems were also launched to meet the specialized needs of military agencies. Today there are a huge and growing number of communications satellites in orbit and thousands of application satellites have been launched since the late 1950s. Some of these application satellites are in Geosynchronous orbit, others are in medium earth orbit, and yet others are deployed as large constellations in low earth orbit. Currently some of the latest versions of low earth orbit constellations envision single networks such as One Web and the Space X Leo system that would include hundreds and perhaps several thousands of small satellites within a single network.

Some of these are multipurpose and support various types of telecommunications services for telephone, radio, and television broadcasting or distribution, plus data networking and Internet-related services. Other satellites are designed and optimized for mobile communications for land, sea, or aircraft communications.

Close on the heels of the telecommunications satellites came other types of applications satellites. Military reconnaissance satellite systems were a very high early priority for both the Soviet Union and the USA in the cold war. Fully half of the first 20 Soviet Cosmos series space launches were for military Zenith imaging systems, and the US Corona satellite system was developed in secret starting in 1959. The Corona program was started under the camouflage of public statements that these satellites were scientific payloads. The Corona program was not publicly acknowledged for many years, and not until well after being out of service. These “spy” satellites set the stage for remote sensing and weather satellites.

First came the weather or meteorological satellites, which were initially developed in order to provide weather and cloud cover information for the military imaging systems. The US President’s Science Advisory Committee reported in 1958 that “The satellite that will turn its attention downward holds great promise for meteorology and the eventual improvement of weather forecasting.” But the potential benefit of weather satellites was evident and the first civil weather satellite launched was the TIROS (Television and Infrared Observation Satellite), which started as a defense department program and was transferred to the new NASA in April of 1959. Its first images in 1960 provided a synoptic view of weather patterns, sea ice, and other features that were immediately analyzed on the ground to great effect, and were the first in an unbroken series of weather satellites that operate to this day. TIROS was in a low Earth (435 miles or approximately 700 km) orbit, but the potential for a permanent Geostationary orbital view was clear. The first Geo weather satellite was the US GEOS-1, launched on October 16, 1975. This satellite demonstrated the benefit of the Geostationary orbit for weather observation. Over the past 30 years, additional weather satellites have been launched by Europe, Japan, India, Russia, and China. These now provide a constant global view of our world and have revolutionized our understanding of global weather patterns and our ability to accurately forecast the weather.

This was followed by remote satellite sensing systems and specialized Earth observation satellites, with the launch of the Earth Resources Technology Satellite in July 1972 (later renamed Landsat 1). The Landsat series led the way in the development of dedicated Earth resources satellites that were specifically designed for a wide range of applications such as agriculture, forestry, and water resources. These systems have continued to develop and have improved their capabilities, with spatial resolution improving from 80 m with Landsat 1 to under 0.35 m with the current ultrahigh-resolution systems.

The most recent class of satellite applications to evolve are those associated with satellite navigation, also referred to as precision navigation and timing systems (PNT) or Global Navigation Satellite Systems (GNSS). These systems were first devised to assist with military and defense-related purposes such as targeting and mapping. This type of satellite system included the early US TRANSIT and Soviet Tsikada Doppler-based systems, which were first fielded in 1959 (Transit) and 1974 (Tsikada). Today, however, there are a wide range of commercial uses for space navigation satellite systems, and these satellites actually represent a multibillion-dollar industry worldwide. This market has grown rapidly and continues to develop new uses. Next-generation system development started with the US Navstar GPS system, first developed in 1973 and fully deployed in 1994. Europe, Russia, India, Japan, and China are all developing and launching their own advanced systems, and the future of this class of satellites is bright.

Later in this handbook the more detailed histories of these various types of applications satellites are provided for those that would like to know how these various types of satellite systems evolved over time.

More than a thousand applications satellites are now in low, medium, or Geosynchronous orbits, and these are being used to provide one or more types of commercial satellite services. Indeed there have been a number of instances where a satellite built for one type of application such as telecommunications had another “package” or “hosted payload” added to the satellite to provide meteorological imaging such as was the case with an Indian “Insat” satellite. Sometimes an operational satellite will have an experimental package attached to test out a new technology. One example of this was the Orion international communications satellite that had an experimental intersatellite link (ISL) package added to it for performance testing. Today most satellites are designed for a particular application because the frequency bands (or radio frequency spectrum) allocated for space applications through the International Telecommunication Union (ITU) are typically different for different types of applications.

Later in this handbook specifics of these allocations are provided. Nevertheless satellites can have a primary payload for one application and then have one or more secondary payload(s) for other applications or to experiment with a new technology or new frequency band. In short a satellite may operate in many different frequency bands. An example of this is the US National Oceanic and Atmospheric Administration (NOAA) polar orbiting POES satellites which also carry the Cospas-SARSAT search and rescue and ARGOS telemetry systems.

For some 50 years now there have been applications satellites in the Earth orbit. Thousands of these various types of satellites have been launched and some of them have come back from their orbit and burned up in the atmosphere or crashed backed into Earth. Others have been pushed out into space above a Geosynchronous orbit where they will stay for millions of years. There are, however, many thousands of defunct and derelict satellites or parts of satellites or launchers still in orbit known as orbital debris or more formally as "non-functioning space objects." Currently on the order of 22,000 "orbital debris" elements about the size of a tennis ball or larger are known to be in orbit and millions of microscopic elements are present – especially in low earth and polar orbits.

The problem of orbital debris crashing into a satellite, space station, or other active space object is thus a growing concern. The September 2015 Solyut mission to the International Space Station was reprogrammed from a 7 h flight to a 2 day journey just to be sure to avoid a collision with space debris.

The Chinese shooting down one of their defunct weather satellites in 2007 that created over 2000 new trackable debris elements and then the 2009 crash between a defunct Russian Kosmos satellite and an active Iridium communications satellite in 2009 both have served to highlight space debris concerns. The United Nations Committee on the Peaceful Uses of Outer Space (COPUOS) that has been addressing this issue for some time has now agreed as of June 2007 to voluntary procedures to reduce the threat of orbital debris creation in future years.[1]

What Does the Term "Satellite Applications" Mean and Why Consider It in a Unified Way?

Why a Handbook of Satellite Applications? One can find handbooks and reference sources in many areas that include the various "fields" of satellite applications. There are reference handbooks on satellite telecommunications, satellite broadcasting,

---

[1] United Nations Committee on the Peaceful uses of Outer Space Voluntary Guidelines on Orbital Debris, http://www.orbitaldebris.jsc.nasa.gov/mitigate/mitigation.html

satellite-based remote sensing, or Earth observation. There are also some reference materials on satellite-based meteorology and space navigation. The key to "satellite applications" is to recognize that while these space-based services all tend to have a different range of users and require different specialists to use the information, the underlying technology with regard to designing, manufacturing, launching, insuring, financing, and getting regulatory approval for "applications satellites" are in many ways quite similar. As noted above an applications satellite designed for one type of service or application can also have a secondary or even tertiary package (i.e., payload) to perform operations in an entirely different field. The platform used for telecommunications, broadcasting, remote sensing, Earth observation, meteorological purposes, or space navigation start out to be remarkably similar in their design, manufacture, testing, launch requirements, and, in many cases, even their deployment.

Common Elements of Applications Satellites

An applications satellite's mission is defined by its payload, which carries out its specialized function, but the platform on which the payload resides is quite often similar in terms of structure, power systems, tracking, telemetry, command and monitoring systems, stabilization, positioning, pointing and orientation systems, thermal control systems, and so on. Many manufacturers of applications satellites are now designing various-size platforms that meet various customer needs. At the smaller end of the spectrum the Surrey Space Centre's microsatellite "bus" or platform has been used for communications, IT-related services, remote sensing, and other scientific purposes. With progressive sizes of larger satellites more ambitious objectives can be met by accommodating larger payloads. Commercial aerospace manufacturers, in short, now have progressively larger platforms that support larger and more sophisticated missions.

Over the last 40 years there have been more telecommunications and broadcasting satellites designed, manufactured, and launched than other types of applications satellites. These "communications satellites" have been deployed to commercial, governmental, and military missions. Just because of their sheer volume, the platforms developed for communications satellites have generally tended to characterize the range of platforms available for other purposes in terms of size, structural integrity, maneuverability, lifetime, power systems, and pointing accuracy. At the very beginning a new platform was designed for each satellite. This custom design process was often driven by the fact that satellites were becoming more capable in size and performance. This constant upgrading of performance and the need for larger satellite antennas required greater pointing accuracy.

In another context, the satellites were also being designed for longer life. Several prime characteristics defined the design of these increasingly complicated and larger platforms. These characteristics were (1) increased capability for the payload; (2) prime or peak power requirements over the satellites' lifetime; (3) the pointing accuracy and orientation requirements for the satellite platform and the size and shape of the antennas or functional elements that must be supported by this platform; and (4) the lifetime desired for the satellites operation and the need to remove satellites from their operational orbit at end of life – factors that required more fuel and larger fuel tanks. These four characteristics were the main drivers that led to a wide range of platform designs.

Over time the satellite manufacturers realized, just like manufacturers of automobiles, that one did not have to design a new "bus" or "platform" every time there was a new order for a communications or other type of applications satellite. Thus manufacturers began to standardize classes of platforms from nano- or microsatellites up through giant 10,000–12,000-kg platforms that are built for the largest type of direct broadcast or mobile satellites that can be launched by currently available launch systems such as the Atlas 5 or the Ariane 5 and soon the Ariane 6.

Although the platforms might be quite different in volume and mass, varying from about 200–12,000 kg, they usually contain many of the same features. For many years the trend has been to design larger and larger platforms to accommodate larger payloads associated with Geo satellites and growing global demand. Ironically, today the shift seems to be on to accommodate large-scale LEO constellations with many small satellites with the smallest of platforms – perhaps with the entire satellite including payload being on 100–200 kg in size. The components of these platforms typically have the same components.

These components are batteries (for emergencies and when the satellite might be in eclipse); solar cell arrays (as a prime source of electrical power); a strong but lightweight structure to hold the satellite and its components together; a thermal control system to keep the components and its payload from becoming too cold or too hot; an electrical system for controlling components; a tracking, telemetry, command, and monitoring system so that people on the ground can actively know how the satellite platform and its payload are operating and send commands to maintain effective operations; a source of fuel and a thruster maneuvering system to aid in keeping a proper orbit; and a finely tuned pointing and orientation system, particularly to help position and point the satellite antennas or onboard sensing system for best performance. Finally there is a star, sun, and/or Earth sensing system to allow people on the ground to know exactly how the satellite is pointing on an X, Y, and Z axis or there is something like an RF alignment system to allow very precise pointing.

This platform or bus will also contain a payload (or in some cases multiple payloads) to perform a particular function such as communications, broadcasting, meteorological sensing, Earth sensing, Earth observation, space navigation, or perhaps a scientific experimental mission. Regardless of the payload and its mission, these elements will largely be common to the "bus" that delivers the payload to where it needs to go and to support the payload's operation 24 h a day, 7 days a week, until the mission is complete. When the mission is complete the payload is then employed to help with the final disposition of the satellite, such as bringing the satellite from low earth orbit back into the atmosphere where it will burn up or crash harmlessly into an ocean. If the satellite should happen to be in Geosynchronous orbit (i.e., a distance that is equivalent to almost one-tenth of the way to the Moon), then the usual maneuver to remove this type of spacecraft from GEO orbit is to raise it to a higher orbit where it will stay for many thousands if not millions of years.

Although there are many common elements to an applications satellite “bus” there still remain quite a large diversity of design elements that are described in the later chapters of this handbook. Some satellites need to be very accurately oriented to perform their mission and others much less so. Thus they range from very simple and low-pointing orientation systems to much more sophisticated ones. The simplest orienting system that is still in use is a gravity gradient system. With this type of platform design there are long booms that can be deployed to extend out into space away from the satellite. Once the booms are extended perhaps 5 m or more away from the satellite in different directions, the pull of gravity from the Earth can more or less orient the satellite toward the Earth. Other designs include satellites that spin around at speeds like 50–60 rpm, while their payloads inside spin in the opposite direction to achieve constant pointing toward the Earth with a stable orientation. These “spinners” were quite common in the early days of satellite communications. Today the most common “bus” is a three-axis-oriented platform that has one to three momentum or inertial wheels inside of the core of the satellite that spin at very high speeds, such as 4,000–5,000 rpm. This inertial wheel serves much like a spinning top to provide very accurate pointing orientation toward the Earth or wherever the platform needs to be oriented.

Just as there are options with the pointing and orienting system there are also options with regard to the thermal control system. Different types of reflective surfaces can be used to control solar heating. There are devices called heat pipes that can transfer heat from the interior of the satellite to the exterior. Despite the diversity of design options, most commercial manufacturers of satellites have a series of four or so basic platforms from which to build desired application satellites again, just as automobile manufactures have four or so chasses from which they build new automobiles. These various elements of the satellite platforms are discussed in great detail in the special section devoted to this subject.

In general, however, smaller and lower-cost satellites will have shorter lifetimes with smaller capacities if they are communications satellites. If they are sensing satellites they will have lower resolution or lesser sensing capabilities. In short, smaller applications satellites will typically have lesser capabilities than the larger spacecraft. Again microelectronics and other recent innovations might reverse this trend and allow smaller satellites to be much more capable. Further, there are often economies of scale that are achieved in the design of larger and longer-lived satellites and they also tend to be most cost efficient to launch on a “per kilogram to orbit” basis.

Organization and Effective Utilization of the Satellite Applications Handbook

This handbook is organized to be useful to a wide range of potential users from design engineers, faculty members and teachers, to reference librarians and students. It is organized into major sections. These sections are each self-contained and provide the history, demand for the various services, and unique technologies associated with the payloads – past, present, and future. Thus there are sections on (a) *Space Telecommunications* (this section includes the main categories of Fixed Satellite Services, Mobile Satellite Services, Broadcast Satellite Services, plus store and forward data services, and data relay); (b) *Satellite Precision Navigation and Timing* (this type of satellite application has now become the second-largest commercial satellite service in terms of market size); (c) *Space Remote Sensing* (this section not only covers remote sensing and Earth Observation, but it also addresses the *Global Information System* (*GIS*) and related software); and (d) *Space Systems for Geosynchronous and Low Orbit Meteorology.* This is the remaining key civilian practical use of outer space. It is different in that the provision of this service is largely by governmental agencies rather than commercial companies. The practical value of this service is more difficult to quantify and commercialize, but each and every year this type of satellite applications serves to save many, many lives and greatly minimizes property damage sometimes in the billions of dollars (US).

Conclusion

This handbook thus addresses the above-described four principal areas of commercial satellite applications and seeks to do so in considerable depth. It does not, however, specifically address classified military and defense-related satellite applications.

What is presented is specific and detailed information about all forms of telecommunications satellites, remote sensing and Earth Observations, satellite navigation, and meteorological satellites. Information is provided for the commercial use by defense organizations of applications satellites in a so-called dual-use mode.

This term applies to civilian or commercial satellites that are also used to meet certain largely “non-tactical” military applications. In this regard it is important to note that military usage of satellites is in many ways quite parallel to civilian space applications and often presages the development of commercial systems. This is to indicate, for instance, that the basic engineering and design characteristics of telecommunications and remote sensing satellites are often quite similar, although military systems may add special features such as radiation hardening, antijamming capabilities, and encryption capabilities.

The aforementioned four satellite applications are today the prime commercial and practical civilian uses of outer space. In future years new applications such as Solar Power Satellites and possibly other applications might be added, but for now these are the areas of satellite applications.

To be comprehensive the handbook also presents current and detailed information regarding global launching capability around the world and also addresses the design, manufacture, test, and deployment of the application satellite spacecraft platforms or “buses” that are launched to support these various types of commercial satellite services. As noted above, the platforms for these various applications satellites are quite similar even though the payloads may be quite different.

Finally the last parts of each section of the handbook address the key economic, regulatory, social business, and trade issues that are associated with applications satellites. Again, although the payloads that are contained on various types of applications satellites are different, the economics, trade, and regulatory aspects in these various commercial systems are often quite similar. This is to say that applications satellites of various types need to use radio frequencies (RF) to send information to and from earth and that accordingly there is a need for RF allocations agreed through the International Telecommunication Union processes for these various operations. There are national processes for the approval of specific frequency assignments to particular spacecraft in a way to prevent undue frequency interference. There are also a host of technical, economic, regulatory, trade, standards, and even social and religious issues that arise from the use of applications satellites since by their very nature these satellites are international in their operation. Thus commercial applications satellites and their operation, for instance, come under some degree of regulatory control by the World Trade Organization (WTO) with regard to how these services are distributed or sold and related international and national regulation and control. Orbital debris is an increasing threat to the safe operation of applications satellites. This issue is before the UN Committee on the Peaceful Uses of Outer Space and its Working Group on the Long Term Sustainability of Space Activities. These final sections thus address these types of issues and especially regulatory, trade, business, economic, and social issues.

There are several other elements of the reference handbook that should be particularly noted in terms of convenience of use. First of all, the various parts of this handbook are divided into major sections and chapters with highly descriptive titles. Secondly, each chapter contains a list of keywords so that if one is interested in “orbital debris,” “photo voltaic solar cells,” “lithium batteries,” “frequency allocations,” “precision timing,” or the “United Nations Committee on the Peaceful Uses of Outer Space (COPUOS)” these terms should be easily identified. In addition there is a glossary of terms that seeks to cover the entire handbook’s contents for easy reference. The organization of the handbook is typically structured to put information about any one subject in a concentrated location. This means that power systems are discussed together rather than in four different sections for each type of applications satellite. Finally the appendices are a key source of information about actual applications satellite systems, launch systems, and relevant information.

Cross-References

- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ Introduction to Satellite Navigation Systems
- ▶ Introduction to Space Systems for Meteorology
- ▶ Introduction and History of Space Remote Sensing

- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Space Telecommunications Services and Applications
- ▶ Satellite Communications Video Markets: Dynamics and Trends

References

J.N. Pelton, *Global Satellite Communications Policy: Intelsat, Politics and Functionalism* (Lomond Systems, Mt. Airy, 1974), pp. 47–50

Satellite Industry Association, The Tauri Group, 2014 State of the Industry Report (2015), http://www.sia.org/wp-content/uploads/2015/06/Mktg15-SSIR-2015-FINAL-Compressed.pdf

Satellite Communications Overview

Joseph N. Pelton

Contents

Abstract

In the 50 years that followed the first satellite launches of the late 1950s and early 1960s, the diversity of satellite services has expanded enormously. Today, there are direct broadcast radio and television services to the home and even to mobile receivers. There are mobile satellite services to airplanes, ships at sea, and even hand-held transceivers. There are so-called fixed satellite services to earth stations of various sizes down to so-called very small antenna terminals (VSATs), microterminals, and even ultra small aperture terminals that can be located on desktops. There are data relay satellites and business to business satellite systems. The age of the Internet and data networking has certainly served to add to the diversity of satellite services. Technology innovation has also led to the growth and development of satellite communications services. Lower cost launch arrangements and development of earth station technology and particularly application specific integrated circuits have been key to driving down the cost and size of ground antennas and transceivers. The development of three axis body stabilized spacecraft, better solar cells and batteries, and more effective on-board antenna systems and on-board switching among multi-beam antennas have also furthered the cause. Finally, the development of not only bigger and better satellites but the evolution of satellite systems design and network architecture

---

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_2

that allowed networks to be deployed in different types of orbits and network constellations has been part of this on-going evolution.

The latest iterations of satellite design have led to almost opposite extremes. On one hand there are large, sophisticated multi-ton satellites, known as high throughput satellites, deployed in traditional geosynchronous orbit locations. On the other hand, there are also small but capable satellites in low to medium earth orbit constellations. These new satellite networks are being designed with more and more mass-produced satellites – up to a thousand or more in a single system – to increase network capacity by means of deploying more and more satellites in lower orbit.

This chapter provides a general introduction to all of these changes and an overview to the entire field. Changes to satellite communication networks over the past half century have come not only in services and technology but also in regulation, standards, frequency allocations, economics, as well as the global reach and impact of satellites on the entire scope of human society.

Keywords

Broadcast satellite service (BSS) • Data relay satellites • Fixed satellite service (FSS) • Geosynchronous earth orbit (GEO) • International Telecommunication Union (ITU) • Intersatellite link (ISL) • Low earth orbit (LEO) • Mobile satellite service (MSS) • Satellite constellations • Store and forward satellite service

Introduction

The serious consideration of the provision of satellite communications from space dates from 1945 when the first technical descriptions were written with regard to launching a spacecraft into geosynchronous orbit and the design of space stations as extraterrestrial radio relays was specifically outlined. In the historical section that follows, however, it becomes clear that the idea or concept had been around many years, indeed centuries before. The 1945 article, however, described the possible delivery of telecommunications services from space and presented detailed calculations as to how this might efficiently be done from a special orbit known as the geosynchronous (or sometimes the geostationary) orbit (Clarke 1945). Today, this orbit is even sometimes called the Clarke orbit.

The new capability that allowed satellites to be launched that came from technology development in the USSR and the USA in the late 1950s and early 1960s expanded into the capability to launch a satellite into geosynchronous orbit – that came in 1963 – allowed the rapid evolution of satellite communications technology. Within a decade, a wide variety of telecommunications services from satellites in different types of orbits became possible. The International Telecommunication Union (ITU), the specialized agency of the United Nations that oversees the use of radio frequencies (RF) for practical and scientific purposes assumed responsibility for satellite radio frequencies. This began with a globally attended Extraordinary Administrative Radio Conference (EARC) in 1959. The ITU thus provided for the first time a formal process by which radio frequency (RF) spectrum could be allocated to support satellite communication (Pelton 1974).

Over time, the ITU defined a number of satellite communications services that might be offered via different satellites in different types of orbits. The ITU international processes also defined a system and a process whereby there could be technical coordination of such satellites to limit interference between and among them. The number and type of satellite communications services have grown and expanded over the years as is discussed in the following sections (ITU 2008).

There are today many types of technical designs for satellite communications, and these technologies are optimized to support a variety of services around the world. A wide range of commercial satellites now operate at the national, regional, and global level. These satellite systems support various types of data, telephone, television, radio, and various networking services around the world.

Overview of Commercial Satellite Services

The services defined by the International Telecommunication Union (ITU) include the following.

Fixed satellite services (FSS): FSS spacecraft support telecommunications services between antennas that are at fixed points. These fixed antennas can be used for reception only, for two-way communication (like a cable in the sky), or for communications within a network that can start with only a few nodes or can grow to a very large network indeed with thousands of interconnected nodes. The first of the commercial communications satellite services were these FSS systems. The Intelsat FSS system was the first to begin to provide international commercial telecommunications services in 1965. Also deployed in 1965 was the FSS system called Molniya, which provided telecommunications services for the Soviet Union, other Soviet Socialist Republics, and Cuba. An Initial Defense Satellite Communications Satellite system was also deployed to provide FSS services to support US defense-related telecommunications services. These initial FSS systems have now multiplied to support satellite telecommunications for over 200 countries and territories around the world (Pelton 2006, p. 30). Figure 1 shows a current generation broadband FSS satellite, the IP Star that operates in the Asian region of the world. Figure 2 depicts the Viasat 2 high throughput satellite. This represents the highest capacity satellite of the current generation of fixed satellite service (FSS) spacecraft with a capability of about 150 GB/s. This is more than ten times the throughput of a large communications satellite of just 5 years ago.

In addition to over 200 commercial communications satellites that supply fixed satellite services, there are now scores of military communications satellites that provide fixed satellite services in support of defense-related missions. Although the largest fleet of defense-related communications satellites are owned and operated by the US military, there are a number of strategic communications satellite systems owned by over a dozen countries around the world. In addition, commercial satellite systems leased capacity to military systems for so-called dual-use purposes to

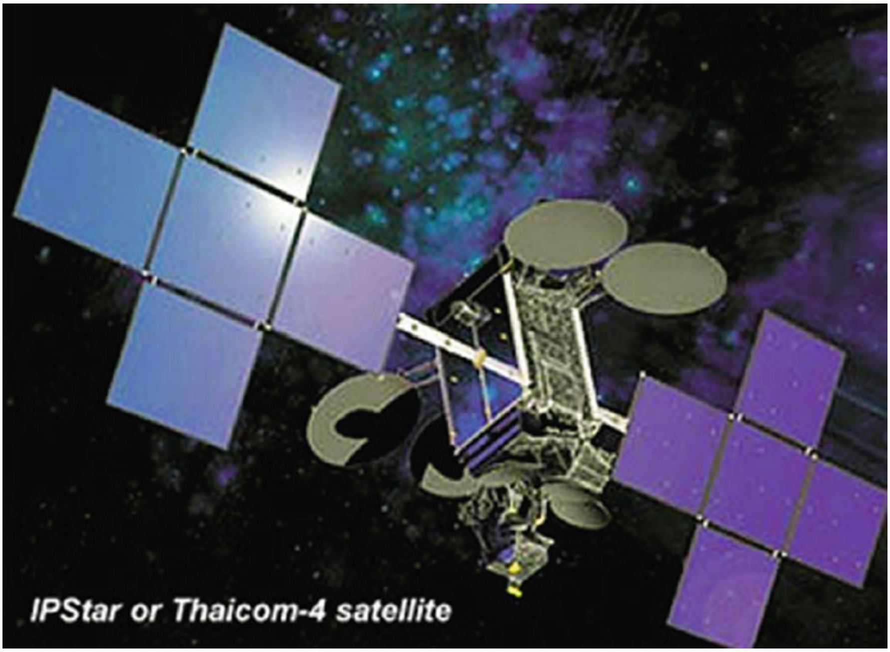


Fig. 1 The IPStar Satellite also known as Thaicom-4 (Graphic courtesy of IPStar)

supplement the capabilities of defense satellite communications systems. Figure 2 shows the WGSS military satellite designed to provide communications services.

Broadcast satellite services (BSS): BSS satellites use very high powered beams to deliver radio or television services directly to end users. In order for this service (also known informally as the direct broadcast satellite (DBS) service) to be economical and efficient, the receiver terminals must be small in size, low in cost, and easy to install and operate. Different RF bands are used for radio or direct audio broadcast services (DABS) in contrast to direct broadcast satellite (DBS) television services. The broadcast satellite service began later than the initial FSS offerings, but this industry has grown rapidly and is now by far the largest revenue generator in the satellite world by a wide margin (Pelton 2011).

The Nimiq BSS satellite (Fig. 3), operated by the Canadian Telesat organization, provides direct broadcast services to Canada and the USA.

Mobile satellite services (MSS): The MSS services provide telecommunications to end-user antennas that move rather than remain stationary. The MSS services today include telecommunications connectivity for maritime, aeronautical, and land-based users. The first MSS satellites were designed for maritime service. Next, these satellites were used to support both maritime and aeronautical services. The last type of mobile communications satellites that has evolved are those designed for land-based mobile services. This is the most demanding of the MSS services technically,

Fig. 2 The high throughput Viasat 2 depicted in orbit (Graphic courtesy of Viasat)

but in terms of market, this is also the most demanding. A variety of different designs in different orbits have evolved with the initial land mobile systems known as Iridium, Globalstar, and ICO experiencing severe financial and market difficulties with their initial service offerings. These organizations have been reorganized and the Iridium and Globalstar systems are deploying their second-generation systems while ICO is developing a new mobile satellite service for the US market (Figs. 4 and 5).

Today, there are a variety of MSS satellites deployed in a variety of different orbits. Some of the latest systems are those designed to work in conjunction with terrestrial cellular telephone services within urban areas. These hybrid systems that integrate mobile communication satellites with terrestrial cellular systems are called MSS with “ancillary terrestrial component (ATC)” in the USA. The equivalent service is called MSS with complementary ground component (CGC) in Europe. These hybrid mobile systems combine urban terrestrial cellular systems in a seamless manner to allow very high powered MSS satellites to cover the rest of a country or region. Unlike the initial constellations like Iridium and Globalstar, these new systems with a terrestrial component are deployed in geosynchronous orbit and are targeted to service to a single country like the USA or a single region like Europe (Pelton 2006, p. 31).

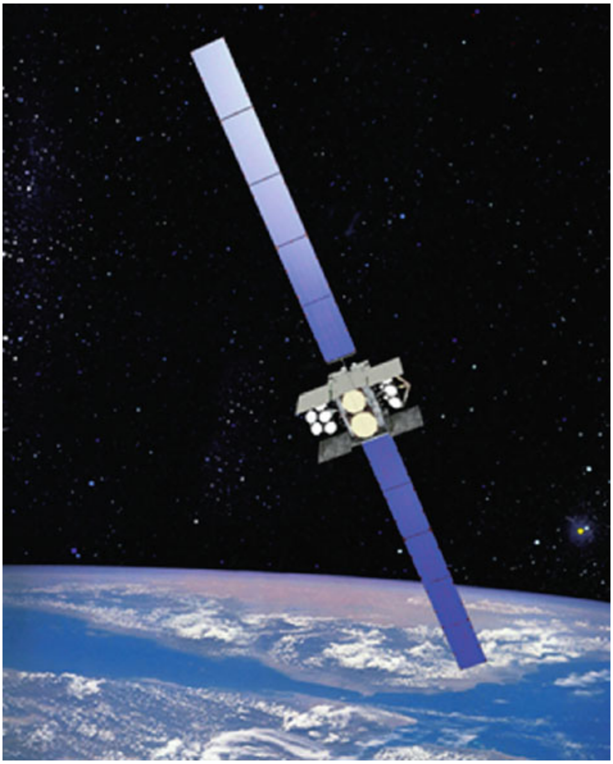


Fig. 3 The wideband global Satcom satellite (Graphic courtesy of the US Military)

These satellite services, FSS, BSS, and MSS, are the so-called big three of the commercial satellite services and represent a very significant part of the total worldwide market for the satellite industry. Nevertheless, there are other types of telecommunication satellite systems that can be, and indeed are, deployed. One additional system is the so-called store and forward type data relay satellite that can support messaging services to remote areas. The more satellites deployed in low earth orbit to support in this type of system, the more rapidly a message can be relayed from one part of the world to another. If there are enough satellites of this type, like in the Orbcomm system, you can have almost instant messaging. In some cases, the receiver can be configured to not only receive short messages but also to receive space navigation signals to support vehicular or ship navigation. One can also design a transceiver to send short data messages as well as to receive them, as has been done with several store and forward satellite systems. There are also data relay satellites that are typically in Geo orbit. These satellites are most typically used to relay data from a low earth orbit system back to a central process so that data can be continuously collected rather than stored for download at a subsequent time. The NASA third generation tracking and data relay satellite that is being deploy in the 2013–2016 time frame is shown in Fig. 6.

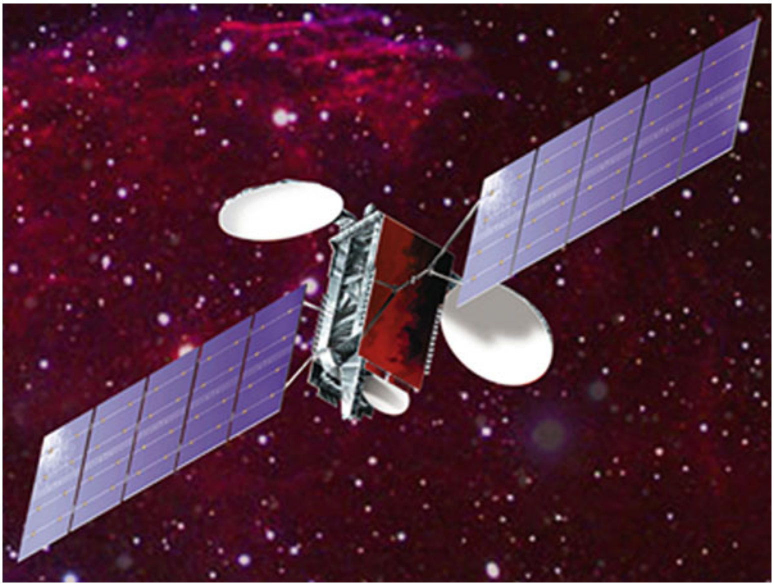


Fig. 4 A Canadian Nimiq direct broadcast satellite (Graphic courtesy of Telesat)

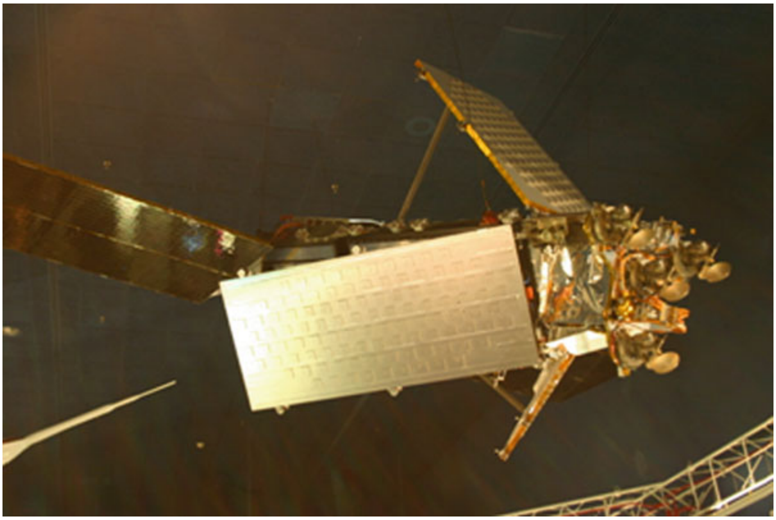


Fig. 5 A constellation of 66 iridium satellites provides global mobile services (Graphic courtesy of Iridium)

Some commercial satellite systems employ what are called cross-links (CLs) or intersatellite links (ISLs), or in ITU parlance intersatellite service (ISS) in order to operate. These can be used in low earth orbit or medium earth orbit to interconnect satellite constellations. ISLs were a part of the design of the low earth orbit Iridium

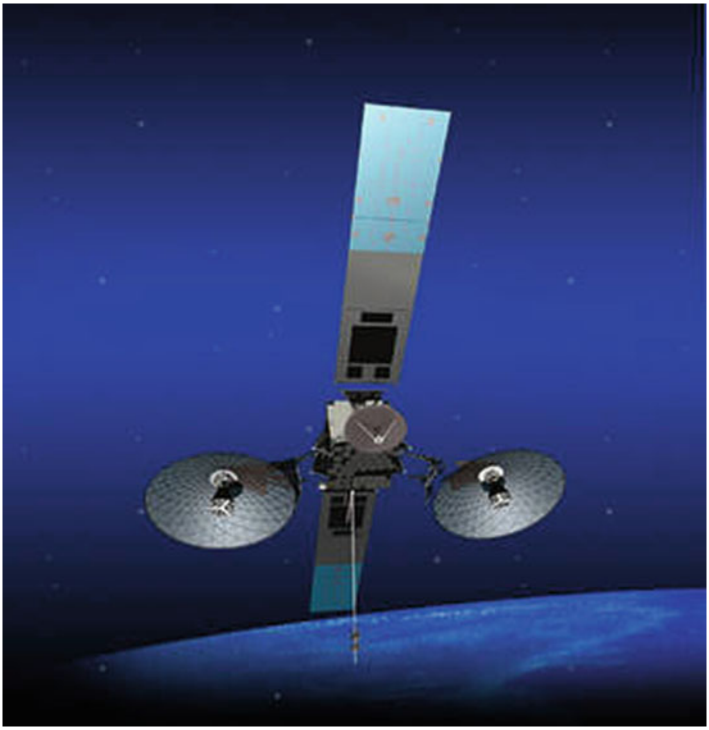


Fig. 6 The NASA third generation of tracking and data relay satellite

satellite network, and they have been used in some military communications satellites. ISLs could also be used to interconnect geosynchronous satellites on an interregional basis in order to avoid double hops when communicating halfway around the world. Today, satellite service connections providing global linkages and thus can often combine with a fiber-optic submarine cable to achieve rapid connectivity across the world. This is true in part since intersatellite links for regional interconnectivity still remain relatively rare. The more common use of ISLs is to interconnect satellites within a large-scale low earth orbit constellation where the satellites are typically hundreds of kilometers apart from one another rather than many thousands of kilometers apart such as the case when geosynchronous satellites are serving different regions of the world (Pelton 2006, p. 31).

Finally, there are satellites for military or defense-related communications. These satellites for military purposes are allocated different frequency spectrum than commercial satellites. These satellites resemble commercial satellites in many of their technical features, but they often have special features. Special capabilities can include radiation hardening, additional redundancy, and special encryption capabilities. Military communications are not operated on a commercial basis for the most part. There is a special chapter in this handbook that does describe such commercial defense-related communications satellite systems such as X-TAR as well as the "dual use" of commercial satellites for civilian requirements as well as defense-related applications.[1]

Other applications satellites are designed for different purposes other than telecommunications. Yet, these too must be able to relay information to various users on the ground. Thus, there are many types of satellite systems, other than commercial satellite systems, which are designed to support scientific communications, exact timing, remote sensing, earth observation, search and rescue, geodetic measurements, or various types of environmental services such as to monitor tsunamis, volcanoes, earthquakes, etc. These types of satellites are not addressed in this part of the handbook but are covered in later sections.

Conclusion

The chapters that follow in this section seek to provide a comprehensive and interdisciplinary overview of satellite communications services and applications, markets, economics, technology, operations and continuity of service, regulation, and future trends. Specific information on commercial satellite systems is provided in the appendices at the end of the handbook. Chapter "▶ Overview of the Spacecraft Bus" addresses the common technical elements found in essentially all types of applications satellites. Thus, this specific chapter addresses spacecraft power systems; thermal balancing and heat dissipation systems; orientation, pointing, and positioning systems; structural design elements; diagnostic systems; tracking, telemetry, and command systems; manufacturing and integration; and quality and reliability testing processes. Chapter "▶ Major Launch Systems Available Globally" also addresses how the various applications satellites are launched by different rocket systems from various launch sites around the world.

References

A. C. Clarke, Extraterrestrial radio relays. Wireless World (Oct 1945)

ITU, Radio regulations as published in 2008 (ITU, 2008). http://www.itu.int/publ/R-REG-RR/en

J. Oslund, Dual use challenge and response: military and commercial uses of space communications, in *Communications Satellites: Global Change Agents*, ed. by J.N. Pelton, R.J. Oslund, P. Marshall (Lawrence Erlbaum, Mahwah, 2004), pp. 175–194. Also see Satellite security and performance in an era of dual use. Int. J. Space Commun. http://www.spacejournal.ohio.edu/Issue6/pdf/pelton.pdf

J.N. Pelton, *Intelsat: Global Communications Satellite Policy* (Lomond Systems, Mt. Airy, 1974), pp. 40–47

J.N. Pelton, *The Basics of Satellite Telecommuications International* (Engineering Consortium, Chicago, 2006), p. 30

J.N. Pelton, *Satellite Communications* (Springer, New York, 2011)

---

[1] See Oslund (2004).

History of Satellite Communications

Joseph N. Pelton

Contents

Abstract

The history of satellite communications is a rich one that began centuries ago with the efforts to interpret the meaning of the "wandering planets" among the stars and to understand the structure of the cosmos. Early scientists such as Sir Isaac Newton and writers of speculative fiction both contemplated the idea that humans might one day launch artificial satellites into orbit for practical purposes. This chapter provides a brief overview of that rich international history up through the early days of global satellite operations. This history continues to provide a

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_35

narrative concerning the different types of satellites that have evolved to offer various kinds of services and the development of competitive satellite networks that are at the core of the communications satellite industry today. A brief history of military satellite systems and the "dual use" of commercial satellite systems to support defense-related communications needs is also addressed.

Keywords

Arabsat • CEPT (Council European Post and Telecommunications) • Clarke, Arthur C. • Cold War • Communications satellite corporation (Comsat) • Courier satellite • Department of defense (DOD) • Divestiture of AT&T • Domestic satellite leases • Early bird satellite • Eutelsat • FR-3 satellite • Galileo • Global information grid (GIG) • Global maritime distress safety system (GMDSS) • Goddard, Robert • Hale, Edward Everett • Initial defense satellite communications systems (IDSCS) • Inmarsat • Intelsat • International maritime organization (IMO) • International telecommunication union (ITU) • Kefauver, Estes • Kennedy, John Fitzgerald • Kerr, Robert S. • MARECS satellite • Marisat • Ministry of post and telecommunications • Molniya • Moon landing • Newton, Isaac • Pickering, William • Relay satellite • SCORE satellite • Shockley, William • Submarine cable systems • Syncom satellite • Telstar • Tsiolkovsky, Konstantin • TV-Sat • United Nations (UN) • van Allen, James • von Braun, Werner • Wells, H. G. • Wideband global satellite (WGS)

Introduction

The idea of satellite communications is a powerful one that has spawned a billion dollar industry on which the people of Planet Earth now depend every day. The idea that humans could actually launch a satellite into Earth orbit, however, was dependent on certain key knowledge about the Solar System that was lacking for many millennia. The concept of an artificial satellite revolving around our home planet was first and foremost dependent on the understanding that Earth itself is a planetary body that revolves around the Sun and that Earth and Moon are subject to universal laws of gravity. It further requires the understanding that the Moon, as a satellite, revolves around the Earth.

In short, before the orbital mechanics of the Solar System were understood and the concept of gravity clearly comprehended, the idea that one might launch an artificial moon or satellite into Earth orbit made no sense. But once one did grasp the basic physics involved, the idea that an artificial satellite could serve as a very high "artificial relay tower" for communications was a quite logical concept to follow. Clearly an artificial satellite circling the Earth might indeed be designed to receive radio waves or some form of signal transmitted up from the Earth out to space and return them to a desired distant location. How then did this historical thought process occur and who were the key players? This chapter not only outlines the history of satellite communications, but also indicates how the current structure of today's complex satellite markets is now evolving.

Early History of Satellite Communications

The starting point in this thought process began with the understanding that the Earth revolves around the Sun and that the Moon revolves around the Earth. This correct conception was confirmed by Galileo Galilei (1565–1642) in the sixteenth and seventeenth century when he was able to look through a telescope to observe Jupiter and note that four satellites were revolving around Jupiter. The limited magnification of his telescope prevented him observing that there were in fact many more artificial moons revolving around this giant planet and that there were other moons circling other planets. The discovery of the moons circling Jupiter provided sufficient physical data to draw a reasonable conclusion about the basic physics of the Solar System's orbital mechanics. Galileo's discoveries aided the thought process to posit that the Earth was also one of the "wanderers" or "planets" that revolved around the Sun. It also helped to confirm that the Earth had its own orbiting satellite, which we call the Moon.

Actually it was Galileo who first coined the term "satellite" that we use today. He applied to these distant moons the Latin word *satelles*. Galileo thought this word might appropriately be used to describe the "moons" of Jupiter. The Latin word was at the time used to describe an attendant or servant who was bound to obey the commands of his master. To Galileo the distant moons flying around Jupiter were bound to obey the commands of this mighty distant planet. Today we indeed have a large number of application satellites which do the bidding of their human designers. Many applications and scientific satellites now launched into Earth orbit carry out communications, navigation, remote sensing, or meteorology as well as various types of scientific discoveries. Galileo, however, did not understand at that time the concept of gravity and thus did not understand what force was used to command the moons of Jupiter to circle in their orbits, nor why the Moon should circle Earth. Indeed, because Galileo's observations ran counter to the dicta of the Catholic Church, it was quite a while until the workings of the Solar System became widely comprehended and understood in a correct scientific sense (Pelton and Madry 2010).

The next key historical step essential to the understanding of how an artificial satellite might be launched into Earth orbit and then provide services to people back on the ground came with the seventeenth-century discovery of gravity. Isaac Newton's discovery had many implications that impacted everything from astrophysics to zoology. He figured out how the mechanics of gravity worked within the Solar System. He did this through his own observations as well as by studying the writings of Galileo and Copernicus. His writings described how a very powerful cannon might shoot an object with enough velocity in order to allow the "launched object" to travel greater and greater distances. He then concluded that if the object could be shot with sufficient velocity, it would overcome the pull of Earth's gravity and would attain orbital speed and thus start circling the Earth. There is even a wonderful illustration from his writings in *Philosophiæ Naturalis Principia Mathematica* that show how this might be accomplished (Pelton 1981).

It is interesting that the next step in the thought process that led to the actual launch of applications satellites came not from the annals of science but from the imaginative literature of the nineteenth century. Writers such as Achille Eyraud (*Voyage to Venus*, 1863), Jules Venus (*From Earth to Moon*, 1865), Edward Everett Hale (*The Brick Moon*, 1869), and H.G. Wells (*The First Men in the Moon*, 1897) inspired popular and scientific thought about the possibility of space travel and the construction of rocket ships that could launch people and things into orbit or even beyond. It is the writings of Edward Everett Hale that today seems to be the most remarkable in its anticipation of today's application satellites. His book in 1869 anticipated the ability to launch a satellite into so-called polar orbit. In his book, he described how an "artificial moon" could be deployed as a practical device for communications, Earth observation, or navigation.

The Modern History of Satellite Communications

By the twentieth century, technology was evolving very rapidly. Konstantin Tsiolkovsky (1857–1935), in Russia at the very outset of the new century, gave careful and deliberate thought to the design of rockets that could carry people into outer space. Robert Goddard (1882–1945) began experiments in the USA to build viable rocket launchers only to be laughed at in a New York Times editorial as the "Moon Man." Goddard persevered and in 1926 proved that viable liquid-fueled launchers were indeed possible. During World War II, the German government assembled a team of scientists to develop rockets as weapons systems based, in part, on Goddard's earlier work. These led to the development in Germany of buzz-bombs, V-1, and then the V-2 rockets with ever-increasing range and accuracy. After the war, a part of the German rocket team was brought to the USA to work on this technology and the other part went to the Soviet Union to develop rocket systems there. From these two efforts came the launcher systems that became so prominent a feature of the so-called Cold War.

In 1945, a young man named Arthur C. Clarke wrote an article that brought into clear focus exactly what a communications satellite system might do, how it might be launched, and even presented in detail the reasons why such a space-based communications network should be place into geosynchronous orbit. Arthur C. Clarke, who spent World War II in the British Radar Establishment, first developed his ideas and sent a detailed letter to colleagues in June of 1945. Then in October 1945, he published his ideas and calculations in the journal *Wireless World*. At the time, this landmark article did not attract a great deal of attention. It was not the cover story of that edition, and he only received only a modest 15 pounds sterling compensation for his efforts. His 1945 article at the time was thus a largely unheralded event, even though his brief eight-page article contained the basic concepts on which a multibillion industry would be born and global television news reporting "live via satellite" would become commonplace only a few decades later (Clarke 1945).

Arthur C. Clarke, who died at age 90 in 2008, explained to colleagues before he died that he did not seek to patent the idea of a geosynchronous communications satellite. This was simply because he anticipated that the space stations he wrote about would be realized many years into the future. He believed that his "space stations" would require a crew to replace the radio tubes that would frequently burn out. In short, Sir Arthur Clarke, who was knighted by the British Government for his many predictions and farsighted writings, did not anticipate the transistor and the integrated circuit. These devices were to make possible not only the reliable solid-state technology that would enable reliable satellite technology but also would facilitate the development of high-speed electronic computers that could calculate the celestial mechanics associated with their accurate deployment into space.

In fact the invention of the transistor came only a few years later at Bell Labs, in December 1947.[1] This fundamental breakthrough by William Bradford Shockley, John Bardeen, and Walter Houses Brattain led to many innovations that ranged from the transistor radio to the modern electronic computer. The "transistor" and the integrated circuitry that followed have transformed the world in almost every conceivable way over the past half century, from the World Wide Web to the cell phone. Certainly the transistor transformed the concept of a communications satellite and the practical utilization of outer space from a far off dream to only a difficult technical challenge.

On October 7, 1957, the Space Age began with the launch by the Soviet Union of the world's first artificial satellite, Sputnik 1. In light of the "Cold War" that then existed between the USA and the Union of Soviet Socialist Republics (USSR), this launch, even though carried out in the context of the International Geophysical Year (IGY) for global scientific research, was broadly interpreted in a political context. Thus, there was an immediate perception that the USA was subject to a so-called missile gap. This led to immediate efforts by the USA to launch and orbit a satellite of its own. Another almost immediate response to the launch of Sputnik was for the US Congress to pass a new law in 1958 to create a new space organization known as the National Aeronautical and Space Administration (NASA).

In the months that followed, the Soviet Union (i.e., the shortened name of the USSR) continued to launch increasingly sophisticated and larger satellites, while the USA experienced a series of embarrassing launch failures. On February 1, 1958, however, the USA did manage to launch the Explorer 1 satellite into orbit. This satellite and the launch team, headed by Dr. William Pickering of the Jet Propulsion Lab, Dr. James Van Allen of the University of Iowa, and Werner Von Braun of NASA, confirmed the existence of the powerful belts of radiation that surround the Earth. The second Soviet satellite, Sputnik 2, had also sensed the presence of orbital radiation.

From the period from 1957 to the early 1960s, a number of satellites were launched by the Soviet Union and the USA – the only two countries with orbital launch capability at that time. The Soviet Union also demonstrated an early capability to launch heavier satellites and to orbit animals and then even people into orbit. On April 12, 1961, Vostok 1 was launched with Yuri Gagarin aboard to become the first person in space. The USA, with lesser launch capability, initially focused on

---

[1] History of the Transistor, http://www.inventors.about.com/library/weekly/aa061698.htm.

miniaturization so that it could launch more capable satellites with a smaller mass. The US presidential election of 1960 hinged in part on the issue of the “missile gap,” and President John F. Kennedy focused one of his first major speeches to Congress on the issue of outer space.

Kennedy’s speech to a Joint Session of Congress on May 25, 1961, is most memorable for his challenge to the USA to send people to the Moon and successfully return them by the end of the 1960s decade. This speech, known formally as the “Special Message to Congress on Urgent National Needs,” was the one which launched the NASA Moon mission known as Project Apollo. In that same speech, Kennedy also called for other space achievements. He called for funding for the Rover nuclear launch system and the rapid development of satellite communications technology and systems. He urged Congressional funding of $50 million (equivalent to perhaps $500 million in 2010) for “accelerating the use of space satellites for worldwide communications” (Special Message to the Congress on Urgent National Needs 1961). Clearly the Moon mission was what dominated the press coverage the next day, but President Kennedy also put great personal stress on the future potential of a global communications satellites network. In September 1961, some 4 months later, President Kennedy went to the General Assembly of the United Nations and called for the establishment of a single global satellite system that would: “. . . benefit all countries, promote world peace, and allow non-discriminating access for countries of the world.”[2] As a result of US urgings, the United Nations adopted resolution 1721 that included Section P, which stated “communications by means of satellite should be available to the millions of the world as soon as possible on a global and non-discriminatory basis” (United Nations General Assembly Resolution of Satellite Communications 1961).

The ongoing political processes led to the creation of the Communications Satellite Corporation (Comsat) in 1962 when the US Congress enacted the Communications Satellite Act of 1962. This led to the subsequent signing in Washington, DC, in August 1964 of the Initial International Agreement to create the International Telecommunications Satellite Consortium known as Intelsat (and its companion Operating Agreement). The creation of Intelsat was largely spearheaded by US initiatives and especially through representatives of the US State Department and of Comsat. This new Intelsat entity, which was initially organized as an international consortium, started with mainly Western countries as members (USA, Australia, Canada, Japan, and most of the Western European nations) and grew to include well over 100 member countries around the world (Alper and Pelton 1986). Seven years later after the 1964 launch of Intelsat, the Soviet Union, in response to its growing international membership, launched another entity known as the Intersputnik International Organization of Space Communications (or simply Intersputnik) with a membership of eight socialist countries, namely, the Soviet Union plus Bulgaria, Cuba, Czechoslovakia, East Germany, Hungary, Mongolia, Poland, and Romania.[3]

---

[2] Op cit. J. Logsdon et al. 1998, p. 42.

[3] Intersputnik International Organization for Space Communications, http://www.intersputnik.com/.

Fig. 1 Courier 1B satellite – world's first active repeater satellite (Photo courtesy of the US army signal corps)

While the national and global political processes were moving along, the related satellite technology was developing at an even swifter pace.

In December 1958, the US Signal Corps launched what might be characterized as the world's first "broadcasting satellite." This satellite, known as SCORE, simply repeated a brief message from President Eisenhower: "Peace on Earth, Goodwill Toward Men." It was launched just before Christmas on December 18, 1958, and its batteries were exhausted just before the end of the year. On August 12, 1960, the Echo I, a giant aluminized balloon, was launched to carry out meteorological experiments, but AT&T Bell Labs experimenters also tested the idea that such a satellite could serve as a passive reflector of radio signals as way to relay signals back to Earth, somewhat like bouncing shortwave radio transmissions off the ionosphere. These experiments were in a way successful by demonstrating that the signal throughput for a "passive communications satellite" would be too modest to serve as a commercially viable communications service.

It was not until October 1960 that the first active communications satellite, Courier 1B, was launched (see Fig. 1). This experimental spacecraft only supported the transmission of 16 teletype channels. Yet this satellite actively demonstrated that the relay of a signal to a satellite and then its retransmission of teletype messages back to Earth could be technically achieved. Its active transponders were powered by solar cells. From this landmark demonstration, quite rapid progress toward more and more capable communications satellites continued apace. Although today's space systems, a half century later, literally possess a billion times more capacity, the basic technical concept is in many ways the same.

By 1962, there was a surge in the technical sophistication of the design of active communications satellites. On July 10, 1962, the Telstar satellite, as designed by Dr. John R. Pierce and his team at Bell Labs, was launched into low Earth orbit (Fig. 2).

For the first time in human history, the Telstar satellite demonstrated how a live and real-time television signal could be relayed across an ocean. This was followed

Fig. 2 The AT&T designed Telstar satellite that first transmitted live television (Photo courtesy of Bell Labs)

by the launch of the Relay satellite on December 14, 1962. This satellite, as built by RCA in accord with NASA Goddard Spaceflight Center design, was similar to the Telstar and conducted similar transmission experiments. This NASA design specified an augmented power system that provided this satellite with a longer in-orbit life. Thus, Relay 1 remained in service through 1965. This satellite, in addition to conducting television transmission tests, was also designed to measure the impact of the Van Allen Belt radiation on the satellite communications subsystem (Fig. 3) (U.S. Congressional Hearings 1962).

The technical feasibility of satellite communications to support teletype, voice, and even television had been demonstrated by the end of 1962. The remaining key technical question was whether a communications satellite could be successfully launched into geosynchronous orbit (sometimes call the Clarke orbit in honor of Arthur C. Clarke) and operated reliably from this great distance – almost a tenth of the way to the Moon.

This question was answered in 1963 when the Hughes Aircraft Company designed and built the so-called Syncom satellites (for geoSYNchronous COMmunications satellites). The three satellites of this design were launched by NASA on Delta launch vehicles. The first launch was a failure, but the second, Syncom 2, was successfully launched on December 14, 1963, exactly 1 year after the launch of Relay. The Syncom 2 and subsequent Syncom 3, as engineered by Dr. Harold Rosen and his team at Hughes, demonstrated that reliable communications to geosynchronous orbit with a return link to Earth was indeed technically and operationally viable. For the 1964 Olympics in Japan, television signals were transmitted from Japan to the USA via Syncom 3 and the signal was transmitted from the USA to Europe via the Relay 1 satellite. The idea of global television relay of major sporting and world events "live via satellite" across the oceans thus date back to the early 1960s.

Fig. 3 The Relay 1 satellite designed by NASA and built by RCA (Photo courtesy of NASA)

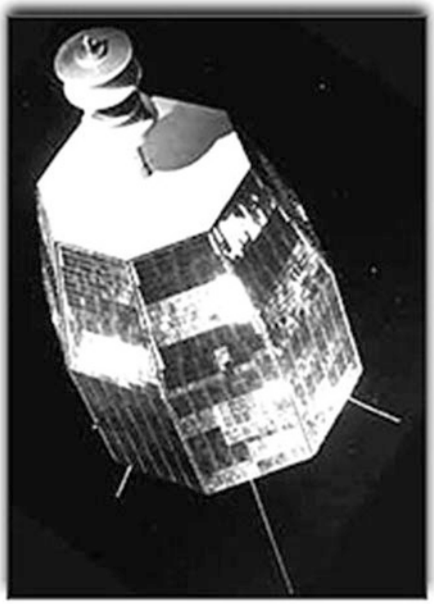
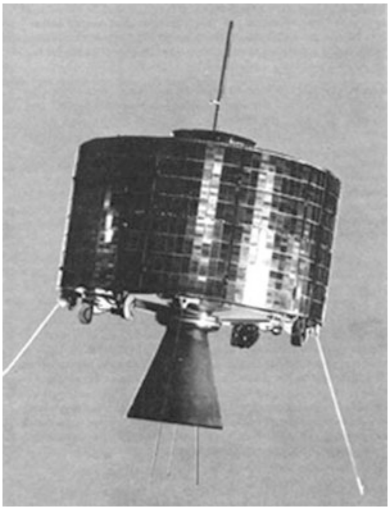


Fig. 4 Syncom experimental satellite that first demonstrated feasibility of operation from a geosynchronous orbit (Photo courtesy of NASA)



Although satellite telecommunications technology was moving swiftly ahead, the political and economic processes to establish a mechanism to provide satellite services to the world were subject to a number of key challenges (Fig. 4). The enactment of the Communications Satellite Act of 1962 in the USA constituted a protracted and very difficult political process. This conflict arose because the telecommunications industries wanted satellite communications services to be completely commercialized, and Sen. Robert S. Kerr of Oklahoma, who headed the powerful public works committee, led the fight in this direction. Sen. Estes Kefauver, who had been the Democratic candidate for Vice President on the ticket with Adlai Stevenson, argued that public expenditures through NASA had brought this new technology to the level of industrial feasibility, and he led the fight for a public agency for satellite communications. President Kennedy, on the other hand, was eager for a bold new space initiative and badly wanted to put the USA in a leadership role with regard to establishing global satellite communications. In addition, he needed to heed the advice of the powerful Senator Kefauver and his colleagues. In short, he wanted a compromise solution. The threatened filibuster in the Senate required skillful action. He relied on John A. Johnson, then general counsel at NASA, to draft a compromise bill that created the Communications Satellite Corporation (COMSAT) as a private corporation, but with half of the shares going to major telecommunications companies such as AT&T, IT&T, RCA, Western Union, and Western Union International and with the other half of the shares to be sold to the public on the New York Stock Exchange. Under this compromise bill, COMSAT was subject to instruction by the US Government on matters of national policy by the State Department, the Federal Communications Commission (FCC), and the White House Office of Telecommunications Policy (OTP). In addition technical advice was also to be provided by NASA. This compromise bill managed to pass and break the deadlock between the Kefauver and Kerr factions and an ongoing filibuster avoided.

The next challenge was the international negotiations to create a framework for international satellite communications services. The original thought within the US State Department was that COMSAT would undertake to establish a series of bilateral agreements with countries that wished to establish satellite links. When the US delegations arrived in Europe to discuss international arrangements for satellite communications, they were confronted with a unified European position via the Committee on European Post and Telecommunications (CEPT). These European telecommunications officials insisted that a new international agency would need to be formed for this purpose.

Two years of tough international negotiations ensued. The final outcome was the signing of the Interim Intelsat Agreements, as described earlier. There were two agreements. One document was an Intergovernmental Agreement signed by nation states on behalf of their governments. The other document was called the Special Agreement, and this was signed by "participating" telecommunications organizations as variously constituted within the countries that signed the Intergovernmental Agreement. The purpose of having this second agreement was to allow private or semi-private companies such as Telespazio of Italy, KDD of Japan, the Overseas Telecommunications Corporation (Australia), the Canadian Overseas Telecommunications Corporation (COTC), or Comsat of the USA to participate directly in the organization as a partial owner as well as governmental agencies such as post and telecommunications agencies.

These documents were deemed to be "interim" in nature because European countries maintained that the USA possessed an unfair advantage due to their technical lead in launch vehicle technology. They successfully maintained that after 5 years of experience, a new set of permanent arrangements should be negotiated to reflect newly gained operational capabilities and strengthened new space technologies that were evolving around the world. These countries (especially European nations) believed, and correctly so, that under the permanent arrangements there would be the opportunity for a more thoroughgoing internationalization of the Intelsat management. In particular, COMSAT was designated as the Manager of the Intelsat system in the Interim Intelsat Arrangements largely due to US official insistence. The supporters of the "interim arrangements" believed that after experience had been gained, the "US dominant technical and operational role" could and would decrease as space capabilities spread around the world.

The two Interim Intelsat Agreements were signed by 15 countries in Washington, DC, on August 20, 1964. Some countries had the ability for these signatures to take immediate effect, and others had to obtain ratification by their national legislatures. In the months and years that followed more and more countries joined this initial satellite communications consortium. An official report on experience gained was completed in 1969, and this led to 2 years of negotiations that concluded with the so-called Final Agreements in 1971. It was not until 1973 that enough signatures were gained for these new agreements to enter into force. By this time, membership had swelled to well over 80 countries.

The Communications Satellite Corporation, COMSAT, in its role as Manager for the Intelsat Consortium, sought to bring the Intelsat system into operation as soon as possible once it was established in 1962. As a result of the successful deployment of the Syncom 2 and 3 satellites into geosynchronous orbit, COMSAT signed a contract with the Hughes Aircraft Company to build a somewhat larger version of Syncom. This satellite with a larger bank of solar cells, a "squinted beam" antenna that provided increased pointing ability back toward the Earth – and thus higher gain – was the result. This satellite once deployed was able to provide the equivalent of 240 voice circuits (or complete two-way voice channels) or alternatively one low-quality black-and-white television channel.

This satellite that was officially known as the Intelsat I (F-1) was actually more popularly known in the world press as "Early Bird" (the "F" stood for flight model and indicated it was successfully launched into orbit). This satellite, which was launched in April 1965 just 8 months after the formation of the Intelsat Consortium, surprised the world by achieving practical commercial satellite communications in a remarkably short period of time.

Exciting satellite video experiments were conducted. For example, Dr. Michael Debakey conducted open heart surgery in Switzerland, and the procedures were watched live via satellite by heart surgeons in Houston, Texas, who were able to ask questions in real time. Coverage of the LeMans auto race in France were beamed to the USA, and Heads of State were able to exchange greetings (Fig. 5).

Early Bird, when it was launched in 1965, was in many ways an experimental satellite. But the Intelsat II series was able to provide multidestination service and video, audio, and data service to ships at sea in support of the US Gemini space

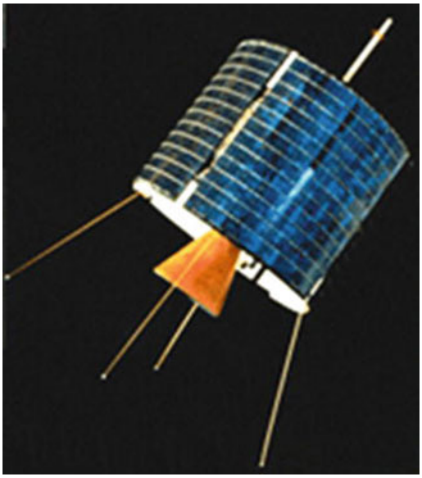


Fig. 5 The early bird satellite, world's first commercial communications satellite (Graphic courtesy of the Comsat Legacy Project)

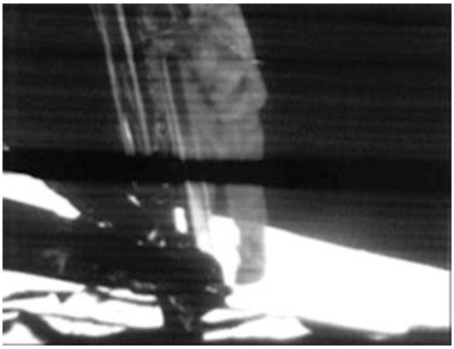


Fig. 6 Low-resolution TV picture of Neil Armstrong's first step on the Moon (Image courtesy of NASA)

program. Next came the Intelsat III series with more than five times the capacity of the early Intelsat satellites. Each of these satellites, with much higher gain antennas, could provide 1,200 two-way telephone circuits plus two-color television channels.

This Intelsat III series was the first to complete a fully global network. It was in June 1969 that a network of these satellites were deployed and fully configured so that they could send voice and television channels not only across the Atlantic and the Pacific Oceans, but even across the Indian Ocean. It was this global Intelsat III network that allowed a worldwide audience of over 500 million people to see the Moon Landing of Apollo 11 on the Lunar surface and the first space walk (Fig. 6).

In the following years, from 1969 to the early 1970s, Permanent Management Agreements were negotiated for the newly named International Telecommunications Satellite Organization (Intelsat). During the period, an international management was established that assumed these responsibilities from the USA-based COMSAT over a transitional period. The communications satellites increased in size, power, lifetime, and performance and migrated from analog to digital communication services. The size of Intelsat Earth Stations decreased in size, and very small aperture terminals (VSATs) supporting what was called the Intelsat Business Service became commonplace. Intelsat provided not only tens of thousands of international voice circuits, data networks, and eventually hundreds of television channels, but it also increasingly leased spare capacity to support domestic voice, data, and television distribution in scores of countries around the world. In the years that followed, the Intelsat system grew in satellite capability and performance, especially as the multiplexing systems migrated from analog to digital.

The number of members and participating countries and territories also expanded and international and domestic traffic surged, despite competition from the greatly expanded channel capacity of fiber-optic submarine cables that were laid across the oceans.

This significant expansion of the satellite communications business and its perception as a viable and attractive business led to efforts to restructure the global system within which these services were provided. There was an increased move, particularly within the countries of the OECD (Organization Economic and Cooperative Development), toward competitive telecommunications service. In the early 1980s, during the Reagan Administration in the USA, there were several competitive filings, particularly with systems known as Orion and PanAmSat, that proposed that they be authorized to compete directly with Intelsat. It took a number of years for this whole issue to be resolved in terms of the restructuring of Intelsat to become a commercial entity and for ground rules to be agreed as to how competitive systems might be authorized by governments and allowed to operate within their borders. This "macro-change" in the structure of telecommunications toward "liberalization" and "competition" led to the conversion of Intelsat and other publicly structured "public monopolies" to become competitive private telecommunications industries by the later part of the 1980s and the early 1990s and for the new commercial satellite systems to be licensed and to deploy their competitive networks.

These efforts to create competitive systems at the international level were, to a certain extent, stimulated by efforts to create at the regional-level communications systems, such as Eutelsat for the European region and Arabsat for the areas of the Middle East and Northern Africa. Further, the decision to create a separate international organization for maritime and aeronautical satellite communications, called INMARSAT, also served to create momentum toward creating satellite systems separate from Intelsat. This thought process argued that systems designed and optimized for specific markets could be better optimized than a single system configured to meet all possible requirements. This thought process, namely, of more competition for telecommunications services began to arise in the 1980s. The argument arose among economists that instead of just having national monopoly communications networks, competition would help improve services and reduce consumer costs. Up until the 1980s, in most countries telecommunications organizations were regulated by so-called rate base oversight. This meant in practical terms that the more they invested in new "allowable communications infrastructure" the more "return" they could realize. In the 1980s, many countries switched over to the idea of competitive telecommunication systems. The thought was that this competitive system might be more responsive and cost less than simply having a monopoly provider.

These various national and regional decisions of the 1980s led to new levels of competition for telecommunications services. A 1983 "judgment" for Federal Judge Harold Greene that settled a suit against AT&T undertaken by the US Justice Department led to the breakup of the AT&T monopoly in the USA as of January 1984. The development of competitive systems in Europe, Japan, and elsewhere followed in the next few years. This context of moving monopoly telecommunications systems to competitive networks clearly set the context for the authorization of competitive international satellite systems.

This process strongly contributed to the ultimate decision among the Intelsat Assembly of Parties to transform Intelsat from a public international organization with national governments acting as members and investors to entirely new arrangements. After the restructure of Intelsat, it became a privately held corporation as of July 18, 2001. Henceforth, Intelsat became just another corporation offering satellite telecommunications services around the world. This led to the "privatized" Intelsat spinning off part of its assets to a new European-based company known as "New Skies." The same sentiments and logic ended with both Inmarsat and Eutelsat also being "privatized" so that these organizations were entirely owned by private equities and no longer owned by national governments. This meant they were no longer international organizations operating under international treaty arrangements, but simply commercial competitors operating in a commercial marketplace along with other competitors with no special rights and privileges.

In the case of both Intelsat and Inmarsat small international organizations were set up to address special concerns about the "public good" and "public services," these organizations as public international organizations had previously performed. These involved services such as the right of access to international communications via satellite for public safety, for other special public needs and also to assist developing countries to achieve equitable access to telecommunication satellite services. In the case of Intelsat, a small part of the former INTELSAT Organization was not privatized on July 18, 2001. This modestly sized residual group remained an international organization, under the acronym ITSO (standing for "International Telecommunications Satellite Organization"). The role of this organization, with 150 members from around the world, and which had previously owned Intelsat when it was "spun off," was officially defined to be as follows:

- Act as the supervisory authority of the new Intelsat Ltd.
- Ensure the performance of Core Principles for the provision of international public telecommunications services, with high reliability and quality.
- Promote international public telecommunications services to meet the needs of the information and communication society.[4]

---

[4] The International Telecommunication Satellite Organization, http://67.228.58.85/dyn4000/itso/tpl1_itso.cfm?location.

This action was taken to assuage those members of Intelsat that had been reluctant to "privatize" Intelsat. As a practical matter the ITSO has limited ability to affect the commercial policies of Intelsat Ltd.

The same parallel was followed in the case of privatizing Inmarsat that in fact occurred before the Intelsat restructuring. In this case, the Inmarsat derivative body became known as the International Mobile Satellite Organization (IMSO). This intergovernmental body was likewise established to ensure that Inmarsat continues to meet its public service obligations, including obligations relating to the Global Maritime Distress Safety System (GMDSS). IMSO is also designated an observer to attend meetings of the UN Specialized Agency, the International Maritime Organization (IMO). In April 1998, the Inmarsat Convention was amended to create this IMSO in its current form when Inmarsat Ltd. was restructured as a privatized organization. In addition to its public maritime safety role, the IMSO seeks to guarantee that services are provided by Inmarsat Ltd. free from any discrimination and in a peaceful way to all persons living or working in locations that are inaccessible to conventional, terrestrial means of communication. IMSO also ensures that the principles of fair competition are observed.[5]

Over time this commercialization or privatization process led to a series of mergers and acquisitions. New Skies was purchased by the group known as SES Global, based in Luxembourg, as part of its global network of satellite assets. Perhaps most ironically of all, the privatized Intelsat eventually ended up purchasing the satellite organization known as PanAmSat. This company, that is, PanAmSat, had originally been its biggest international competitor and driver of the competitive process that led to Intelsat being restructured as a private competitive satellite provider.

Apart from the move to create a competitive global industrial structure for the provision of worldwide fixed satellite services starting in the late 1980s and 1990s, the overall history of satellite communications was punctuated by several key events. These will be addressed in appropriately titled sections ahead, and these events relate to: (1) the creation of separate satellite systems for maritime and mobile satellite services, (2) the evolution of regional and domestic satellite systems, (3) the development of satellite systems to support infrastructure for defense- and military-related services, and (4) the development and launch of direct broadcast satellite systems, known in the parlance of the International Telecommunication Union (ITU) as the Broadcast Satellite Service (BSS).

Separate Systems for Maritime and Mobile Satellite Services

The success of satellites for international communications, and especially the new ability to provide broadband service across the oceans, quickly led to interest in using satellite technology for maritime communications (Fig. 7).

---

[5] The Creation of the International Maritime Satellite Organization (IMSO) in its current form, www.imo.org/conventions/contents.asp?doc_id=674&topic_id=257.

Fig. 7 The Marisat satellite, the world's first dedicated maritime communications satellite (Picture courtesy of Comsat Legacy Project)

As noted above, the Intelsat II satellite series was sponsored by the National Aeronautics and Space Administration (NASA) essentially to support communications between launch vehicles ascending from Cape Canaveral and to establish links with tracking ships in the Atlantic Ocean in support of the Gemini Manned Space Program. But this was an inefficient system because of the small antenna size of the ship-mounted reflectors. Intelsat satellites, at least in earlier years, were designed to communicate between and among larger-scale fixed location Earth Stations. The desire by the US Navy to communicate more effectively with its globally deployed fleet led to the planning of a dedicated maritime satellite known as Marisat. This satellite as pictured above was manufactured by the Hughes Aircraft Company (now the Boeing Corporation). It was deployed in 1976 on the basis that half of the Marisat system capacity would dedicated to meeting US Navy fleet communications needs and the other half to commercial maritime communications needs. COMSAT General, a subsidiary of Comsat created to enter into other satellite ventures, served as the operator of the system and marketed the additional maritime capacity to other entities desiring maritime services.

The success of fixed satellite services stimulated worldwide interest in "the next step" in terms of maritime satellite services. The European Space Agency had developed and launched some experimental fixed satellites known as the European Communications Satellites (ECS). It followed this program with the European Communications Satellites (ECS) for Maritime Service, known as (MARECS). These satellites were launched and performed a number of successful tests and demonstrations. Within Intelsat, there was active discussion as to whether it should expand its services into the maritime mobile communications satellite services area. Under Article XIV of the definitive Intelsat Arrangements, Intelsat was granted by its member states the right to enter into what were characterized as "specialized services" that included maritime, aeronautical, or land mobile services. There was, however, a quite important caveat added to this authorization. This required an active determination by the Intelsat Assembly of Parties (the plenipotentiary body of all member states) that the provision of such specialized services would not involve an economic penalty to member states that were not users of these additional services. In the mid to late 1970s, Intelsat was in the process of acquiring its fifth generation of satellites known as the Intelsat V. These satellites, with 12,000 voice circuit capacity and two television channels, were procured from the Ford Aerospace Corporation (now Space Systems/Loral) with an initial purchase of six satellites. After considerable discussion and a vote within the Intelsat Board of Governors and finally a favorable decision by the Intelsat Assembly of Parties, it was decided to acquire three additional Intelsat V satellites with a maritime communications package aboard. These Intelsat V-MCS satellites were also launched successfully into orbit.

The launch of the Marisat, MARECS, and ISV-MCS capacity into geosynchronous orbit created a great deal of maritime mobile satellite capacity, but the institutional and organization situation was certainly quite unclear. The Intelsat organization had a strong interest in extending its worldwide sway over maritime and possibly aeronautical and other mobile services. However, the institutional situation was complicated by several factors. One key factor was that the Soviet Union was not a member of Intelsat. The USSR was not a major user of international telecommunications services, but it was certainly a key player when it came to maritime communications. Another key factor was that a number of countries tended to see Intelsat, even after the negotiation of the permanent management arrangements and the creation of an internationally staffed Executive Organ headed by a Secretary General (and later a Director General), to be largely controlled and staffed by the USA.

These "complications" led to preliminary discussions held in the UK about the possibility of creating a new international organization to provide maritime communications satellite services. The Safety Committee of the International Maritime Organization (IMO), a United Nations specialized international organization, also endorsed the idea of creating a separate international organization dedicated to maritime satellite communications and safety. This led to a series of three Conferences in 1975 and the formal signing of an international agreement to establish INMARSAT in 1976 that actually went into force in 1979.

In this new organization European, Soviet Union, and other major shipping interests would have a predominant voting share in contrast to Intelsat where the USA had predominant control. In short, the thought was to create a new organization that would be structured around maritime shipping interests, maritime fleets, and maritime safety and not international telecommunications usage as reflected in the Intelsat Organization.

In its structure and its enabling agreement, however, this new international organization closely resembled the Intelsat organization. Like Intelsat, INMARSAT had an Assembly of its members, a Board and specialized advisory committees of its Board. Its membership, however, was focused on the major maritime powers and most notably differed from Intelsat by including the Soviet Union in its membership and ownership. It was also headquartered in London rather than Washington, DC, and thus it was largely seen as a “European” entity rather than an “American” institution. Unlike Intelsat, that had to gradually build up its space infrastructure that took from 1965 to 1969 to establish a global network, Inmarsat “inherited” a global space network that included the Marisat satellites, the MARECS satellites, and the Intelsat V-MCS that together covered most of the world’s oceans except for a thin strip of the Southern Pacific off the coast of Chile.

Once the Inmarsat Agreement was in place, the issue of mobile satellite communications to support aeronautical services began to arise in the 1980s. Inmarsat not only began to plan its own dedicated satellites to support future maritime needs but also began working toward space segment capability that could not only meet maritime needs but also provide communications to aircraft with appropriately designed antennas that could be easily mounted on airplanes.

In 1994, the INMARSAT Assembly proceeded to amend its charter to create the International Mobile Satellite Organization (IMSO) that would address the needs of maritime and aeronautical satellite communications for safety. In 1998, it also spun off the new “privatized” commercial Inmarsat Ltd. that would own and operate the Inmarsat satellite system and would provide commercial mobile satellite services.

During the early 1990s, the Motorola Corporation initiated a project to provide a global satellite network to provide land mobile satellite services on a global basis. From the outset, officials from Motorola met with officials of both Intelsat and Inmarsat to explore whether either organization would like to engage in a joint venture to deploy such a global land mobile satellite system. Intelsat and Inmarsat both declined, but in the case of Inmarsat it decided to not only privatize and commercialize its maritime and aeronautical satellite services under the name Inmarsat Inc. but also to create an entirely new commercial organization first known as the International Circular Orbit (ICO) Ltd.

This new ICO corporation was capitalized by an Initial Public Offering (IPO) and had the objective of deploying a global land mobile communications service. The Secretary General of INMARSAT before it was privatized and restructured as a private corporation, Mr. Olof Lundberg, decided to resign as head of INMARSAT and to become the head of this new ICO commercial entity. At the time, this new commercial land mobile satellite business seemed to be quite promising and was projected to grow more rapidly than the maritime or aeronautical satellite communications business. The prospects seemed so bright that the billion dollar IPO offering for ICO was oversubscribed. Motorola proceeded on its own and formed a new global satellite consortium of commercial partners known as Iridium to launch a low Earth orbit land mobile satellite constellation. Further, the aerospace corporation Space Systems/Loral also formed yet another consortium with telecom partners around the world to launch the Globalstar low Earth orbit satellite consortium for land mobile satellite services. On the order of US $15–$18 billion were put at risk to create these new satellite systems for land mobile satellite services at the time terrestrial-based cell phone systems were expanding and maturing at a rapid pace. Unfortunately all three of the dedicated land mobile satellite systems, Iridium, ICO, and Globalstar, failed financially and commercially, and the ICO system, as originally conceived, was never launched even though several satellites for this system were designed and manufactured.

In 1995, 1996, and 1997, the Iridium consortium launched, deployed, and began operating a network of 66 satellites Leo constellation that was also supported by a number of operational space satellites. The projected satellite cell phone traffic in the millions of circuits did not materialize. Marketing and licensing agreement problems and technical performance problems associated with the inability to call reliably from buildings and automobiles plus the large size of the user handheld transceivers led the Iridium system going into bankruptcy in 1998 after less than 2 years of operation. An estimated \$7–\$8 billion of losses were incurred by Motorola and its many international partners around the world. The Globalstar satellite that deployed some 48 satellites in a low Earth orbit constellation plus spares was deployed just shortly after the Iridium system. It also declared bankruptcy in 1998. Finally the ICO system that had purchased medium Earth orbit satellites from Boeing also declared bankruptcy after the failure of Iridium and Globalstar without ever deploying its network. The staggering losses of \$7–\$8 billion for the Iridium system, the \$6–\$7 billion losses for Globalstar, and the over \$2 billion losses for ICO had a dramatic impact on the overall satellite communications industry in the late 1990s and early 2000s.

Ironically, the Inmarsat Ltd. commercial venture continued to expand its maritime and aeronautical satellite services successfully and proved to be quite financially viable. Particularly with the deployment of its latest quite powerful and large aperture Inmarsat 4 satellites, Inmarsat Inc. has managed to expand into the land mobile satellite services market in most recent years. Inmarsat has continued to deploy larger and more capable satellites from geosynchronous orbit to support all of these services. Today, New Iridium (which is the name of the commercial entity that took over the assets of the original bankrupt Iridium consortium) and a reorganized and restructured Globalstar have both recovered from the catastrophic failures of the late 1990s and are providing global services from low Earth constellations. Further, on a regional and global basis, there now are two geosynchronous-based networks. These are the Inmarsat system, already discussed, as well as the geosynchronous-based Thuraya system that serves not only the Middle East but parts of Europe, North Africa, and Asia. These satellite networks both offer broadband land mobile services to a large number of customers. The Inmarsat Ltd. system supports maritime and aeronautical communications as well as land mobile. Iridium has ordered a second generation of satellites that will also allow higher powered and broader band services, and Globalstar has also ordered new satellites to upgrade their capabilities as well.

The latest innovation in land mobile satellite services comes from nationally based services for the USA, which the Federal Communications Commission (FCC) has designated as mobile communications satellites with ancillary terrestrial component (ATC). (Note this mobile satellite service is known as Complementary Ground Component (CGC) in Europe.) This concept involves the active marriage of terrestrial cellular service (i.e., land mobile services using terrestrial towers to cover

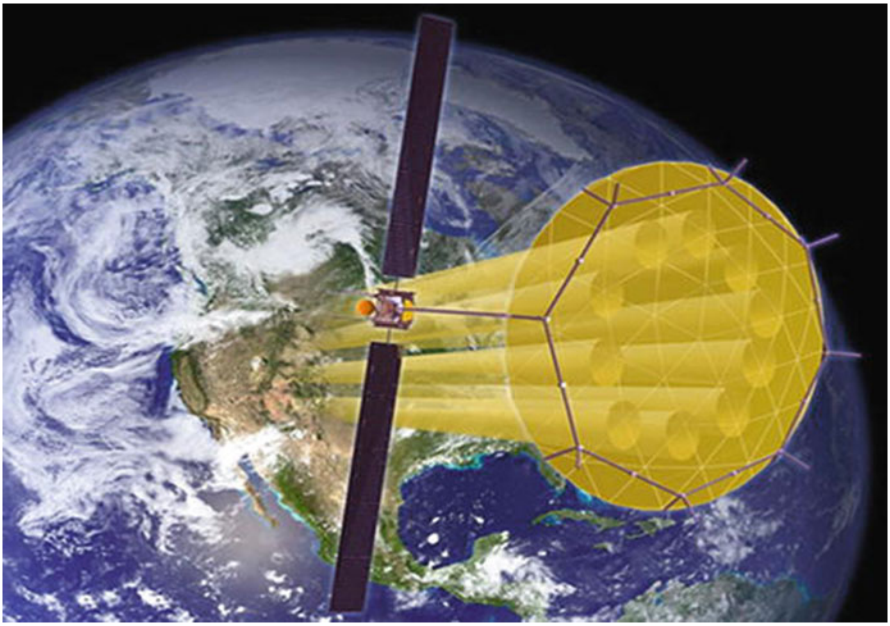


Fig. 8 The Light Squared land mobile satellite with ancillary terrestrial component with its huge deployable antenna system (Note: This system was formerly known as SkyTerra and before that MSV) (Photo courtesy of Light Squared)

the largest urban areas), which is then linked together with extremely high-powered, multiple-beam satellites. As with all new systems, there are key issues to be solved. In the case of Light Squared, the key issue that has arisen is how to avoid undue interference between its ground networks and GPS satellite signals. Indeed, these new types of communications satellites currently deploy the world's largest communications satellite antenna systems. These antennas effectively cover with relatively high power all of the rural parts of the USA. There are two of these systems now deployed and they are known, respectively, as Light Squared (formerly SkyTerra and prior to that MSV) and TerreStar. These satellites with their extremely large antennas with a total area almost equivalent to a soccer field can generate very powerful beams to support the service demands of mobile consumers anywhere outside the coverage of the terrestrial cell towers in urban areas. The Light Squared satellite with its huge multibeam antenna is shown in Fig. 8.

Evolution of Regional and Domestic Satellite Systems

At the time the original Intelsat Agreements were formed, the USA was essentially the only source of launch services since the Soviet Union chose not to participate in the consortium and launched its own network known as the Molniya satellite system.

There was considerable feeling in Europe and especially in France that the Intelsat arrangements were too much under US control. Specifically, they maintained that other satellite networks should be allowed. When the definitive arrangements were negotiated between 1969 and 1971, one of the more contentious issues was over the possibility of separate regional satellite systems and if this should occur what type of coordination process would appropriately be employed.

After months of negotiations, with a block of countries largely composed of the USA and developing countries on one side and European nations largely on the other, a deadlock of opinion occurred. One of the prime barriers to agreement was Article XIV that sought to address what services Intelsat might provide in the future and the coordination processes that would be employed in the event of other satellite systems. One provision that was generally conceded to be appropriate was that there would need to be technical coordination between Intelsat and other communications satellite systems owned or operated by Intelsat members. The provisions of Article XIV(d), however, required that any separate communications satellite system would also be subject to "economic coordination" and that the members participating in another such system would need to demonstrate that there would be no economic harm to Intelsat.

The creation of Inmarsat in the late 1970s set the stage for serious consideration of what other satellite systems might be deployed. This was particularly relevant to the European region, because the European Space Agency (ESA) and the French Space Agency (known as CNES) had seriously begun the development of the Ariane, a European launch vehicle. Earlier efforts to create a four-stage launcher within what was called the European Launcher Development Organization (ELDO), where different countries developed different stages of the launcher, had failed. This new effort under a unified management and a consolidated technical design and fabrication capability proved to be successful. Thus in 1977, the agreements under which a European Telecommunications Satellite Organization (Eutelsat) could be established were signed. This organization, just like Intelsat, was initially created as an intergovernmental organization (IGO) to develop and operate a satellite-based telecommunications infrastructure for Europe. One of the Deputy Director Generals of Intelsat, Mr. Andrea Caruso, formerly of Telespazio of Italy, left Intelsat to head up this organization, and the Eutelsat Agreements were not surprisingly much akin to Intelsat in their nature, and many of the European members of Intelsat also were members of Eutelsat. (Also just as the case with Intelsat and Inmarsat, this organization was later "privatized" and is now a private enterprise with private equity ownership.)

The question of Article XIV coordination, under the Intelsat Agreements, immediately arose with regard to Eutelsat. Documents were presented to Intelsat by the European organization describing the technical characteristics of the proposed Eutelsat satellites and indicating how and why these satellites would not pose harmful technical interference to Intelsat satellites. The more challenging issue was that of Article XIV(d) regarding economic coordination. These documents showed the traffic currently carried on Intelsat and indicated that there was very little traffic between and among European countries on the global system. This economic coordination document indicated that only very minor potential streams of traffic such as between Norway and Turkey would be involved and that these potential streams would be much less than 1 % of the traffic carried by Intelsat at that time, and the bulk of intra-European traffic for which Eutelsat was designed would very likely never be economically viable streams for the Intelsat system.

This proved to be a very contentious issue for Intelsat and its Board of Governors. It was perhaps very much the case because Intelsat officials knew that this was not only a test case, but that the decision would set precedents for other regional systems that could and indeed would follow. Finally with carefully worded language, the Board of Governors and then the Intelsat Assembly of Parties agreed that if a number of restrictions were observed, the Eutelsat satellite system would not constitute technical nor economic harm to Intelsat.

Eutelsat was thus clearly on its way to deploy its regional system without any significant legal or technical constrictions to its operations. It proceeded to launch its first satellite in 1983 on an Ariane launch vehicle. This was some 18 years after the launch of the first Intelsat satellite that was manufactured entirely in the USA and launched on an American launcher. Eutelsat thus demonstrated that European industry was now able to launch its own regional satellite, manufactured entirely in Europe. This Eutelsat spacecraft was launched from the Ariane equatorial-sited launch facility in French Guyana in South America. As was to be expected, a number of different proposals for separate satellite systems ensued that followed the European precedent. The next of these regional systems was called Arabsat. This regional system was designed to cover the Arab world, within the Middle East, and also covered the Arab states of Northern Africa.

The most dramatic shift to the world of satellite communications came in 1983 from filings for new satellite systems proposed to the FCC in the USA. The RCA Corporation plus two new start-up firms filed applications to create commercial satellite systems to provide international services directly in competition with Intelsat. When the first of these applications were officially filed, starting with Orion, and then PanAmSat, RCA, and others, the Intelsat Board of Governors was meeting in Sydney, Australia. At this meeting, there was general dismay that such a direct attack on the single global system had come so unexpectedly. Part of the dismays originated from the fact that Intelsat was largely created through the efforts started by the John F. Kennedy Presidency in the USA.

The leadership for the Orion system came from top Congressional aides who were in close contact with the Reagan White House Office of Telecommunications Policy (OTP). The Reagan White House and the leader of OTP Tom “Clay” Whitehead were advocates of the breakup of the American Telephone and Telegraph (ATT) monopoly and its divesture to create competition that they believed competition, whether domestically or internationally, would fuel innovation and drive down the cost of service. In short, the Reagan administration favored a pro-competitive policy for services both within the USA and abroad and provided a favorable attitude toward international satellite telecommunications competition.

On the other hand, the PanAmSat initiative came from television broadcasters from Mexico and South America who had found the cost of television broadcasting under Intelsat tariffs and especially under the ultimate price charged by Intelsat Signatories around the world to be exceptionally high. Other broadcasters such as CNN in America had also particularly encouraged the creation of competitive satellite systems, again because they found the television broadcasting and distribution tariffs of Intelsat and its signatories to be quite expensive.

Since USA was the largest member of Intelsat and since the US Government had played a key role in the formation of Intelsat, government officials knew they had to proceed cautiously. Thus, the process whereby the applications for competitive systems was considered within the US Government and discussed within the Intelsat Intersystem Coordination processes took a number of years. Lobbyist organizations and politically savvy law firms entered the fray as the applications for competitive satellite systems took center stage within Intelsat during the period 1983 through the end of the decade.

The following report from the conservative think tank, the Cato Institute, that describes events at this time and provides its views about the evolution of the thought processes that moved forward toward “competition” and “liberalization” summarizes these events and the political context that evolved from the late 1970s through the course of the 1980s.

In the 1970s, however, US telecommunications policy began to take a path that brought cold-war-era concerns about world leadership and a single global system into conflict with domestic trends favoring competition and diversity. An increasingly pro-competitive US government [sic: i.e., the Reagan Administration from 1981 to 1988] deregulated satellite communications for domestic traffic. Later, the United States allowed its domestic satellites to carry trans-border traffic on an ancillary basis. The latent policy conflict came to a head in 1983 when the US government received applications from RCA, Orion, PanAmSat, and others to launch and operate private satellite systems that would carry international traffic in direct competition with Intelsat (Mueller 1991).

The regional systems such as Arabsat, Asia Sat, along with Eutelsat that proceeded them, resulted in successful technical and economic coordination with Intelsat through the normal Board of Governors and Assembly of Parties processes. The issue of competitive international satellite systems raised a wide range of more fundamental issues. This ultimately led to proposals to restructure and “privatize” Intelsat, Inmarsat, and Eutelsat and move from state ownership to private ownership. Inmarsat was the first to make this transition, but within 2 years Eutelsat and Intelsat also soon followed. The restructure of Intelsat proved the most complicated with not only Intelsat moving from an intergovernmental organization (IGO) but also led to the spin-off of a number of Intelsat satellites to create a new European-based operator called “New Skies.” This action was designed to create a more competitive market structure more quickly. The claim by competitive systems was that the official status of Intelsat as an IGO gave it an unfair competitive advantage in the market place.

Satellite systems for international or regional satellite communications services were not the only types of systems to evolve over time. Intelsat had created tariffs and commercial arrangements for a number of countries to lease spare capacity on the Intelsat system for domestic purposes. Some of this "domestic traffic" – such as traffic between the US Mainland and Hawaii and Alaska, or between Denmark and Greenland, or between France and its Overseas Departments such as French Polynesia and Martinique, of course – seemed very much like international traffic. But beginning in the mid-1970s, Intelsat began to lease spare capacity to countries such as Algeria, the Sudan, Nigeria, Saudi Arabia, Malaysia, the Philippines, China, India, Brazil, Argentina, Colombia. Over time up to nearly 100 countries or territories leased capacity from Intelsat for domestic telecommunications and television services. This was "spare capacity" on Intelsat satellites that had been launched to restore service outages that might occur, but was not currently needed to provide international satellite services. These lease arrangements allowed countries to establish national long-distance telecommunications networks as well as national television broadcast networks. Not all of this leased capacity was by developing or industrializing countries. Some developed countries such as Australia, Germany, and the UK leased capacity to provide national television distribution or to otherwise supplement their terrestrial communications networks.

It was only a matter of time before the countries with the largest market needs proceeded to deploy their own dedicated satellite communications systems. The Soviet Union had indeed deployed their Molniya satellite network back in 1965. Canada was the first of the Western nations to deploy a domestic satellite network, but then a host of countries, both developed and newly industrializing, followed suit. France, Germany, the USA, Australia, Japan, Indonesia, Argentina, Brazil, China, India, Malaysia, Nigeria, Pakistan, Spain, Sweden, Taiwan, and Turkey among others have at least one if not several dedicated national satellite systems for telecommunications services. Many of these national systems are for business services to interconnect large-scale VSAT (very small aperture terminal) networks, others are for radio and television services, and some are for both.

The combination of international, regional, and national satellite systems today results in a quite large number of satellites in orbit. Virtually all of these types of satellites supporting fixed and broadcasting television and radio satellite services are in geosynchronous orbit and nearly 300 of such "Geo satellites" for communications services are in orbit today. The latest development for Geo satellites is the new generation of high-throughput satellites that have enormous throughput capabilities of up to 140 Gbps. These systems include KA-SAT (December 2010), Yahsat Y1A (April 2011), ViaSat-1 (October 2011), Yahsat Y1B (April 2012), EchoStar XVII (July 2012), HYLAS 2 (July 2012), Astra 2E (July 2013), O3b Constellation (2014), Inmarsat Global Xpress Constellation (2015), HNS Jupiter (2015), ViaSat 2 (2016), and Intelsat Epic (2017).

Communications Satellite Constellations as a New Option

The other significant new stage in communications satellite services involves low Earth and median Earth orbit constellations for a wide range of services. First this was new systems for land mobile satellite services in the late 1970s and 1980s. But most recently these new constellations are for fixed satellite services and particularly for Internet networking and for service to countries in the equatorial regions. The operation of these satellite constellations requires careful technical coordination of these satellites to lessen the problem of inter-satellite interference between low earth orbit constellations and GEO networks. These systems also create new challenges in order to avoid collision with orbital debris. This means that they seek to maintain the orbital positions of the satellites in their constellations with some precision, as assigned through a process established through the United Nations specialized agency, the International Telecommunication Union.

The use of low earth and medium earth constellations primarily for mobile satellite service appears on the verge of significant change. The O3b satellite network was deployed in medium earth orbit in 2013 to provide a service optimized for Internet networking. Even more radical change is slated to come if the proposed One Web constellation of some 700 small satellites and the Space X constellation of up to 4000 small satellites were to be deployed to provide Internet-based service in underserved portions of the world. These new type networks would radically change 50 years of satellite service by deploying a huge number of satellites in low earth orbit to support a new type of fixed satellite service (FSS) to support Internet networking primarily in the equatorial regions of the planet.

For the first two decades of satellite communication, services through about 1985 communications satellites were primarily deployed for international and regional satellite service, but this changed to also provide satellite services for domestic television and radio distribution and even telephone and data services. Today over 100 countries around the world either lease satellite capacity to meet domestic telecommunications needs or have established one or more separate satellite systems to meet their telecommunications needs. In some cases countries have separate systems for domestic needs but are still leasing capacity to meet additional needs that they might have, such as France which uses international FSS satellites to reach several of their overseas departments in the Caribbean and Pacific regions.

Table 1 provides a summary of countries that are either leasing capacity from Intelsat or other international or regional systems or now have separate networks to meet domestic needs. In total, the number of countries that today rely on satellite networks for long-distance overseas communications or for domestic links constitutes over half of the countries and territories in the world.

The regional and domestic satellite systems now deployed are designed to provide a wide range of services that include telephone, data, VSAT networking, virtual private networks to support corporate tele-working, television distribution of

Table 1 Domestic satellite systems and domestic satellite leases around the world (Countries with one or more separate systems are designated with SS)

1. Afghanistan	36. Guinea	71. Paraguay
2. Algeria	37. Guyana	72. Peru
3. Angola	38. Hong Kong (China)	73. Philippines (SS)
4. Argentina (SS)	39. India (SS)	74. Poland
5. Australia (SS)	40. Indonesia (SS)	75. Portugal
6. Austria	41. Iran (SS)	76. Qatar
7. Barbados	42. Iraq	77. Romania
8. Belgium	43. Israel (SS)	78. Russia (SS)
9. Bolivia	44. Italy (SS)	79. Saudi Arabia
10. Bosnia-Herzegovina	45. Ivory Coast	80. Senegal
11. Brazil (SS)	46. Japan (SS)	81. Solomon islands
12. Bulgaria	47. Kenya	82. South Africa, Rep. of
13. Cameroon	48. Korea, Rep. of (SS)	83. Spain (SS)
14. Canada (SS)	49. Kuwait	84. Sri Lanka
15. Central Africa Rep.	50. Libya	85. Sudan
16. Chad	51. Madagascar	86. Surinam
17. Chile	52. Malaysia (SS)	87. Sweden (SS)
18. China (SS)	53. Mali	88. Switzerland
19. Colombia (SS)	54. Martinique (France)	89. Taiwan (SS)
20. Congo, Democratic Rep.	55. Mauritius	90. Tanzania
21. Congo, Rep. of	56. Mexico (SS)	91. Thailand (SS)
22. Costa Rica	57. Mongolia	92. Trinidad and Tobago
23. Croatia	58. Mozambique	93. Turkey (SS)
24. Cuba	59. Myanmar	94. Tuvalu
25. Czech Republic	60. Namibia	95. Uganda
26. Denmark	61. Nepal	96. Ukraine
27. Egypt	62. Netherlands	97. UK (SS)
28. Equatorial Guinea	63. New Zealand	98. USA (SS)
29. France (SS) + depts.	64. Nicaragua	99. Vatican State
30. Gabon	65. Niger	100. Venezuela
31. Georgia	66. Nigeria (SS)	101. Vietnam (SS)
32. Germany	67. Norway	102. Zambia
33. Ghana	68. Oman	103. Zimbabwe
34. Greece	69. Pakistan (SS)	
35. Greenland (Denmark)	70. Papua New Guinea	

Pelton (2005) and from Intelsat

programs to support cable television systems, and direct satellite broadcasting for both radio and television services.

In the satellite communications field, however, new history is constantly being made. The latest trend is the rapid evolution of high-throughput satellites (HTS). These satellites with the latest multibeam antenna technology and the latest encoding systems can now provide data rates that are ten to even a hundred times those of past generations of communications satellites. In addition, the new so-called megaLEO constellations of small satellites optimized for Internet services also promise a time of major change in the satellite industry over the next few years.

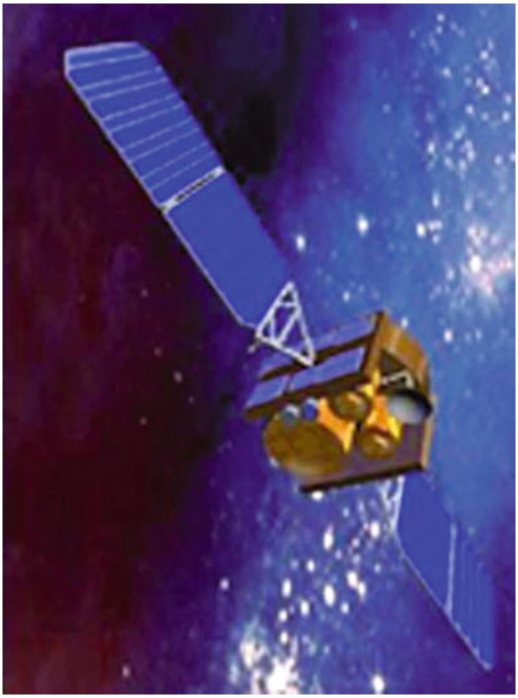
Satellite Systems to Support Defense- and Military-Related Services

The launch of the Intelsat system beginning in 1965 was not the only system deployed in that year. The Molniya satellite system, launched by the Soviet Union, was the world's first domestic satellite system. The US Department of Defense also launched the Initial Defense Satellite Communications System (IDSCS). This was a series of low Earth orbit (Leo) satellites in a random orbit constellation that allowed for more or less global communications, although there were some periodic service interruptions due to gaps in the satellite coverage. This initial limited capacity system led to a wide range of military communications satellites being launched, not only by the US military but by a number of other defense forces around the world in the years and decades that followed. Most of these systems are classified and most of them are deployed in geosynchronous orbit, but there are also some defense-related communications satellites in low and medium Earth orbit and even in super-synchronous orbit. In addition to the US defense-related systems, there are a wide range of other satellites for defense purposes that are now deployed and in service on behalf of Russia, the UK, Spain, France, and China. In addition, the Japanese Government has now authorized such systems to be deployed to support Japanese defense as well. The most complete defense-related satellite networks are those deployed by the USA. These satellites are collectively known as the Military Satellite Communications Program (MILSATCOM), and their various functions can be summarized as follows.

The MILSATCOM architecture has three major elements: (a) There is one type of satellite system for mobile tactical support services that operates in the ultrahigh frequency (UHF) band and has limited throughput capacity. (b) Another type is for long-haul protected communications. This type is represented by the three different generations of the Defense Satellite Communications System (DSCS), sometimes known as Discus. (c) Thirdly there is the type of satellite for Wideband Defense Communications Services, known as the Military Strategic, Tactical, and Relay (or MILSTAR) satellites (Fig. 9).

The latest version of this type of architecture is the Wideband Global Satellite (WGS) network. This was once known as the Wideband Gapfiller Satellite. These satellites today represent the most capable and most rapid throughput system in the US military communications satellite network. There were also plans until 2009 for a so-called Tranformational Satellite System (T-SAT) that would provide worldwide connectivity to the Global Information Grid (GIG). The GIG is the name for the entire global network of all forms of ground, air, and space communications systems for the US military. The T-Sat system is now on hold. Figure 10 provides an integrated view of all of these US defense-related space communications systems,

Fig. 9 The Phase III, or third-generation, DSCS satellite (Graphic courtesy of the US department of defense)



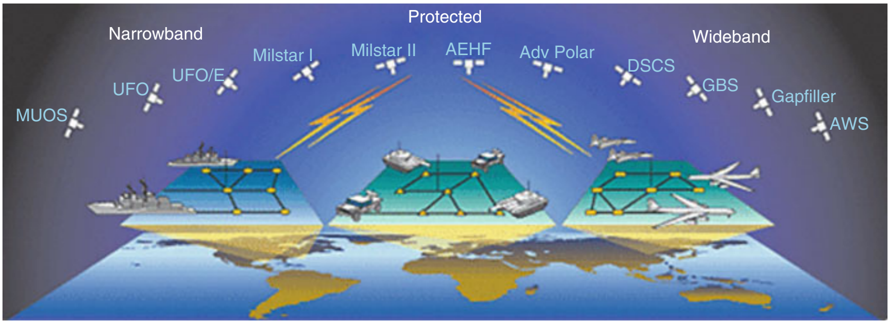


Fig. 10 US military communications satellite systems (Credit to US Department of Defense)

starting with the "narrowband" systems and ranging to the "protected" and then the "wideband systems." Except for the Advanced Polar Satellite and the planned low Earth orbit constellation Mobile User Operator System (MUOS), all these satellites are in geosynchronous orbit.

Despite all of this considerable global communications satellite capability, there still are gaps in the information and communications networks of the US satellite defense-related systems. The reason for these gaps is, in part, because military conflicts occur in different and sometimes unexpected parts of the world. The solution that has been found by the USA and other defense forces around the world has been to rely on commercial satellite systems. This so-called dual use of commercial satellite facilities has been to adapt commercial systems to military, defense, or emergency rescue use as special needs and demands arise.

For many decades now, going back to the 1970s, commercial communications to support defense-related services have been leased from satellite operators around the world. These "dual use" services support routine communications or even television or radio broadcasts to overseas personnel. In many of the applications, special capabilities such as jamming and "rad hard" protection (i.e., capability to survive severe radiation) are not required. For these and other reasons, commercial services can often meet demand and do so at lesser expense. Specific examples of such "dual use" satellites by the US Department of Defense (DoD) are the distribution of entertainment to overseas troops or to support e-mail and video messages to families. Such services often require a good deal of bandwidth but not special security.

Other "dual use" applications such as reliance on commercial mobile satellite services in Afghanistan and Iraq, however, often represent more strategic defense communications. The Enhanced Mobile Satellite Services (EMSS) project provides the DoD with a secure, global, handheld communications capability. DoD now employs New Iridium, Globalstar, and the Inmarsat 4 commercial mobile satellite systems, among other systems, to provide key "in-the-field" mobile communication service. These commercial mobile satellite systems thus provide connectivity to the Defense Information Systems Network (DISN).

More detailed information about the dual use of commercial satellite systems around the world and the specifics of how these arrangements work is provided later in chapter "► An Examination of the Governmental Use of Military and Commercial Satellite Communications." This discussion also provides detailed information about how other countries have found new and innovative ways to meet their military satellite communications needs in conjunction with commercial suppliers.

Direct Broadcast Satellite Systems

The Intelsat organization evolved a system to sell television services over extended periods of time rather than just on a short-term basis. Although Intelsat from the outset sold telephone and data circuits on a full-time basis, it began by selling television services on a minute-by-minute basis with a 10-min minimum. This was because television commanded a great deal of the satellite capacity and indeed the entire satellite for Intelsat I and II. During the Intelsat IV and Intelsat V era, however, Intelsat evolved a new charging principal of selling spare capacity for domestic systems.

This type of sales began with Algeria in the mid-1970s, and in the 1980s Intelsat decided to sell full-time transponders to support television service. This full-time television service lease began with Australian broadcaster Kerry Packer and quickly expanded to other types of full-time television leases around the world. This led to the idea of forming a company that could lease a full-time transponder and then sell spot capacity to broadcasters. The first company to do this was named Brightstar, a joint venture of Western Union in the USA and Visnews in the UK (Visnews was later acquired by the Reuters news agency). In time, other companies such as Wold International, Bonneville, AP_TV, and Globecast (the merged entity that combined the holdings of Wold and Bonneville) expanded this type of satellite television business greatly. These services provided the "distribution" of television news, sports, and entertainment among terrestrial television systems worldwide. In short, television programs and news material was delivered to cable television networks, to terrestrial microwave distribution systems, or over-the-air terrestrial broadcast systems for distribution to consumers. A number of scientists, engineers, and businessmen around the world and especially in North America and Europe began to ask why not send the satellite signal directly to the home and bypass the terrestrial networks. As domestic satellite systems were deployed in Canada, the USA, Europe, and Asia, many consumers, particularly those in rural and remote areas without access to cable television or over-the-air broadcast, began buying backyard television receive only (TVRO) dishes to receive satellite TV. In many cases, they also bought "descramblers" and began to watch so-called premium channels such as HBO and Cinemax. After a few years, there were literally millions of these backyard dishes and it became obvious that so-called direct broadcast satellite television services would be a viable business.

The Importance of Broadcast Satellite Services as the Largest Market

Indeed, today the direct broadcast satellite service represents the largest commercial satellite market as measured in the size of its global revenues. Because of this history, whereby the initial service evolved from Fixed Satellite Service (FSS) satellites providing service to backyard dishes and then new types of more powerful direct broadcast satellites designed for direct-to-the-home services (DTH) which evolved later, there is a confusion as to where one service ends and another begins. Today DTH services covers both types of services, namely, backyard dishes that can obtain programming from so-called FSS satellites that operate in several downlink frequency bands and Broadcast Satellite Services (BSS) spacecraft that operate in another downlink band.

BSS is the formal terminology used by the International Telecommunication Union (ITU) for this direct-to-the-consumer television service. This BSS offering is considered by the ITU to be a separate type of offering. Thus, for BSS offerings higher transmit powers are authorized, different frequencies are allocated for the downlinks, and the user terminals are typically much smaller and compact and accordingly cost less money than the backyard dishes. BSS terminals are typically 30 cm to 1 m in size, while backyard dishes can be 3 m to even 7 m in size.

The "true" BSS direct broadcast service is one in which many television channels are uplinked to a broadcast satellite in geosynchronous orbit and then downlinked at very high satellite broadcast power via a highly concentrated beam to a country or region so that the signal can be directly received by quite small home- or office-mounted satellite antennas. Under the ITU allocation of frequency bands for this type of service, there are different spectra used in different parts of the world. These high-power downlink transmission bands are as follows: The spectrum 11.7–12.2 GHz is allocated in what is known as ITU Region 1 (this region includes Europe, the Middle East, Africa, and parts of Russia). The downlink spectrum of 12.2–12.7 GHz is allocated for ITU Region 2 (this region includes all of the Americas). Finally the spectrum band of 10.7–12.75 GHz is allocated for downlinking BSS services in Region 3, which includes Asia and Australasia.[6]

Many countries, especially developing countries and those without satellite communications capacity, thought that the initial process by which frequencies were allocated for FSS services in Special and Extraordinary sessions of the World Administrative Radio Conferences that were held in 1959 and 1963 were biased in favor of the most advanced countries. Thus, when sessions were held to allocate frequencies for this new Broadcast Satellite Service in the 1970s, a number of countries supported the idea that allocations of frequencies for this service should also include allotments for countries that did not yet have satellite capabilities against their future needs.

Thus, in the 1977 international BSS Plan each country (at least in Regions 1 and 3) was allocated specific frequencies at specific orbital locations for domestic service. In this ITU negotiation process, a number of BSS channels were assigned for specific countries regardless of their current ability to launch and deploy BSS satellites. This was not the result for Region 2, however, because the USA in particular contended that this was too arbitrary of a process and that some allocations would likely never be used. Thus, in Region 2 allotments ended up being made on the basis of actual need with the opportunity for new entries being accommodated as new systems have arisen.

Not surprisingly, the much more static plans for Regions 1 and 3 have needed to be amended as a result of changes that resulted from the shift from analog to digital technology, the shift of national systems to regional coverage, and many other changes (including accommodating new countries that have emerged in Eastern Europe, Africa, and Asia). Satellite systems that provide this type of service today include, among others, BSkyB system in the UK and Europe, Europesat and Eutelsat Hotbird Satellites in various parts of Europe, plus the latest Astra satellites by SES. In the USA, these systems include Dish, DirecTV, and SES Americom services; Anik and other DBS systems in Canada; Sky Perfect JSAT and NHK in Japan; and Insat, Koreasat, Asiasat, and Chinasat satellites in Asia. (See the section on broadcast satellite markets for more complete details. Also see the Section on ITU allocations for more complete details of how this process actually transpires.)

In all there are over 20 broadcast satellite systems around the world transmitting many thousands of television channels to individual subscribers. Virtually all of

---

[6] ITU Radio Regulations, Article 1, Definition of Radio Service, Section 1.25, http://www.ictregulationtoolkit.org/en”practiceNote.aspx?id=2824.

these are now digital television channels and an ever-increasing number of these are high-definition television channels.

Today there is increasing clarity about the Broadcast Satellite Service (BSS) or Direct Broadcast Satellite systems in terms of what frequencies they use around the world and the digital transmission standards they use. These broadcast systems virtually all use efficient digital compression standards in order to send more television channels through available transponder capacity. These systems use the standards developed by the so-called Motion Picture Expert Group or MPEG standards. These are known as the MPEG 2, 4, or 6 digital compression standards. When this type of broadcast television service via satellite first emerged, there was considerable debate about which analog standard would be used and whether a global high-definition television standard for use by BSS could be adopted. The conversion to the more efficient digital transmission systems and the development of the MPEG 2, 4, and 6 standards have served to bring standardization to the BSS world.

In the late 1970s and early 1980s, the European Space Agency was interested in launching an experimental direct broadcast satellite that was over time given various names, that is, L-Sat, then H-Sat, and eventually Olympus. This project was complicated by a procurement process in which the French and German governments decided, after the contractor was selected without a participant from their country, that they would not sign on to fund their “voluntary allocations” associated with this project. Instead, they decided to proceed with their own joint project known as TV-Sat in Germany and the FR-3sat in France.

As these various “official projects” proceeded to develop direct broadcast satellites in accord with ITU allocations, the Luxembourg-based company known as SES decided that it would use a high-powered Fixed Satellite Service (FSS) satellite to distribute a quasi-direct broadcast service directly to consumer homes and multidwelling units. SES proceeded to provide this FSS-based quasi-direct broadcast television service via a Ku-band satellite that had complete coverage for Europe. They provided what became known as direct-to-the-home (DTH) service. With this service, consumers could have small antennas installed at their home and receive a wide range of television programs – far wider than that offered by national terrestrial broadcasters.

In fact, it turned out that consumers cared little about ITU allocations or service definitions. In short SES, via its Astra satellites, stole a march on the official BSS alternatives. In the UK, the company known as Sky Television PLC also designed and launched a BSS system on an Astra platform. When there were some financial difficulties, Sky Television merged with another new project, namely, British Satellite Broadcasting (BSB) backed by Rupert Murdoch, and this TV service became known as BSkyB. Today direct-to-the-home (DTH) television satellite service is the largest identifiable satellite market and has grown consistently around the world in both developed and developing economies. Nevertheless, actual BSS or DBS systems continue to grow and have become predominant in the USA and Japan. In Europe, the competitive dual between BSS systems and DTH systems continue on in a significant way.

Satellite Radio Broadcasting

Satellite Radio Broadcasting Services represents an important additional aspect of the satellite broadcasting industry today. Commercial satellites from the earliest days were able to use broader band channels to send high-quality audio, music, and radio shows from one location to another. As the transition was made from analog to digital satellite transmission, the idea that satellite radio systems might be deployed came into much clearer focus. The motivations for this type of service came from many different perspectives. One motivation was from "official" national radio broadcasting systems that might be characterized as sending favorable "propaganda" on behalf of one country or another. During the Cold War years, very large amount of money was spent on establishing high-power terrestrial broadcast systems to send music, entertainment, and news to locations across "Cold War boundaries."

It was not surprising that many satellite planners thought that a global or regional satellite system that was able to broadcast to small compact shortwave radio receivers could be a technically more efficient and cost-effective method to send this information to millions of listeners. From another perspective, broadcasters, educators, and news people in the developing world recognized that there were more radios than television sets in some of the poorest countries. They believed that a satellite radio broadcasting services might be an effective way to reach a new and broader audience in these parts of the world. In the most economically advanced areas, entrepreneurs envisioned that a satellite radio broadcasting system might be a way to reach a broad new audience in their automobiles and even home listeners and office workers who wanted high-quality news, entertainment, and sports on a commercial-free basis by paying just a small monthly subscription fee. For these various reasons, the ITU allocated frequencies for a satellite broadcast service that is variously known as Digital Audio Broadcast Service (DABS), Direct Access Radio Service (DARS), and Broadcast Satellite Service-Radio (BSS-R).

It took over a decade of efforts to get such a new allocation for the service through the ITU processes. Part of the difficulty was that one needed to be able to have reliable service to a very small and low-cost receiver to make this offering viable. Allocations in the higher frequency bands would have problems with "rain fade" during times of very rapid rain rates. Also, signals in these regions, because of very, very small wavelengths, would need to be in direct line of sight connections to the satellite. This would make reception in an automobile or inside a building quite difficult if not impossible. The lower bands used for mobile service, that is, UHF (300–3000 MHz) do not require direct line of sight unlike the SHF band (3–30 GHz) and especially the EHF band (30–300 GHz), which require direct and uninterrupted access. After a very lengthy process, the result was to allocate the 2.3 GHz frequencies in North America for downlink broadcasting of this service. (These frequencies are sometimes known as being in the S Band.) In the rest of the world, the lower frequencies for downlinking in the 1.4 GHz band were allocated to the Digital Audio Broadcast services (this is sometimes known as the "L" band). These UHF frequencies are well suited for sending signals to vehicles and other locations without having a direct line of sight to the receiving antennas and are not subject to any significant atmospheric disruptions even with high rain rates, snow, or fog.

Once a frequency allocation for satellite radio broadcasting was finally agreed a number of companies actively pursued this business. At the lead was a company called Worldspace, which was headed by a charismatic Ethiopian visionary, Noah Samara, who had actually been a leading advocate of this new radio service and who had led the fight for new satellite frequency allocations within the ITU processes. Worldspace proceeded with the immediate design, manufacture, and launch of Worldstar satellites in partnership with the French firm of Matra Marconi. The first of these launches put Worldstar 1 into geo orbit to provide this new radio broadcasting services with coverage for all of Africa, the Middle East, and parts of Europe.

The business model for Worldspace was to offer a lease of one or more individual radio channels to broadcasters who would then provide their own programming for these satellite broadcasting downlinks. These channels could be used not only for radio news, sports, and entertainment, but at nighttime (or even daytime) might optionally be utilized to provide educational programming. Indeed, these digital radio channels could even be used to download short video educational programming over longer transmission periods. (This actually required a period of some hours to send a “slower and narrower signal” using the smaller “digital pipe.”) The cost of the radio sets and the attendant national import tariffs (that doubled the cost of the satellite radio receivers) ranged from about US $100 to US $200. This was unfortunately a high enough cost to create an unsuccessful business model. As a result, Worldspace encountered major financial difficulties almost from the very beginning.

In the USA, two systems known as Sirius (using highly elliptical orbits) and XM Radio (using geosynchronous satellites) were licensed by the FCC and both companies managed to successfully deploy radio broadcast satellite systems. The very high cost of building and deploying very large aperture satellites for this type of service plus high overhead and programming costs led to financial difficulties for these satellite radio services as well. This type of service was marketed largely to automobile owners on a subscription basis. Consumers were offered, on a 24/7 basis, a very wide range of radio programming and emergency communications and antitheft services. XM was offered via General Motors automobiles and Sirius was offered via Chrysler and Ford automobiles. When the financial difficulties mounted, these two systems merged in 2009 with XM Radio essentially acquiring Sirius.

In Europe, there were also plans for a radio broadcasting satellite service, but the financial problems with the other systems delayed the deployment of a European system.

Economic and Political Evolution of Global Telecommunications

The history of satellite communications is most often told in terms of the evolution of the technology that has developed rather continuously for the last half century. This has allowed the creation of larger, more powerful, and more proficient spacecraft with longer lifetimes and greater opportunity for automated operation. The evolution of the technology has allowed the ground devices to become simpler, easier to use, and less costly. It is quite remarkable that the initial communication satellite Earth stations were 30 m giants that require a 24 h a day crew of 40 people or so, and today, almost a century later, there are handheld satellite transceivers that any individual can purchase and carry around much like a cell phone. The development of the technology and the allocation of new frequency bands has also allowed the creation of a wider and wider array of satellite communications services that include fixed satellite services; large-scale networking among multinode corporate business satellite networks; broadband digital services based on the Internet Protocol (IP) standards; various types of television, audio, and radio broadcasting and media distribution; aeronautical, maritime, and land mobile satellite services; various types of search and rescue services; as well as various types of satellite communication links to support other satellite applications for remote sensing, precision timing, satellite navigation, meteorological and geodetic services, as well as scientific satellite missions.

ITU Key Role in Satellite Communications

The history of satellite communications also has an important economic, political, and regulatory dimension that is important to recount as well.

The International Telecommunication Union (ITU) made its first important effort to bring international order to the allocation of radio frequencies for the purpose of satellite telecommunications when it convened the Extraordinary Administrative Radio Conference in 1959 to address the issue of allocation of radio frequencies for the purpose of satellite communications. The ITU had for many years previously held its periodic World Administrative Radio Conferences (WARCs) to allocate frequencies for terrestrial radio-wave applications and applications such as radio astronomy, and it continues to do so although the name is now simply World Radio Conferences. The ITU is actually the largest specialized agency of the United Nations with over 200 member countries, and its role is to internationally coordinate all matters related to telecommunications and all forms of broadcasting such as technical standards, radio frequency allocations, interference mitigation, and telecommunications development for developing countries.

With the launch of Sputnik in 1957, the issue of allocation of frequencies for satellite usage was clearly an important new matter to address. At that meeting initial radio frequencies were agreed and another WARC in 1963 established a more mature framework. The International Frequency Registration Board (IFRB) was assigned the responsibility for recording each national administration's allotment of frequencies for satellite communications, and they also recorded the orbital locations assigned to geosynchronous satellite networks and the orbital characteristics of low or medium Earth constellations. The ITU also established the procedure for registration of national, regional, or global satellite networks. In the case of regional or global networks, a member country is designated to provide the information to the ITU. For instance, by way of example, the USA was designated to register intersystem coordination information for Intelsat and the UK was designated to register information on behalf of Inmarsat. The ITU also established procedures for the circulation to all members of official notices concerning new satellite system registration and procedures. This process also defined how technical coordination would be conducted in the event there were concerns with regard to technical interference between or among the various satellite systems in close proximity to the new satellite network.

At the start of the satellite age, most countries assigned the responsibility for post, telephone, and telegraph (PTT) services to a government ministry that had the monopoly right to provide these services. When the Intelsat agreements were negotiated, these international agreements had to be structured into two parts – one agreement for the governments and another for the operators. At that time, most operators were Ministries of the PTT although there were a few commercial companies such as Comsat (USA), Telespazio (Italy), and KDD (Japan) as well as what were called "crown companies" such as the COTC (Canada) and OTC(A) (-Australia). The "Operating Agreement" allowed for the commercial operators within Intelsat to assume an official role.

The dramatic increase in capacity that communications satellites represented over the transoceanic submarine cables in the late 1960s led to a significant decrease in the cost of overseas calls. The initial annual cost of an Intelsat two-way telephone circuit was set at $64,000 in 1965 when service first started, but within 7 years this rate had dropped to $8,000 and continued to drop as satellite capacity increased and satellite lifetime was also extended. The introduction of digital service (i.e., Time Division Multiple Access (TDMA) multiplexing) to replace analog service (i.e., Frequency Division Multiple Access (FDMA) multiplexing) drove costs and pricing even lower. At first there was a thought that the high capacity, low cost, and multiple destination satellites might replace submarine cables altogether as satellite costs and pricing plunged. But in the 1980s and particularly the 1990s, coaxial submarine cables grew in performance and throughput quite rapidly as well. Then fiber-optic submarine cables together with digital multiplexing and what was called wave division multiple access (WDMA) served to give a cost-efficiency edge back to the cable side of the telecommunication industry – at least for all of the heaviest routes of traffic.

In the 1950s through the mid-1960s, international overseas telephone line connections worldwide were measured in hundreds of circuits and an international call could be $15 a minute or more. By the end of the 1990s, the overseas connections were measured in the hundreds of thousands and the cost of an international call had dropped to levels equivalent to national long-distance calls. Today subscribers who use Voice over Internet Protocol (VoIP) services can call all over the world at virtually no additional cost other than a monthly connection fee. The annual cost of a submarine cable or an international satellite telephone circuit is so low (now well under US$5 to $10 per annum on the most efficient routes) that it is a minor part of the cost structure. Today the major costs associated with an international telephone or data connection, whether it is via cable or satellite, relate to marketing and billing and not to transmission costs. With the new high-throughput satellites, the cost per circuit has dropped significantly and is only slightly above that of a fiber link. If the automobile cost curves had followed the satellite and submarine cable industry efficiency gains on a proportionate basis over the last 50 years, one could today purchase a Rolls Royce for under $100 and one could drive over 100 km on 1 L of gasoline.

Submarine Cables and Communications Satellites

Much has been written about the economic competition between submarine cable systems and satellite communications networks, but in many ways there has been a co-development of both systems. In many ways, these systems have been complementary as often as competitive. First of all the two systems over time have tended to be mutually available for emergency restoration of service. There have been so-called Mutual Aid Working Groups (MAWGs) that have coordinated the ability to switch from one facility to another in case of loss of service and to respond to various emergencies that might occur. Cable systems (both in national terrestrial systems or international submarine systems) are quite vulnerable to earthquakes, volcanoes, or other natural disasters while satellites are not. Japan after the great Kobe earthquake that disrupted most communications and transportation on the southern part of Honshu decided it must undertake a fundamental change. It thus undertook to create a satellite system operating to a new network of Earth stations provided just to provide emergency backup in the case of natural disasters. These emergency Earth stations are largely installed at post office buildings all over the country of Japan.

Submarine cables with very high-efficiency fiber-optic transmission and dense wave division multiplexing (DWDM) are extremely efficient for very heavy telecommunications traffic between the USA and Europe. When it comes to television distribution or broadcasting over very large areas that are thinly populated, fiber-optic networks on the other hand are not well suited to such applications. Satellites, of course, also excel over cable systems for large-scale networking and mobile applications. Tables 2 and 3 both indicate the relative strengths and weaknesses and the relative performance levels of satellite networks vis-a-vis fiber-optic cable systems. Table 3 in particular provides a comparison of the cost-efficiencies and performance of different types of satellite systems when compared to cable.

Cable and communications satellite systems strengths and weaknesses more often than not complement each other. The advent of the World Wide Web and a global Internet has only strengthened this ability for the two technologies to reinforce each other. The other way of viewing the relationship between communications satellites and broadband cable systems is to examine technical performance. The following

Table 2 Comparing cable and satellite networks strengths and weaknesses (Chart courtesy of the author)

Relative performance strengths of communications satellites and cable systems			
Communications satellites		Coax and fiber-optic cable systems	
Strength	Weakness	Strength	Weakness
Mobile services	Point-to-point trunks	Point-to-point trunks	Mobile services
Large-scale business networks	Thick routes for concentrated cities	Dense urban networks	Dynamic large-scale business networks
Multicasting	Fixed distribution	Fixed distribution	Multicasting
Broadcasting for TV and radio	Fixed distribution in small geo. areas	Fixed, urban TV, or radio distribution	Broadcasting over very large areas
Connecting rural and remote areas	Intra-urban services	Intra-urban services	Connecting rural and remote areas
Dynamic networks with changing nodes	Stable and fixed node networks	Stable and fixed node networks	Dynamic networks with changing nodes
System reliability and rapid restoration	Systems with low tolerance for delay	System reliability and rapid switch over to backup systems	Vulnerable to natural disasters, construction digging, etc.

table shows the technical performance of a typical satellite and a typical cable transmission system. These performance levels will vary from year to year and from system to system around the world, but the relative scale has tended to be similar for the last decade or so. Fiber systems have considerably faster throughput and higher quality of service as measured in bit error rates, but their reliability (or system availability) can be less than satellites, particularly in transoceanic submarine cable installations. Satellites can be reconfigured more rapidly and are at their strongest for broadcasting, networking, or mobile services.

Satellites and the Internet

The advent of the Internet changed the world of global telecommunications more than any single factor in the past two decades. More and more telecommunications systems are digital and use the IP protocol to support every service whether it is for telephone, data, television, high-definition television, or mobile telecommunications. The world of satellite communications has had a more difficult time adjusting to this IP-based digital service because of the satellite transmission delay associated with geosynchronous satellites. New techniques to compensate for this transmission delay and the creation of IP over Satellite (IPoS) standards have allowed satellite systems to adapt to the global use of IP protocols.

Table 3 A technical comparison of satellite communications and broadband cable systems (Chart copyrighted by author and provided as a courtesy by the author)

Comparing satellite and fiber characteristics

Capability	Fiber-optic cable systems	Single GEO satellite in a global system	Single MEO satellite in a global system	Single LEO satellite in a constellation
Transmission speed	10 Gbps–8 Tbps	1–140 Gbps	0.5–5 Gbps	0.01–2 Gbps
Quality of service	10^{-11}–10^{-12}	10^{-7}–10^{-11}	10^{-7}–10^{-11}	10^{-7}–10^{-11}
Transmission latency	25–50 ms	250 ms	100 ms	25 ms
System availability	93–99.5 %	99.98 % (C-Ku band) 99 % (Ka band)	99.9 % (C-Ku band) 99 % (Ka band)	99.5 % (C-Ku band) 99 % (Ka band)
Broadcasting capabilities	Low	High	Low	Low
Multicasting capabilities	Low	High	High	Medium
Trunking capabilities	Very high	High	Medium	Low
Mobile services	None	Medium	High	High
Cost-efficiency	Very high ($5–$10 per year per transoceanic voice channel)	High ($20–$50 per year per transoceanic voice channel)	Low (more than $100 per year per transoceanic voice channel)	Low (more than $150 per year per transoceanic voice channel)

Pelton (2002)
The rate of 3.2 Tbps assumes 100 monomode fibers with each being able to transmit 80 Gbps

Conclusion

Satellite communications have now been in commercial service since 1965 and over 500 telecommunications satellites in a variety of low, medium, and GEO orbits now provide every conceivable service for commercial telephone, data, networking, audio, and video service. These satellites heavily support governmental and military communications as well. This history has been punctuated by several key drivers that have led to the rapid evolution of communications services. These drivers include the following.

- *The rapid evolution of the satellite technology.* These developments have allowed satellites to become more capable, higher in capacity, more powerful, longer in lifetime, and able to support more and more services as the user devices on the ground, on the sea, or in the air have become simpler, smaller, and lower in cost. The high-throughput satellites on one hand and the new MegaLEO systems represent the latest technology trends. In-orbit servicing may represent the next key technological step.

- *The evolution of new satellite services.* Today satellite communications support fixed and mobile telephone and data services plus television and radio distribution and broadcast services. These are known under their official ITU definitions as the Fixed Satellite Services (FSS), the Broadcast Satellite Services (BSS), and the Mobile Satellite Services (MSS) for aeronautical, maritime, and land mobile. There are also other communications satellites services such as short messaging services (also known as machine to machine (M2M) relay) and hybrid communications services directly linked to space navigation services as well as data relay satellites, but these represent a very small percentage of the overall market revenues for the industry.
- *The development in the satellite world of a competitive services industry.* Although commercial services started with one global enterprise known as Intelsat, the demand for different types of services, a global change in telecommunications regulation that brought about competitive systems, and market innovation have all led to many competitive systems in the satellite world.
- *The development of special needs to use satellites to support military- and defense-related purposes.* These satellites tend to have special features such as anti-jamming and radiation hardening and operate in different frequency bands. Dual-use satellites, wherein commercial satellite systems support defense communications, have become a key part of how such services are provided.
- *The parallel and competitive development of satellite communications in space and broadband cable systems on the ground has reshaped both industries.* Satellite communications initially had greater capacity than telephone submarine cables and could offer services at lower rates. Then fiber-optic technology reversed the trend. The different strengths and weaknesses of satellites and fiber-optic technology have led to a parallel deployment of these systems that are sometimes quite complementary and sometimes competitive.

The decision to use satellite communications systems today hinges on many factors such as the type of service needed; whether the service demand is for heavy (and concentrated) or thin streams of traffic; whether the service involves two-way traffic, multicasting, large-scale networks, or a broadcast service; and whether the service is to a fixed location or a mobile service. Another factor is whether the service is always to the same location or whether it is a dynamic network with nodes being added or subtracted on a continuous basis.

This is an overview history of the satellite communications industry since it started some 50 years ago. This history provides many of the key events and indicates the general technical, operational, regulatory, and market trends. For specific details of historical events, one can also consult various sources, as noted in the endnotes below (Pelton et al. 2004; Pelton and Alper 1986; Whalen 2002; Logsdon et al. 1998).

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Satellite Communications Video Markets: Dynamics and Trends
- ▶ Satellite Orbits for Communications Satellites
- ▶ Space Telecommunications Services and Applications

References

A.C. Clarke, *The Space Station* (Wireless World, London, 1945)

J.M. Logsdon, R. Launius, D.H. Onkst, S.J. Garber, *Exploring the Unknown Volume III: Using Space* (NASA, Washington, DC, 1998)

M. Mueller, *Intelsat and the Separate System Policy: Toward Competitive International Telecommunications, CATO Institute Policy Analysis 150*. (CATO Institute, Washington, DC, 1991), http://www.cato.org/pubs/pa-150.html

J. Pelton, *Global Talk* (Sihjthoff and Noordhoff, Alphen aan den Rijn, 1981), pp. 15–16

J.N. Pelton, *The New Satellite Industry: Revenue Opportunities and Strategies for Success* (International Engineering Consortium, Chicago, 2002)

J.N. Pelton, *Future Trends in Satellite Communications: Markets and Services* (International Engineering Consortium, Chicago, 2005), pp. 138–141

J.N. Pelton, J. Alper, *The Intelsat Global Satellite System* (AIAA, Washington, DC, 1986)

J. Alper, J.N. Pelton (eds.), *The Intelsat Global Satellite Organization* (AIAA, New York, 1986)

J.N. Pelton, S. Madry, How satellites service the world, Chapter 6, in *The Farthest Shore: A 21st Century Guide to Space*, ed. by J.N. Pelton, S. Madry (Apogee Press, Burlington, 2010)

J.N. Pelton, J. Oslund, P. Marshall (eds.), *Satellite Communications: Global Change Agents* (Lawrence Erlbaum Associates, Mahwah, 2004)

Special Message to the Congress on Urgent National Needs, John F. Kennedy address to a Joint Session of Congress, 25 May 1961, John F. Kennedy Library and Museum, http://www.jfklibrary.org/Historical+Resources/Archives/Reference+Desk/Speeches/

United Nations General Assembly Resolution of Satellite Communications, Resolution 1721, Section P (1961)

U.S. Congressional Hearings, House of Representatives, Communications Satellites Hearings, Part II, Interstate and Foreign Commerce Committee Hearings (1962), pp. 683–685

D. Whalen, *The Origins of Satellite Communications, 1945–1965* (Smithsonian Press, Washington, DC, 2002)

Space Telecommunications Services and Applications

Joseph N. Pelton

Contents

Abstract

This chapter examines the ever-increasing number of services and applications that are now provided by the commercial satellite industry. It explains basic types of satellite services as defined by the ITU for the purpose of radio frequency allocations – particularly the broadcast satellite service (BSS), fixed satellite service (FSS), and mobile satellite service (MSS). This section further explains that regulatory, standards, and policy actions by various international and regional organizations, plus commercial competition also leads to the development of different terms to describe new and emerging satellite services. Key to the development of satellite services and applications within the global

---

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_4

telecommunications market is not only the development of new satellite technology but also the competition between satellites and terrestrial wireless, coax, and fiber-optic networks. Satellites and terrestrial systems, despite being competitive, are nevertheless often complementary because they have particular strengths and weaknesses that do complement each other. Further, these systems are also used to restore each other against outages – particularly during natural disasters. Satellites have evolved in their offerings for nearly 50 years and will continue to do so in the future including services to interplanetary distances and perhaps beyond.

Keywords

Amateur satellites services • American National Standards Institute • Broadcast satellite service (BSS) • Broadcast satellite services for radio (BSSR) • Business networks • Digital audio broadcasting services (DABS) • Direct to home (DTH) television • Enterprise networks • European Telecommunications Standards Institute • Fixed satellite service (FSS) • International Electrical and Electronics Engineers (IEEE) • International Electrotechnical Commission (IEC) • International Telecommunication Union (ITU) satellite service definitions • Internet Engineering Task Force (IETF) • Internet protocol (IP)-based satellite services • Machine to machine (M2M) satellite services • Microsatellites mobile satellite service (MSS) • Nano satellites • Satellite telephony • Space communications for interplanetary and cislunar links • The International Standards Organization (ISO) • Very small aperture terminals (VSATs)

Introduction

The world of satellite communications services and applications can seem complicated due to the fact that there are a number of service descriptors that come from quite different sources. Some of these service definitions come from the International Telecommunication Union, while others come from other standards making bodies, national or international governmental regulatory agencies, or perhaps most frequently from the various communication satellite industry markets around the world. Sometimes, the differences in the names of satellite services are regional in nature; in other instances, copyright or trademark restrictions lead to differences in terminology. Despite a difference in the names for various services, in many cases, the actual satellite service in question may exactly describe precisely the same offering. Figure 1 in chapter "▶ Satellite Applications Handbook: The Complete Guide to Satellite Communications, Remote Sensing, Navigation, and Meteorology" indicates in a synoptic way the very wide range of commercial satellite services available today.

The International Telecommunication Union (ITU) is a specialized agency of the United Nations that has the prime responsibility for creating standards for satellite communications. It is also responsible for the allocation process for radio frequencies that are essential to satellite communications services. This organization establishes official definitions for a wide range of telecommunications satellite services and indeed does the same for the various other types of satellite applications described in later sections of this handbook (The International Telecommunication Union). These official definitions of satellite communications services, however, are not always used by national regulatory agencies or the commercial satellite communications markets around the world. Further, there are many other standards making bodies that create standards that relate to the provision of satellite services. These bodies and their standards sometimes lead to confusion concerning service definitions and uncertainty as to whether one particular term is equivalent to another. This problem can be compounded by commercial organizations which because of copyright and trademark restrictions often resort to using different phrases to describe the same service.

There is a lot of complexity just in the global standards making arena. The ITU and its regulatory and standards making processes, plus regional standards groups, address satellite issues from the perspective of telecommunications efficiency. Then there is the Internet Engineering Task Force (IETF) that is organized under the auspices of the Internet Society to address satellite networking issues. The IETF, however, develops its standards and terminology not from the perspective of satellite communications but in the context of improving Internet connectivity standards. Other standards groups that also develop standards, and in the process sometimes develop services or service requirements, are the International Electrical and Electronics Engineers (IEEE), the International Electro-Technical Commission (IEC), the International Standards Organization (ISO), as well as national and regional bodies such as the European Technical Standards Organization (ETSO) and the American National Standards Institute (ANSI). The complications extend further in that military- and defense-related organizations also often define their own terms for various specialized satellite services.

The commercial- and defense-related terms for applications and services are often driven by marketing and sales personnel. These terms are thus much more dynamic and change not only year to year but even month to month. These market-driven terms for satellite applications change for a variety of reasons such as trademark restrictions, a perceived market advantage, a new or altered national governmental regulatory restriction, a new way to set a tariff, or a dozen other reasons.

The various types of satellite communications services and applications will be first presented here in terms of ITU definitions and frequency allocations. This is simply because ITU terminology often serves as the “common language” of satellite communications and allows a basis for some global commonality when it comes to satellite communications services.

The remainder of this section provides an overview discussion and analysis of the commercial satellite communications services broken down by the various markets, including “dual use” of commercial satellite networks to support military or defense communications services. Later in this chapter, the actual development and growth dynamics of these markets are addressed in greater depth with regard to the major ITU-defined service categories of fixed satellite services (FSS), broadcast satellite services (BSS), and mobile satellite services (MSS). There is also a section with regard to the development and evolution of military- and defense-related satellite services that are carried on commercial satellite systems.

The first presentation and service definition relates to FSS applications. This section thus describes how satellites of this type provide connection between ground antennas that are fixed in their location. These FSS satellites typically provide two-way voice communications, various types of information technology (IT) relay, data services and data networking, video and audio distribution services, and commercial defense-related satellite communications. These are all services that are provided between two "fixed" earth station antennas or a network of fixed earth stations. Later chapters of this Handbook provide background and analysis as to how FSS services and applications relate to other commercial communications that are sometime competitive, and in other cases, satellites play a complementary role to terrestrially based telecommunications systems. These sections thus address how coaxial cable, Ethernet systems, submarine cable, and fiber-optic networks relate to and sometimes compete with commercial satellite communications systems in terms of economics and global division of markets. Most FSS applications provide interactive communications in "real time" between two antennas or perhaps among a network of fixed earth stations. In some cases, however, the FSS applications can include so-called store and forward and supervisory control and data acquisition (i.e., SCADA) noncontinuous communications. For these types of services simple data messaging is sufficient since the satellites in this type of configuration or constellation are typically not able to "see" at the same time the various locations that are to be connected. In these types of systems, a satellite picks up a data message at one point and then delivers the stored message at another location. This is why it is called a "store and forward" system. These are also sometimes referred to as "machine to machine" (i.e., M2M) systems.

The second broad category of commercial satellite service relates to "BSS applications" and the direct broadcast of satellite television, high-definition television, and radio. In this instance, the service is uplinked to a high-powered satellite that sends a one-way signal to a large broadcast audience. These broadcast satellites are capable of sending a signal directly to the consumer where it is received by a home or office micro terminal or in the case of broadcast satellite service radio to a small receiver that is sufficiently small that it can be installed in an automobile or other type of vehicle.

The third presentation and service definition relates to MSS applications and the type of satellite service that provides interactive communications with antennas for users that are on the move and can perform mobile communications. These MSS satellites can be used for communications to and from ships, aircraft, or land mobile vehicles.

Satellite Communications Services as Defined by the ITU

The membership of the International Telecommunication Union (ITU), headquartered in Geneva, Switzerland, includes virtually every country and territory in the world. Although it is a specialized agency of the United Nations, the ITU membership is actually larger than the United Nations Organization itself. This is not surprising, since today virtually everyone needs to use radio frequencies for a wide range of applications, and there is a global need for information and communications technology (ICT) services. One of the many functions of the ITU is to allocate frequencies for satellite communication services as well as to help with intersystem coordination so that the various satellite networks do not provide excessive interference to one another. One of the ways that the ITU functions with regard to radio frequency allocations is to define different types of satellite services and provide for the use of different radio frequency bands for these services. The various defined services, as agreed through global meetings, now known as the World Radio Conference (WRC) are specifically identified in Table 1 (World Radio Conference, of the International Telecommunication Union).

About half of these various ITU-defined services relate to commercial communications satellite offerings. All forms of satellite activities, whether for defense-related applications, remote sensing, space navigation, satellite meteorology, radio astronomy, time synchronization, space research, or space operations, need to operate active communications links to convey information to the Earth and to receive information and commands from Earth locations. ITU allocations of frequencies associated with these various satellite services is a complex activity. In most cases, there is a primary allocation of one or more frequency bands for these various services. There can also be a secondary and tertiary allocation for radio frequency spectrum (Pelton 1998; ITU Frequency Allocation Table).

Further organizations can also use certain bands, on a noninterference basis. In a number of cases, countries will place an "asterisk" against a frequency band allocation for a particular service in their own national boundaries. This means that they do

Table 1 ITU-defined satellite services with specific frequency allocations (Derived from Pelton 2006)

ITU-defined satellite services
Fixed satellite services (FSS)
Inter-satellite services (ISS)
Broadcast satellite services (BSS)
Broadcast satellite services for radio (BSSR)
Radio determination satellite services (RDSS)
Radio navigation satellite services (RNSS)
Mobile satellite services (MSS)
Aeronautical mobile satellite services (AMSS)
Maritime mobile satellite services (MMSS)
Maritime radio navigation satellite services (MRNSS)
Land mobile satellite services (LMSS)
Space operations satellite services (SOSS)
Space research satellite services (SRSS)
Earth exploration satellite services (EESS)
Amateur satellite services (ASS)
Radio astronomy satellite services (RASS)
Standard frequency satellite services (SFSS)
Time signal satellite services (TSSS)

not accept this allocation within their own country. This will be explained in greater detail later in the section on ITU functions and, especially, its role with regard to frequency allocation.

All of the communications satellite services are defined by the International Telecommunication Union. In many cases, however, other terms are used to describe these same services as they are marketed within the commercial world.

Fixed Satellite Services

The "official" ITU definition of fixed satellite services (FSS) is as follows: "A radio-communication service between earth stations at given positions, when one or more satellites are used; the given position may be a specified fixed point or any fixed point within specified areas; the fixed satellite service may also include feeder links for other space radio-communication services" (ITU Radio Regulations, Article 1, Definition of Radio Service, Section 1.21).

This was the initial form of commercial service that began in 1965 with the Intelsat global satellite network for international links and the Molniya satellite network that provided domestic services within the Soviet Union. Fixed satellite services today provide applications that include telephone, facsimile, various data services, audio distribution, videoconferencing, video distribution, multi-casting services, and corporate enterprise services such as virtual private networks. One of the more rapidly growing data services on fixed satellite systems support IP networking for Voice over IP and broadband IP data services. This type of IP-based service via satellite can be broadband data transmission to support heavy "trunking" links between cities or countries or it can support relatively wideband Internet connections directly to end users at small office/home office locations (i.e., so-called SOHO connections). These commercial satellite services typically operate in the so-called C-band (6 GHz uplink and 4 GHz downlink), the Ku-band (14 GHz uplink and 12 GHz downlink), and most recently the Ka-band (30 GHz uplink and 20 GHz downlink). Additional frequencies are used for military satellite services. All these frequency band allocations by the ITU will be discussed in detail in the section of the Handbook addressing this issue.

Broadcast Satellite Services

The "official" ITU definition for broadcast satellite services (BSS) is as follows: "A radio-communication service in which signals transmitted or retransmitted by space stations are intended for direct reception by the general public. In the broadcasting-satellite service, the term 'direct reception' shall encompass both individual reception and community reception" (see Fig. 1 above for current BSS type satellite) (ITU Radio Regulations, Article 1, Definition of Radio Service, Section 1.25).

Despite the fact that this ITU definition of the broadcast satellite service (BSS) is quite detailed, and despite the fact that specific frequencies are allocated by the ITU

Fig. 1 State-of-the-art Direct Broadcast Satellite-Echostar XIV – Manufactured by Space Systems/ Loral (Graphic Courtesy of Space Systems/Loral)

for this broadcast service, there are nevertheless ambiguities in terms of types of satellite services actually delivered in the marketplace. The problem is that there are overlaps that occur between BSS and FSS systems in practice when it comes to how satellites in orbit are utilized. Some very high-powered FSS systems have been used not only to distribute television signal to cable television head ends and other locations for redistribution to the public but also to provide service directly to the consuming public. This is sometimes referred to as "direct to home" or DTH service.

The Home Box Office (HBO) system began delivering television programming to cable television "head ends" via satellite in 1975 using FSS satellites. Consumers in rural and remote areas responded to this "opportunity." They bought "backyard satellite dishes" and sometimes "decramblers" to see the video signal and hear the audio. Thus an informal type of direct broadcast satellite or direct to home service was born in the United States, Canada, and many other countries. The Astra system in Europe was the first FSS system to launch entire satellites based on the DTH business model of selling television service directly to consumers (A global overview of Direct to Home Television services).

In what might be called the "reverse situation," BSS systems have been used to deliver what have been traditionally considered the domain of FSS applications. In this case, individuals and distributed business offices have utilized BSS systems to obtain two-way high-speed data services using protocols such as Digital Video Broadcast with Return Channel Service (DVB-RCS) or Digital Over Cable System Interface Standard (DOCSIS) to deliver digital video and high-speed data "to the edge" of business enterprise networks. To obtain both television services and asymmetrical data distribution with thin route message return, the business user only needs to install very small aperture antennas (VSAAs). Such “asymmetric satellite ground systems” can not only receive high-quality digital audio and television as well as high-speed IP-based data but also transmit low data rate return messages. This type of operation is almost ideally suited to support typical digital telecommunications services to an asymmetrical corporate-type data network with interactive services to the “edge” of an enterprise business hookup. Companies with national, regional, or even global business networks such as automobile manufacturers with highly distributed dealerships, oil companies with a large network of gasoline service stations, department stores with many outlets, etc. would use such satellite business networks to support communications to their corporate headquarters or to credit card authorization offices. Such entirely digital networks can be operated on either FSS or BSS networks.

In such cases, there will be a high-speed downstream blast of data at many megabits/s (typically at speeds like 30–75 Mbps) and then there would be a thin route uplink response capability. These types of asymmetrical data distribution service with a return channel link actually operate most typically on an FSS system but can now also operate on BSS systems as well. In a telecommunications service world with data networking supporting video, audio, and high-speed data (that can be telephony or digital streaming), the division between FSS and BSS service tends to be increasingly blurred. One example is that the Boeing Corporation manufactured three high performance digital satellites known as Spaceway 1, 2, and 3. Two of these satellites are deployed by the Direct TV network to provide direct broadcast BSS services in the United States, but the third satellite in the series, the Spaceway 3, is utilized by the Hughes Network Systems (HNS) to provide high-speed digital services to customers under its Hughes Net offering. The three satellites are essentially the same, but the utilization is dramatically different (The Boeing Company Manufactures and Deploys the Spaceway Satellites).

There is currently no ITU enforcement of how various satellite systems are used. Indeed, over the past decade various proposals have been made at various ITU conferences to allow more flexibility in service definitions. The argument behind these proposals is that if there were “multiple usage allocations” provided for satellite communications frequencies, this would encourage the most efficient and effective use of satellite spectrum. Currently, the reverse is true. ITU allocations for satellite services tend to be what might be called “overly precise.” The ITU allocations and the frequency planning process for the BSS services tend to be geared to particular orbital locations for satellites and the available frequencies vary over the three ITU regions around the world. Detailed information on the frequencies available for this service is provided in later sections.

The ITU has also defined a service known as broadcast satellite service radio (BSSR) with frequencies allocated in a lower band to accommodate this service that is more narrowband than broadcast satellite-television services. This service can also be called simply satellite radio or direct broadcast radio satellite. Since most systems use digital broadcasting to support efficient transmission, it can also be called digital broadcast radio satellite systems or direct broadcast radio systems or digital audio broadcast satellite (DABS) system. In this type of satellite radio service, the audio broadcast signal is digitally encoded and broadcast to Earth-based receivers from an orbiting satellite directly to a receiver in a home, office, or vehicle. In some cases, it is sent to a "repeater station" that then rebroadcasts the signal to receivers. The Worldspace, XM Radio, and Sirius Radio satellite systems operate in these bands. The BSS or direct broadcast satellite systems that operate in the higher Ku-band transmit a large number of video and high-definition television channels, however, also broadcast high fidelity audio channels as well. The XM Radio and Sirius Radio satellites (now merged into a single system under the ownership of XM Radio) primarily transmit their programming to vehicles on the move equipped to receive their signals – with subscribers paying a monthly subscription fee. World space, which is nearing the end of its operations after experiencing financial and revenue stream problems, transmits its digital audio signals largely to developing countries where television service is limited or nonexistent.

Mobile Satellite Services

There are, in fact, a number of mobile satellite services that include land mobile satellite services (LMSS), aeronautical mobile satellite services (AMSS), and maritime mobile satellite services (MMSS).

The official definition by the ITU of mobile satellite services (MSS) is as follows: "A radio-communication service (a) between mobile earth stations and one or more space stations, or between space stations used by this service; or (b) between mobile earth stations by means of one or more space stations. This service may also include feeder links necessary for its operation." (ITU Radio Regulations, Article 1, Definition of Radio Service, Section 1).

Today there are some satellite systems, such as the ones operated by Inmarsat Ltd. or Thuraya, that have satellites with sufficient power and flexibility to support service to aircraft, ships, and land mobile units. Because these systems are equipped with enough power to operate mobile units, they can clearly also be used at fixed locations. This is because it is much easier to complete a satellite link to a fixed location where there is a clear line of sight to the satellite that to a constantly traveling mobile unit. In the case of mobile satellite service, blockages to a clear line of sight between the satellite and the mobile antenna can occur at any time due to tunnels, forestation, utility poles, or buildings. Since this is the case, broadcasters that cover a wide variety of events from the field may set up a temporary "fixed" location and uplink and downlink from a mobile satellite system. Offshore platforms that are "fixed" often rely on mobile satellite systems as well.

In short, ITU definitions are precise and seek to separate the three main satellite communications services into the categories of FSS, BSS, and MSS. These exact divisions of the three primary satellite services are not always observed in practice. Since there are no "enforcement officials" or fines imposed these distinctions have no practical implications. In most cases, the prime use of the satellite system is consistent with the defined service and the spectrum band allocated to that service.

There is a new type of mobile satellite system that has originated in the United States and is under planning in other regions of the world. This is the mobile satellite service with so-called ancillary terrestrial component (ATC). This type of mobile satellite service is known in Europe as a mobile satellite system with a complementary ground component (CGC). The first of these satellite systems for service in the United States, namely, the SkyTerra system and the Terrestar system (as pictured in Figure 2), need to deploy extremely large satellite antennas. These gigantic antenna systems with huge apertures more than 15 m across allow a very large number of powerful beams to be generated from space. These systems are unique in that they are designed to operate in tandem with terrestrial mobile satellite systems. The concept is that terrestrial wireless mobile services provide high-quality mobile connections within well-covered urban areas, but mobile satellites are used to provide broad and complete coverage outside the city with the service seamlessly switching from terrestrial wireless to satellite as the consumer moves from one point to another.

The financial status and viability of businesses providing mobile satellite services in North America has changed dramatically in the past 2 years. Both Terrestar (formerly Sky Terra) and DBDS North America (formerly ICO Ltd.) experience financial setbacks and both formally declared bankruptcy in 2010. Then, both were acquired by Dish Inc. in 2012 for $2.3 billion out of bankruptcy court in order to provide mobile satellite services on a consolidated basis. LightSquared has also experienced major difficulties because its frequencies and GPS frequencies have significant interference. The Terrestar 1 satellite is shown in Fig. 2.

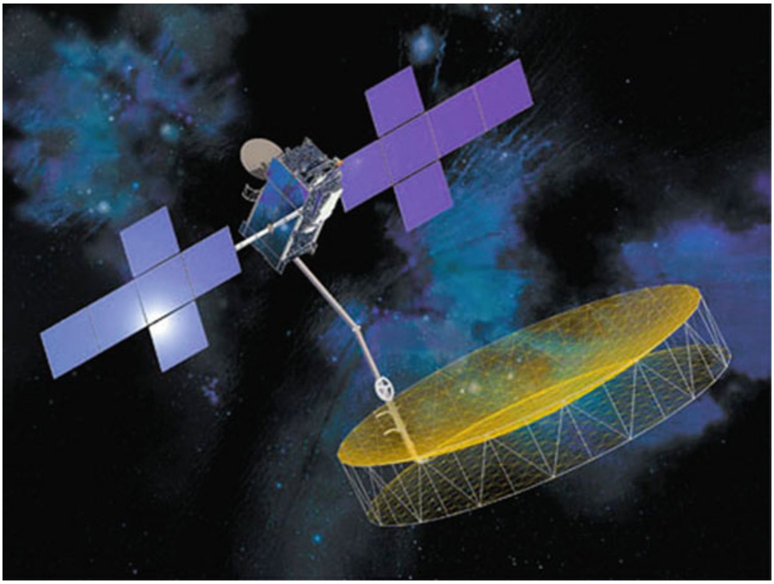


Fig. 2 The gigantic Terrestar 1 satellite pictured as deployed in space (Artwork Courtesy of Terrestar)

The radio frequencies allocated for this service typically start at around 1,700 MHz and range up to 2,600 MHz in an assortment of bands. The lower radio frequencies allocated for this service (as opposed to FSS and BSS service) are important because of signal blockage. When mobile satellite users are on the move, there is a problem of signal blockage. Factors such as foliage from trees, buildings, signs, utility poles, and especially the roof tops of vehicles can serve to block a direct line of sight signal from mobile satellites to the user. The lower frequencies with longer wavelengths that can “bend” around obstacles are more tolerant of partial pathway blockage. These lower frequencies in the so-called L-band and the UHF frequencies can still close a link to a mobile satellite without having a direct line of sight between the satellite antenna and the user antenna device. In contrast, the higher frequencies used for FSS and BSS services typically need to have a direct line of sight to the satellite to complete a transmission.

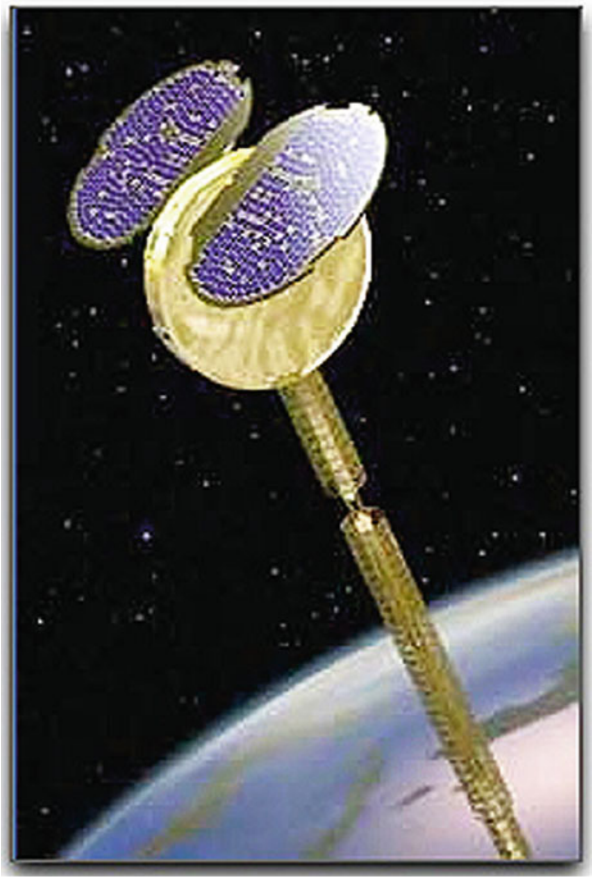
Higher frequencies in the Ka-band are also allocated to mobile satellite services, but these are not typically used. This is because the frequencies in the 30 GHz band with quite tiny wavelengths are not tolerant of interference and essentially require direct line of sight to close a link between a satellite and a user’s mobile antenna. Specific information on these allocations is provided later in Handbook.

Other Types of Commercial Satellite Services

The three most important commercial satellite communications systems are thus those providing fixed, real-time broadband satellite services, broadcast satellite service, and mobile satellite services. These services represent the overwhelming amount of revenues associated with satellite communications. There are, however, some other forms of communications satellite systems that exist and provide services to various users around the world. In some cases these are thin route commercial services and in other cases these are satellite that are used to serve educational, medical, first responder and emergency communications, and amateur “ham radio” operators.

1. *Store and Forward Satellite Systems* (also known as M2M): These systems, that rely on microsatellites, are typically deployed to provide non-real-time data relay. The most well known of the M2M commercial satellites systems is the Orbcomm satellite network that provides a global data relay satellite service for what is often called “machine to machine (M2M)” communications. This is a commercial service that can relay data around the world with a minimum of delay. The M2M types of satellite system deploy ground transceiver units that can be positioned at fixed locations (such a remote mine or offshore drilling rig) or as vehicular mounted systems. These vehicular mounted units are designed not only to transmit and receive data, but they are also often paired with a space navigation system that can indicate the vehicle’s location in real time as well. One of the more common application for this type of satellite network is for “supervisory

Fig. 3 One of the first generation of Orbcomm satellites that created a "fast" store and forward global LEO constellation (Graphic courtesy of Orbital Science Corporation)



control and data acquisition" (SCADA) services that are used for control of power stations, monitoring of oil and gas pipe lines, and managing various types of mobile fleets from trucks, buses, or rail systems to ships at sea.

The Orbital Sciences Corporation designed and built the first generation satellites for this system and deployed 35 of these micro satellites into low earth orbit using its Pegasus and Taurus launch vehicles. These lightweight satellites could be "stacked" together in a launch vehicle and launched eight at a time (Fig. 3).

This network was first established as a part of Orbital Sciences but divested as a separate entity when the initial system experienced financial difficulties. Orbcomm Ltd. is now a separate company and independently owned. The second generation of this system involves a new constellation of 18 satellites that supplements the initial system. These satellites are under contract for launch by the Space Exploration Technologies Corporation (known as Space X). This next generation of satellites were built by the Sierra Nevada Corporation and its Microsat subsidiary plus a team of contractors that included Boeing and ITT (Orbcomm 2008).

Applications associated with commercial store and forward satellite networks can be to send and receive information from trucks, buses, trains, dispatch vehicles, and even ships and airplanes and also fix the location of the mobile unit so this can be relayed to a home office. Although different frequencies and different satellites are used for the data relay on one hand and for position location on the other, the mobile unit is consolidated and relies on a common battery. These consolidated units that provide this type of service are popular with car rental agencies, bus, and rail systems and shipping lines. This machine to machine (M2M) type of service provides a reliable form of data relay that can often relay a message on a global basis in a matter of minutes and provide space navigation services on an instantaneous basis using GPS satellites. The M2M satellites typically operate in the 137–138 MHz and 400 and 435 MHz frequencies. Since these are very narrow bands, they have very limited service capacities that relay only short messages.

In addition to commercial M2M services such as Orbcomm provides, there are a number of small satellites that are used to support the mission of nongovernmental organizations operating around the world for educational or research purposes. These satellites are often designed and built by the small satellite program at Utah State University in the United States or by the Surrey Space Center in the United Kingdom. In some cases, they are built by newly emerging national space programs that are built in cooperation with the Surrey Space Center as a way of launching their initial application satellites programs. There are also commercial companies that design and build so-called microsatellites, such as Microsat Systems Inc. and Space Dev who often design small satellites for defense-related missions. There are also number of small firms that manufacture nano satellites (i.e., 50–100 kg) including companies like ISIS, GomSpace, and UTIAS-SFL.[1]

An example of the small store and forward satellite systems designed by the Surrey Space Center is the two-satellite low earth orbit “Lifesat” system. This microsat system was designed to provide medical information on demand via satellite to rural and remote parts of the world with a guaranteed response time of 2 h or less using Surrey-designed satellites. This type of microsatellite designed and built at the Surrey Space Center or Utah State University can be designed for many applications. Thus, in addition, to store and forward microsatellites for communications, these small satellites can also be designed to provide remote sensing, scientific missions, or other space operations. These satellites, which typically range from 10 to 100 kg, are often launched as a “piggy back” operation added on to the launch of one or more larger satellites. There are also very small “cube satellites” or “nanosatellites” in the 1–10 kg range that are typically built by students at universities to carry out short-term experiments that are capable of relaying small bursts of data communications, but because these are often restricted to battery power only, they have only a short duration lifetime. These satellites typically operate at 137 and 400 MHz.

---

[1] For more information on micro satellites and nano satellites see www.satellite-links.co.uk/links/satman.html

2. *Amateur Satellites*: These satellites are generically known within the industry as Amsats but to the world as OSCARs. These are microsatellites designed to be used by amateur "ham radio" enthusiasts around the world to relay messages. This is also essentially a "store and forward" satellite data service, but they also include voice transponders to support short bursts of conversations. Since these satellites are launched into low earth orbits, they cannot support continuous communications because the satellites come quickly in and out of range as they pass overhead of the ground based amateur satellite operators and their transmit and receive equipment. This represents one of the largest applications of this type of technology in terms of the number of satellites launched and the number of participants. These satellites are known as "OSCARs," which stands for operating satellites carrying amateur radio. Figure 4 shows OSCAR I that was launched on December 12, 1961. This was the first microsatellite to be launched as a "piggy-back" operation. In short, it was launched in tandem with a much larger satellite as a much smaller ancillary or "secondary" payload. Since that time there have been hundreds of launches wherein microsatellites were launched along with a larger primary mission. Over 70 OSCAR satellites have been launched since the first launch half a century ago.

These satellites are built typically to operate in the following radio frequencies: 28–29.7 MHz, 144–146 MHz, and 435–438 MHz (on a noninterference basis).[2] Thousands of amateur radio operators all around the world use these satellites to relay messages. This is not only as a hobby, but many times this has proved a very important way to relay messages during disasters and other types of emergencies where conventional communications systems may have failed or are temporarily out of service.

Overview of FSS Services: Telephony, Information Technology (IT) Services, and Enterprise Networks via Communications Satellite

The fixed satellite services (FSS) began commercial service in 1965 with the Intelsat I (or Early Bird) satellite leading the way. These first commercial communications satellites were able to support telephone, telex, radio and audio service, and for the first time live international television service – although of quite low quality.

---

[2] The International Amateur Radio Users Guide to Frequency Coordination provides a full listing of approved frequency bands under the International Telecommunication Union Radio Regulations for all three ITU Regions. Provision 5.282 of the ITU Radio Regulations specifies the following: "In the bands 435–438 MHz, 1,260–1,270 MHz, 2,400–2,450 MHz, 3,400–3,410 MHz (in Regions 2 and 3 only) and 5,650–5,670 MHz, the amateur-satellite service may operate subject to not causing harmful interference to other services operating in accordance with the Table (See 5.43) [i.e., The ITU Allocations Table] Administrations [of the ITU] authorizing such use shall ensure that any harmful interference caused by the emissions of a Station in the amateur-satellite service is immediately eliminated in accord with the provisions of the "Spectrum Requirements for the Amateur and Amateur-Satellite Service, International Amateur Radio Union, August 2008" http://www.iaru.org/ac-08spec.pdf

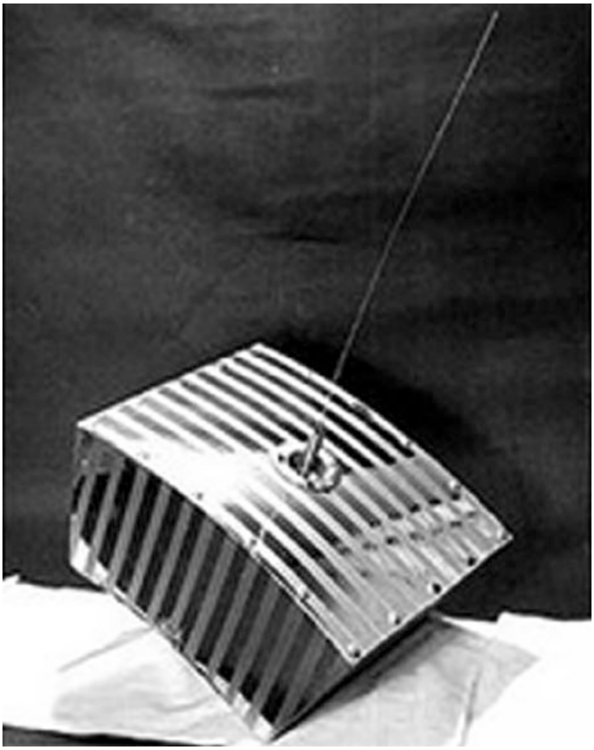


Fig. 4 The OSCAR 1 satellite launched in 1961

Overtime satellite transmissions from space were sent with higher power and higher gain antennas and the quality of the television signal improved. These next generation satellites, as deployed a few years later, could support full color and reasonably high-quality audio and television in all of the international standards then used, namely NTSC, SECAM, and PAL. The range of telecommunications and information services that are offered by FSS networks today is quite extensive. A listing of some of the most important types of services is provided in Table 2.

It should also be noted that commercial satellite systems also often fall into the categories of international satellite networks, regional networks, and domestic networks. International networks can, in fact, be utilized to provide not only global interconnectivity around the world but also capacity, often on a transponder lease basis, for either regional or domestic services. Examples of global networks are Intelsat, SES Global, and Eutelsat. Arabsat is a prime example of a regional system and there are many dozens of domestic systems around the world, although in some instances such as Koreasat these networks serve as a regional system as well. A worldiwide inventory of FSS systems is provided in the Appendices to this handbook. When the first FSS systems were deployed, they represented the widest band capability for transoceanic service and dominated the overseas markets for video, voice, and data services.

Table 2 Examples of satellite applications divided by generic service category

FSS satellite–based telecommunications and information services
Voice- and telephony-based services
Rural telephony
Telephone connections for locations in hostile natural environments
Transoceanic and regional telephony
Remote connectivity to research stations, offshore drilling stations, mines, etc.
Submarine cable and terrestrial telecommunications network restoration
Audio-based services
High-quality audio and music downloads (8–32 kHz channels)
Radio programming distribution
Video-based services
Television distribution to cable television head ends (conventional, high definition, and 3D high definition)
International television relay
Remote newsgathering and transfer to television production centers
Direct to home (DTH) television (i.e., quasi-direct broadcast services)
Television to ships at sea and commercial aircraft
Video-based tele-education and tele-health services
Digital networking services
Broadband Internet "trunking" or heavy route interconnection
Internet services directly to the Small Office/Home Office (SOHO)
IP-based telephony (Voice over IP) and multimedia over IP
Digital Video Broadcasting and network data distribution
Digital Video Broadcasting with Return Channel Service (DVB-RCS) or DOCSIS services (i.e., high data rate distribution with narrow band return)
Multi-casting IP-based services
Enterprise networking via VSAT and micro-terminal networks
Business television networks
Remote office and plant management systems
Scientific network connections
Digital networking for defense and military applications (including Comms on the Quick Stop)

The emergence of fiber-optic cable systems with their even greater throughput capabilities and lesser transmission delays in comparison to geosynchronous communications led to these terrestrial systems reestablishing market dominance for virtually all of the heavy telecommunications traffic streams around the world – especially for voice traffic. With the transition from analog to digital traffic and the growth of Internet Protocol (IP) transmissions, satellite systems have needed to make special adjustments. This is because satellite transmission delay can be "perceived" as system congestion and lead to slow recovery modes and thus transmission interference. Also IP Security (IPSec) over satellite routing leads to difficulties when "header information" is stripped from a satellite transmission.

A great deal of progress has been made to develop new standards to allow efficient IP-based satellite transmission over geosynchronous satellites where the greatest transmission delay occurs, but these issues will be discussed in greater detail below. A significant number of FSS systems in the developing world continue to operate in the 6 and 4 GHz frequencies (known as the “C band”), while many of the networks that are providing services to enterprise networks and support large scale VSAT corporate networks today operate in the higher spectrum bands, namely the 14 and 12 GHz frequencies (i.e., the “Ku bands”) and the 30 and 20 GHz frequencies (i.e., the “Ka bands”).

Commercial FSS satellite systems can be used for virtually all types of telecommunications services that include among other things telephony, fax, data, terrestrial and submarine cable system restoration, interconnection of rural wireless telecommunications networks, television, audio, and IP-based data networking. The satellite-based IP services can be used for both heavy route IP-based major system interconnection (or “trunking” services) and for services that are provided directly to individual subscribers and small businesses. On a global basis, FSS satellites are still employed for all these services and more. Despite this broad range of potential applications, satellites for communications are most efficient when they operate in the following modes and for the reasons provided in Table 3.

Overview of Video and Audio Broadcast Satellite Services

By far, the predominant market in the field of satellite communications relates to video and audio services. A significant portion of the FSS market comes from satellite television and audio distribution services. Some of the higher-powered FSS satellites actually transmit television and audio services directly to homes, offices, multi-dwelling units, and even to ships at sea. These FSS satellites are sometimes also employed to send video programming directly to third and fourth generation broadband cell phones. The direct broadcast satellite market or, in ITU terminology, the BSS market, as indicated in Table 3, represents over 50 % of the total satellite communications revenues worldwide. These types of direct broadcast services are (except for the exceptions noted above) different from FSS services in that they typically connect to end users and bill consumers directly, as opposed to the FSS services that are connecting business, governments, and other large scale users and usually are not offering services to general consumers on a “retail basis.” The BSS services are thus offering retail services to millions of consumers and billing them directly. These BSS service providers are further up the “value added ladder” of commercial offerings than the FSS service providers. Rather than having millions of direct consumers as customers, the FSS providers have a much smaller pool of large commercial customers or governmental agencies that in turn relate to the actual public or end user.

Audio services are typically provided in tandem with television distribution and/or broadcast television services. In addition to the BSS television and related audio services, there are now a growing number of special digital audio broadcasting-satellite services, such as those associated with Worldspace, XM Radio and its Sirius subsidiary service. These direct radio or audio broadcast satellite

Table 3 Prime usages for fixed satellite services

The most efficient FSS satellite applications	
FSS application	Reason for satellite efficiency in this application
Television distribution over broad areas	Satellites can provide broad coverage over very large areas, including locations with difficult terrains or locations that are hard to reach via terrestrial media. An antenna for transmission or reception can be quickly added or removed at modest expense. Hundreds, thousands, or tens of thousands satellite receivers can be added to a television distribution system at virtually no additional cost to the satellite transmission system in space
Communications services to rural and remote areas	Satellites, by covering large areas with a signal, do not require a concentration of users at a particular point to make a satellite transmission cost-effective. In comparison, laying a fiber-optic or coaxial cable or building a microwave relay tower system to reach only a few users in isolated areas is often not an economic proposition. In some cases, satellites are used to interconnect rural wireless telecommunications networks
Communications services to islands with small populations	Islands with only a few inhabitants are not economic for fiber-optic or coaxial submarine cable systems. Thus satellites that cover broad areas of the oceans are economic ways to serve islands and colonies with modest populations. Again, in some instances satellites are used to interconnect rural terrestrial wireless telecommunications networks
Communications services to large-scale networks over broad areas	Large-scale networks for such applications as credit card verification for gasoline/petrol stations, grocery store or retail chain stores, or for networking together car dealerships, hotel/motel chains are cost-effective satellite applications. Satellites work well for these applications because they can provide connectivity over broad areas and for many flexible locations at low cost and constant updatability of service locations
Multi-casting	This is quite similar to large scale networking, except that data messages can be selectively sent and received at multiple addresses that vary within a large network. This is useful for functions like inventory control messaging, retail pricing by geographic area, etc. The advantages are thus the same as for large-scale networking
Short-term events or remote news gathering	Satellites do not require long-term installation of antennas. Temporary access to a videoconference, up-linking of television news or other short-term events do not require the laying of a terrestrial cable or other "permanent infrastructure" in the ground. A truck mounted satellite earth station, a fly-away terminal or other temporary antenna can be used for a brief period and then relocated to the next place where temporary access is needed

services represent an additional and growing satellite communications market. In the case of this type of service, radio programming is often provided directly to vehicles and individual mobile consumer receivers. These audio broadcasting services are often offered along with associated security or road assistance programs such as the “Onstar” service provided via XM radio to the owners of General Motor vehicles that subscribe to the service. These systems are also often designed to combine the “radio” and “security related services” with a space navigation service as well. In the case of an accident, this allows emergency services to know exactly where to go. In the case of a criminal issue such as car theft, a stolen car can be disabled so it will no longer operate.

Defense-Related “Dual Use” of Commercial Communications Satellite Systems

When commercial satellite services began, the first applications were essentially focused on governmental agencies, television and radio broadcasters, and telecommunication companies seeking international and transoceanic services. Overtime these markets expanded to cover additional regional, domestic, rural, and island telecommunications services. Defense-related communications via satellite were quickly seen as major applications since satellites could provide broad coverage and transportable earth stations could be set up at remote locations. For similar reasons satellites were also seen a key way to respond to natural disasters and other emergency situations.

As early as 1965, at the same time as the first commercial satellite system, namely Intelsat, was deployed, the US military launched a low earth orbit constellation known as the “Initial Defense Satellite Communications System.” It was designed to test the feasibility of having a dedicated satellite system for defense and security communications. The Soviet Union, in 1965, also deployed a system known as “Molniya.” This three satellite system was designed to provide continuous coverage of the Northern latitude regions of the world. This was accomplished by launching three satellites in highly elliptical and inclined orbits (i.e., orbits that were highly inclined with respect to the Equator. Note: A satellite that travels in the plane of the Equator has a 0° inclination, and a satellite that orbits around the Earth from the North to South Pole is essentially inclined 90° to the Equator.). This allowed each of the Molniya satellites to be visible above the horizon for at least 8 h a day. This Molniya system was used for a mixture of domestic governmental and civilian communications for the Soviet Union as well as for military and defense purposes.

The military forces of countries around the world quickly concluded that many nontactical communications requirements related to defense operations, such as entertainment for troops and personal communications for overseas personnel who wished to talk to families, etc., could be more easily provided and at lower cost by commercial satellite systems. Thus many defense-related communications of a non-tactical basis that required the use of satellite networks migrated to commercial systems. Such usage became known as “dual use” of commercial systems to provide civilian as well as defense-related communications.

Article XIV of the now defunct Intelsat Agreements that covered the coordination of the Intelsat system with other separate satellite facilities simply stated that: “This Agreement shall not apply to the establishment, acquisition or utilization of space segment facilities separate from the Intelsat space segment facilities for national security purposes” (Intelsat Organization 1973). These Agreements also explicitly indicated that Intelsat should not create “specialized satellite service facilities” for military purposes, but it was silent on what forms of satellite services might be carried on the Intelsat systems for military or national security purposes. In practice, Intelsat carried a range of “dual use” defense-related services but did not provide “tactical military communications services involving the direct application of military forces in a combat environment.” Other commercial satellite systems have tended to follow that same pattern.

Over time, however, commercial satellite systems have become more adept at providing “dual use” services since these types of applications represent a significant revenue stream. Particular ways that commercial systems have adapted to military-related types of services include:

- The ability to provide efficiently encrypted communications services
- The creation of special units within the commercial satellite operators that are especially designed to handle governmental or military communications satellite services
- The ability to accommodate specific military- or defense-related requirements such as “communications on the move,” “communications on the quick stop,” and coverage of isolated or littoral areas
- Tailored lease arrangements so that communications satellite services can be “called up” and be available when required
- Arrangements for the quick launch of capacity when a particular part of the world requires additional satellite communications services

Despite these specific steps by commercial satellite system providers to accommodate military, defense, or emergency communications requirements, there have been new innovations in recent years to respond to specific requirements. These innovations fall into two new categories of commercial satellite services.

One is the case where specific defense-related satellite service requirements are procured from commercial contractors on the basis of a long-term lease, with any additional capacity be available for sale to regular commercial customers. The first such instance was the so-called Leasat, where capacity was made available to the US Navy and additional service was sold to maritime service customers. Today several European countries are obtaining defense-related services in this manner. The United Kingdom “Skynet” and other similar defense-related satellite service programs will be discussed in chapter “▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications.”

The other case is where commercial satellite consortia or companies design and build a communications satellite designed to operate exclusively in military- or defense-related frequency bands on the basis of selling capacity – usually as a transponder lease – to various military organizations or units around the world. The first such project is known as X-TAR. This system is backed by the US-based Space Systems Loral Corporation and a consortium of investment companies in Spain, and the capacity from the two X-TAR satellites is being leased to US, European, and South American defense and emergency communications organizations. The specific characteristics of these types of defense-related satellite systems, which offer these services on the basis of transponder leases to a number of countries on a regional basis, are also described in further detail in a later chapter.

Evolution of New Digital Services and Applications

As described earlier, the evolution of communications satellite services over the past few decades has been marked by several key trends. One of the most important trends has been the continued subdivision of communication satellite markets into a growing number of service categories. Today there are fixed satellite services (FSS); land, maritime, and aeronautical mobile satellite services (MSS); broadcast satellite services (BSS), direct broadcast satellite (DBS) services, and direct to home (DTH) services. There are also store and forward satellite services and a wide range of military, defense-related, and emergency satellite services that range across all of the above service categories. Some satellite service providers even provide a hybrid type of service such as a combined store and forward messaging services plus a space navigational service to support shipping, trucking, bus, and railway customers. The key to the ability to provide all of these different types of satellite communications services efficiently, reliably, and cost-effectively is the new digital services that have largely transplanted analog services around the world.

The use of digital modulation, digital encoding, and digital multiple access techniques, and especially time division multiple access (TDMA) and code division multiple access (CDMA) coupled with new highly efficient coding concept such as "Turbo-coding," has allowed satellite systems to grow and become more reliable and cost-effective. The greatest challenge to the growth of satellite services in the digital age of the Internet has been adapting to the widespread application of the Internet Protocol (IP) across the planet. The transmission delay associated with connecting between the Earth and geosynchronous orbit created a special problem for networks operating on the Internet Protocol. This is because the original design of IP networks was based on terrestrial networks and perceived delays were registered as "network congestion" rather than transmission delay. This led to a "slow recovery" process that undermined the efficiency of satellite transmission. Even satellites in medium earth orbit (MEO) had this sort of problem. The development of IP over Satellite (IPoS) standards has led to new efficiencies in satellite transmission. Another problem for IP-based satellite services has involved the creation of virtual private networks (VPNs) within business enterprise networks. The main problem in this regard that IP Security (IP Sec) strips off "headers" as a VPN is created. This also requires an innovative solution for VPNs delivered via satellite. The optimization of satellite transmission to operate efficiently within IP networks and to accommodate corporate VPNs is described in section “Satellite Communications Services as Defined by the ITU.”

Limits to the Growth of Satellite Networks

The development of satellite communications services and applications over many decades since the start of such services in 1965 has been impressive. The most important stimuli to growth have been diversification. The field of satellite communications has expanded by constantly finding an ever-wider circle of applications. This began with just fixed satellite services, but services quickly diversified into such areas as video and audio broadcasting, maritime, aeronautical and land mobile satellite services, and even store and forward and supervisory control and data acquisition (SCADA) applications. Many of these applications today are “hidden” from the consumer since the satellite network is remotely orbiting thousands of kilometers out in space. When a consumer makes a purchase at a gas station, a grocery store, or retail outlet, a credit card is most likely validated via satellite connection. Most television signals, even those received via cable television, most likely originated or traveled by one or more satellites at some point in their transmission.

The second most important source of growth has been the prodigious expansion of digital satellite communications services and the adaptation and optimization of satellite for IP-based transmission. Many countries connect to the worldwide Internet via backbone or trunk satellite transmissions. Increasingly satellites can support high-speed broadband Internet connections to the small office/home office (SOHO), especially where fiber, cable, or terrestrial broadband wireless is not available.

Despite the growth of satellite communications services, there are limits to this expansion as terrestrial networks (fiber, coaxial cable, or broadband terrestrial wireless systems) are installed across the planet. This suggests two paths forward. One path is the increased integration of satellites with terrestrial systems. The other is the off-world use of space communications for interconnection with scientific satellites and in time even colonies in space.

The integration of satellite and terrestrial systems is seen in various ways. Fiber-optic systems and satellite networks have often been planned in tandem and satellites have been used as backup to terrestrial cable systems in case of emergency outages or natural disasters. In large and geographically diverse countries, satellites have frequently been used to cover large and thinly populated areas such as mountain ranges, deserts, and forested areas or wetlands in conjunction with terrestrial cable systems. In recent years, the new trend is to use satellites in conjunction with terrestrial broadband wireless systems. There are now combined satellite and land mobile systems that provide fully integrated wireless broadband Fourth Generation (4G) service in the United States and North America. These systems and their current financial status were described above. The official generic name for this service, as defined by the US Federal Communications Commission (FCC), is land mobile satellite service with “ancillary terrestrial component” (ATC). In Europe, such type of integrated satellite and terrestrial wireless services is known as mobile satellite service with “complementary ground component” (CGC).

Another similar effort to provide integrated satellite and terrestrial broadband services for the developing world is system being planned under the name O3b. This unconventional name stands for the “Other Three Billion” people who largely live in the equatorial regions of the planet and have limited access to potable water, health care, educational services, electricity, and communications. The origin of this idea is that the density of broadband Internet traffic in continents such as Africa and large parts of South America and Asia still do not yet justify the installation of broad band terrestrial fiber or coaxial cable systems. Instead, however, a broadband wireless terrestrial systems linked to an IP-optimized satellite system would be a way to provide service to the underserved areas of these continents. Indeed, the logic that initially suggested that this might be a viable solution for Africa led to the further idea of a global satellite system optimized for the developing countries largely concentrated in the equatorial band of the planet. This new type of satellite system could be deployed to meet the traffic requirements of the three billion plus people in developing countries that would like to have economical and reliable access to broadband terrestrial wireless.

Although the O3b system is intended to be optimized for this service, other satellite systems such as Intelsat, SES Global, Asiasat, etc. can also support this type of rural connectivity architecture. Consumers with broadband wireless Internet access could then connect via satellite to national and even global service – especially if the service was truly low cost and suited to the affordable pricing realities of developing economies. The O3b system is currently in planning and capital financing is still being put in place. The economic difficulties that have been experienced by satellite systems geared to significant dependence on developing country markets, such as the “Worldspace” audio broadcasting system, the initial Iridium and Globalstar systems, however does suggest that such types of satellite systems will face significant economic challenges. Nevertheless, the concept of combining satellite wide area coverage with terrestrial wireless networks in town and city areas has a great deal of logic behind it, in terms of serving regions with a lower density of broadband IP-based traffic.

The future of satellite services and applications does seem to be constrained by major parameters and yet stimulated by specific opportunities suggested by satellite technology with its broad areas of coverage.

Satellites thus seem to have particular service opportunities for broadcasting and multi-casting services (both audio and video coverage) for a good time to come. This is the largest source of revenues to commercial satellite services today with well over 70 % of revenues coming from this source. Satellites, for similar reasons of broad coverage at economic rates, are well suited for large-scale corporate networks, often referred to as enterprise networks. Since these networks tend to be very dynamic with nodes often being added or subtracted all over a country, a region or indeed the world, satellite links remain quite well suited to such services. Satellites likewise remain very well suited to mobile communications or communications on the move, especially where the density of traffic is low or traffic needs suddenly appear such as in the case of an emergency, natural disaster, or area of armed conflict.

The reverse condition works in favor of terrestrial fiber-optic or coaxial cable. In areas where there is a high density of users within a developed economy, the installation of such terrestrial infrastructure tends to occur. Also, fiber-optic submarine cables are also installed between high-density international telecommunications traffic routes. In these cases, satellite services tend to provide backup capability in case of cable breaks or natural disasters. The new land mobile satellite systems (with ATC or CGC) represent an excellent example of how satellite communications design engineers have adapted to these broad trends related to the strengths and weaknesses of satellites and terrestrial telecommunications networks to design systems that are optimized to both forms of technologies.

Finally, the other major trend in commercial communication satellite services known as "technology inversion" is also expected to continue. This trend is the consumer mandate to develop, for the individual user, ever more compact ground communications units that require less operating power, are lower in cost, and easier to operate. As the number of users of satellites has expanded from thousands to millions to billions, the economics of satellite communications have changed dramatically. The volume of users has allowed large investments in powerful and sophisticated satellites that allow user transceivers to shrink from huge 30 m earth stations with large operating crews to quite small hand-held units. Although there might be some small satellites for experimental purposes, to support ham radio relay, or message relay, the main commercial trends will continue to support mass consumer needs and very small and low cost user terminals and transceivers.

The truly longer-term future of satellite telecommunications services will relate to the need for truly long distance communications through space to the Moon and lunar colonies and eventually even to other planets and beyond. Today the various space agencies, particularly NASA (USA), ESA (Europe), JAXA (Japan), CNSA (China), Roscosmos (Russia), CSA (Canada), and ISRO (India) have had the need to create space communications systems to support exploratory missions. NASA has developed and expanded a very sophisticated Deep Space Network (DSN) for decades to receive signals from scientific missions to the Moon and beyond. The important element to note is that technology developed for the demanding requirements of sending and receiving signals over the vast distances of space have often led to technical or service innovations that can be used to improve commercial communications satellite systems here on Earth. The three axis body stabilized satellites that are now in common use around the world, that provide higher levels of pointing accuracy and better solar illumination of solar cell arrays, were first developed at the US Jet Propulsion Labs in terms of improved communications requirements associated with planetary missions. In coming years, commercial satellite communications systems may indeed evolve to provide interplanetary and cislunar communications services.

Conclusion

The commercial satellite communications market has proven to be very dynamic and new services and applications have diversified and grown as the field has matured over the decades. Digital communications technology and Internet Protocol-based services have helped to accelerate this diversification. The dynamic interrelation of satellite technology to terrestrial telecommunications overtime has been another key factor that has influenced both the size of commercial satellite markets and the shape of the market in terms of services offered. Coaxial cable systems, fiber-optic networks, and most recently terrestrial broadband wireless systems have both limited the growth of commercial satellite communications markets and services and also defined new opportunities for satellite systems to complement terrestrial telecommunications systems.

Satellites continue to have particular market opportunities in terms of broadcasting and multi-casting services, offerings to rural and remote areas, to island countries and developing countries – particularly when the terrain or topology of these countries create significant barriers to deployment of conventional terrestrial telecommunications systems. In short, countries with a large number of islands (i.e., Indonesia, the Philippines, or Micronesia), countries with significant mountainous terrains (i.e., Chile or Nepal), countries with extensive jungles (Brazil, the Republic of Congo, Malaysia, and Thailand), or countries with major deserts (i.e., Algeria, Libya, or Mauritania) will find satellite technology and services well adapted to their needs. The ability of new, more powerful satellites to operate to smaller and smaller user terminals of lower cost and greater mobility will also continue to stimulate the growth and diversification of commercial communications satellite markets. Finally, the next frontier for satellite telecommunications services in the decades ahead will relate to communications across the solar system and beyond as human exploration scientific studies and practical space applications extend further and further beyond Planet Earth. For further readings related to materials covered in this section, please consider the materials covered in the endnotes.[3]

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Economics and Financing of Communications Satellites
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ History of Satellite Communications
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Satellite Communications Video Markets: Dynamics and Trends
- ▶ Satellite Orbits for Communications Satellites

---

[3] For further reading concerning satellite telecommunication services, See J.N. Pelton, A. Bukley (eds.), *The Farthest Shore: A twenty-first Century Guide to Space* (Apogee Books, Burlington, Canada, 2010). Particularly Chapter Six on satellite applications. D.K. Sachdev, *Business Strategies for Satellite Systems* (Artech House, Boston, 2004). R. Jakhu, *National Regulation of Space Activities*. Space Regulation Library Series (Springer, Dordrecht, 2010). And J.N. Pelton, *The Basics of Satellite Telecommunications* (Professional Education International, Chicago, IL, 2006).

References

A global overview of Direct to Home Television services. http://www.electronics.ca/publications/products/Direct%252dto%252dHome-(DTH)-Satellite-Television-Services-%252d-A-Global-Strategic-Business-Report.htm

Intelsat Organization, *Article XIV of The Agreement of the Intelsat Organization, As it Entered into Force as of Feb. 12* (Intelsat, Washington, DC, 1973)

ITU Frequency Allocation Table. http://www.itu.int/ITU-D/tech/spectrum-management./SMS4DC_AM_TM_4.pdf

ITU Radio Regulations, Article 1, Definition of Radio Service, Section 1.21. http://www.ictregulationtoolkit.org/enpracticeNote.aspx?id=2824

ITU Radio Regulations, Article 1, Definition of Radio Service, Section 1.25. http://www.ictregulationtoolkit.org/enpracticeNote.aspx?id=2824

ITU Radio Regulations, Article 1, Definition of Radio Service, Section 1. XX. http://www.ictregulationtoolkit.org/enpracticeNote.aspx?id=2824

Orbcomm, *Orbcomm Signs Next Generation Satellite Constellation Contract for 18 Satellites* (Orbcomm, Fort Lee, 2008)

J.N. Pelton, *The Satellite Revolution: The Shift to Direct Consumer Access and Mass Markets* (International Engineering Consortium, Chicago, 1998)

The Boeing Company Manufactures and Deploys the Spaceway Satellites. www.boeing.com/defense-space/spaceway/spaceway.html "Direct TV's Spaceway F1 Satellite Launches New Era in High Definition Programming". www.comspacewatch.com/news/viewpr.html?pid=16748

The International Telecommunication Union. www.itu.int/net/about/index/aspx

World Radio Conference, of the International Telecommunication Union, Final Acts of the 2007 Conference. www.itu.int/ITU-R/go/index.asp?wrc-07/en

Satellite Orbits for Communications Satellites

Joseph N. Pelton

Contents

Abstract

One of the key elements of a communications satellite service is the ability to launch satellites into precisely defined orbits and to maintain them in the desired orbit throughout the lifetime of the satellite. The system control and oversight of satellite orbits both require not only the technical ability to launch and maintain the orbit, but the ability to attain the proper legal authority, at the national and international level, to transmit and/or receive radio signals from these orbits. This regulatory process means a number of specific steps associated with registering for the allocated frequencies from the International Telecommunication Union

---

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_5

(ITU)through a national governmental administration, obtaining assignments of those frequencies in the required orbits in accord with national licensing procedures, and coordination of the use of the specific frequencies through intersystem coordination procedures.

There are a wide range of different orbits that are currently used in communication satellite services although the most common are geosynchronous Earth orbits (GEO), medium Earth orbits (MEO), and low Earth orbits (LEO). This chapter explains the various orbits that can be used and the advantages and disadvantages of each of the orbits most often employed for satellite communications. This analysis indicates some of the primary "trade-offs" that are used by satellite system engineers in seeking to optimize a satellite systems performance both in its design and subsequently over its operational lifetime. The activities involved in selecting an orbit; designing and achieving an operational satellite network; and optimizing its technical, operational, and financial performance over the systems lifetime involve a wide range of issues. These start with selecting a desired orbital framework, obtaining authorization for orbital access (including the registering and precoordination of the satellite and its orbit with other systems), launch, deployment and test, systems operation, and end-of-life disposal of a satellite from its orbit.

Keywords

Antenna gain • Antenna pointing • Command and control of satellites • Figure-8 orbit • Geostationary earth orbit • Geosynchronous satellite orbit • Inclined orbit • Loopus orbit • Low earth orbit • Medium earth orbit • Molniya orbit • Omni antennas • Polar orbit • Quasi-Zenith orbit • Radio astronomy • Satellite constellations • Space weather • String of pearls orbit • Sunspot activity • Sun-synchronous orbit • Supersynchronous satellite orbit • Tracking of satellites • Van Allen belts

Introduction

Sir Isaac Newton first discovered the laws of gravitational attraction and created understanding of the planets revolving around the Sun and why satellites revolve around the planets. Newton even recognized that it would be possible to "shoot" artificial satellites into orbit around the Earth or other planetary bodies. His universal law of gravitational attraction is still fundamental to understanding basic orbital mechanics. This law is expressed as follows:

$$F_g = GMm/r^2$$

The above formula expresses Newton's universal law of gravitation and shows how to calculate the force of attraction between the Earth and another object. This mutual attraction (although the Earth's pull is obviously very much greater) is determined by taking the mass of the Earth (M) and then knowing the distance between the center of the Earth's mass and the center of gravity of the orbiting mass. This shows that the gravitational pull decreases as the object moves further from Earth. Indeed, this mutual attraction decreased by the square of the distance it is away from the Earth's center. In order for an object to be in Earth orbit, it must have sufficient velocity or centripetal force to overcome the gravitational pull (Pelton 2006).

In the early experimental days of satellite communications, there was a wide range of opinion about what types of orbits might be used most effectively for telecommunications services. There were a number of perceived advantages and disadvantages envisioned for different orbital configurations. Once the experimental Syncom 2 and 3 satellites established the viability of launching and operating a satellite in geosynchronous (GEO) orbit, however, the worldwide practice – in a short period of time – concentrated on this orbit. This was because of several factors such as the simplicity of not having the Earth station systems to track the satellite and the high gains of those types of antennas over those with low-gain omni antennas that could receive signals across the full open horizon. Also the fact that satellites in GEO could cover over one-third of the world's surface was a strong economic factor (Pelton 1974).

Over time, other applications rekindled interest in other orbits. These factors included interest in mobile satellite systems, defense satellite systems with the need for communications on the move, store and forward applications, and the desire to achieve low latency (or less delay in satellite transmission time). These elements and more served to revive interest in other types of orbits.

This is not to say that GEO presented the only orbital option. Special conditions such as the northern latitudes of the Russian landscape allowed a special highly elliptical orbit, known as the Molniya orbit, to be utilized (Pelton 1974, p. 55). The limitation of available radio frequency allocations and the crowding of the geosynchronous orbit that over time served to move the "spacing" of communications satellites in GEO orbit closer and close together affected the technology of the satellites and Earth stations as well as the active consideration of different orbits. There is, however, a technical coordination difficulty when satellite systems that use the same allocation of frequencies for similar services attempt to use disparate orbital constellations.

For instance, when constellations of satellites in low and medium Earth orbits and using the FSS frequencies cross over the equatorial plane, they can cause substantial interference to the geosynchronous satellites that utilize the same radio spectrum bands. The new large-scale MegaLEO systems, such as OneWeb, have developed a way to repoint their LEO satellites on a temporary basis in order to reduce interference as they cross the equatorial belt to minimize interference to satellites in the GEO orbit. Further LEO and MEO satellites with high-powered beams can also cause significant interference with radio astronomy – particularly when they are using frequencies for mobile satellite services (MSS).

The following discussion presents the various types of orbits and their uses and applications including the advantages and disadvantages that are involved from a technical, operational, financial, and regulatory perspective. This is followed by consideration of particular issues that are involved with Earth station design, technical coordination between different types of orbital configurations, as well as terrestrial and other services (such as radio telescope services, terrestrial microwave, and high altitude platform systems (HAPS), and finally disposal of satellites at end of life). Most of the issues involving technical coordination and registration of radio frequencies for satellites are addressed in the chapter that explains ITU allocation procedures, but the issues particular to different orbital configurations are addressed here.

Key technical concerns that accompany the selection of satellite orbits include the extent of their geographic Earth coverage and the so-called path loss that is determined by how far the satellites are away from the Earth's surface. Another major concern, however, is the problem of destructive radiation that can disable or even completely end the useful life of communications satellites. The Van Allen belts contain intense radiation which actually helps to shield the Earth from radiation from the Sun or the stars but these "structured belts" around the Earth can be destructive to satellites that must fly through them. In addition, radiation from the Sun, so-called space weather, is also a concern to maintaining satellites effectively in orbit. Indeed, during intense "sunspot activity" or solar storms, satellites are typically shut down to prevent failure to satellite electronics (Charles et al. 2009). Further, satellites in orbit are also subject to cosmic radiation and are especially vulnerable to the solar wind and to the most intense radiation from the Sun that comes with occasional solar storms that follow a multiyear cycle that peaks during what is called the Solar Max period.

Different Orbital Configurations for Different Communications Satellite Services

There are a large number of orbits that can be used for satellite communications and an infinite number of constellations that can be created using multiple satellites. The most common orbits are geosynchronous or geostationary orbit (often called Clarke orbits in honor of Arthur C. Clarke), medium Earth orbits, and low Earth orbits.

Deployment of satellites into these various orbits require progressively more energy as they are positioned further away from Earth since they are continually overcoming the Earth's gravitational pull as they ascend to higher orbits. Also highly specialized orbits, such as a GEO orbit, that require the satellite to be placed into a circular plane above the equator and at great height require greater orientation and positioning capability as well as the need for nearly constant station-keeping maneuvering. In the simplest terms, low Earth orbits that involve "direct insertion" are the easiest to attain. Polar orbits and medium Earth orbits are the next easiest to achieve. GEO orbits are the most difficult. At the outset of the satellite industry, satellites were deployed into very highly elliptical orbits with the apogee at the desired height of 35,870 km (22,230 miles). After this orbit was firmly established, something called an "apogee kick motor" (AKM) was fired at the appropriate apogee in order to push the satellite from transfer orbit into a perfectly circular orbit around the equator at the desired longitude.[1]

Today's rocket systems with advanced propulsion systems can insert the satellite into the final GEO orbit without the hazard of firing a solid motor rocket. In the future, ion engines might achieve GEO orbit in entirely different ways by continuously firing "electronic thrusters" that slowly spirals the satellites outward from Earth to achieve the desired circular GEO orbit after weeks of firing these much lower energy thrusters. The subject of rocket launchers and orbital deployment is discussed later in this handbook.

The following discussion describes the characteristics of the various orbits utilized and their various strengths and weaknesses for various types of communications services.

Geosynchronous or Geostationary Satellite Orbits

The most common orbit for satellites providing fixed satellite services and broadcast satellite services and quite a few mobile satellite services are those that are called the geosynchronous, geostationary, or Clarke satellite orbit. This is a unique orbit where the orbital velocity is sufficient to maintain the satellite in this circular equatorial path with the centripetal force away from Earth that exactly overcomes the pull of gravity at this altitude. The "g" force or gravitational pull at this orbit is approximately one-fiftieth (1/50th) that experienced at the Earth surface. This is to say that at a distance of 22,230 miles or 35,870 km the accelerative pull of the Earth's gravity is (0.22 m/s^2 rather than 9.8 m/s^2). What makes this orbit so special is that the orbital velocity that creates the angular momentum (and thus the centripetal force) needed to overcome the pull of gravity just happens to constitute the exact speed needed to complete a revolution around the world exactly every 23 h and 56 min and 4 s. The "odd missing 4 min" of a 24 h day represents rather exactly the 1/365th of the time the Earth uses to revolve around the Sun. In short, in celestial (or sidereal) time, a spacecraft in GEO Earth orbit revolves exactly once around the world every day. It thus appears as if it were indeed a very, very tall tower in the sky with the satellite at the top of the imaginary tower (Pelton and Madry 2009).

This special orbit identified in 1928 by Herman Potočnik (who wrote under the German pseudonym Hermann Nordung) as a location for an inhabited space station, and more famously identified by Sir Arthur Clarke as a location a "geostationary communications satellite" has been in continuous use by artificial spacecraft since 1965.[2] There is always a difference between theory and practice. The geostationary orbit of theory would keep a satellite moving west to east exactly with the rotation of the Earth as it remains stable exactly about the Equator. There are gravitational affects of the Moon and Sun that tend to tug a satellite in this orbit to move North

---

[1] (Pelton 2006, pp. 73–87).

[2] Hermann Nordung, the Slovenia Scientist, www.astronautix.com/astros/noordung.htm.

above the equatorial plane and the tug it back South of the equatorial plane. It turns out that the "station keeping" required to keep a spacecraft exactly in geostationary orbit in terms of North and South migrations demands approximately ten times more energy (i.e., firing of station-keeping jets) than maintaining the East–West stabilization. In short, a spacecraft to be maintained within its assigned "GEO orbit box" – in terms of its desired longitude (East–West location) and its desired "0" degree latitude – requires active station keeping. The excursions North and South both above and below the equatorial plane look like a very low amplitude sine wave. These are the most difficult to control and require by far the most fuel use in the firing of thrusters.

The buildup of the excursions to the North and South of the equator are called the inclination of the satellite orbit. Most "GEO satellites" tend to move perhaps a degree or so North and South of the equator each day. As long as the spacecraft does not move more than plus or minus a very few degrees above the equatorial plane, this does not create more than a very minor problem for an Earth station on the ground pointing to the satellite receiving the signal. The slight variation in the gain for the Earth stations on the ground are most pronounced in the equatorial regions, but since the locations at the so-called subsatellite point receive the strongest signal (i.e., the least path loss), this is not a particular problem. In practice, therefore, satellites in "GEO" orbit are thus more or less "geosynchronous" but not really "geostationary" because of these small excursions off the equatorial plane. Toward the end of life, satellite operators tend to let "GEO" satellites build up their inclination (i.e., movement North and South of the equator) because this saves station-keeping fuel and allows the extension of the satellite's practical lifetime (Williamson 1990; Pelton 2006, p. 76). The ITU that maintains the global registry of satellites and their location recognizes the registration up to 5° inclination above and below the equatorial plane. (Note this is of significance in terms of problems of interference between GEO satellites and those in low and medium Earth orbits when operating in the same frequency bands.)

A final note on orbits is that concerning eclipses of satellite in different orbits. Satellites in polar or near polar orbit in low Earth sun-synchronous orbits will in a 90 min orbit be behind the Earth and shielded from the Sun for about 35 min for each revolution. Batteries must provide power for this part of the orbit. In the case of geosynchronous satellites, the issue of eclipses represents a more complicated issue. The Earth is "tilted" 23.5° on its axis and as such during the solstice times a GEO satellite and its solar cell arrays are fully illuminated. GEO satellites either "see the sun" from over the North Pole or under the South Pole during the winter and summer months. Some 22 days before the equinox period, however, GEO satellites will experience a small eclipse that builds up to a maximum of some 70 min a day during the spring and fall equinoxes. These eclipses then dissipate in the same manner until 22 days after the equinox. The issue of an eclipse from the shadow of the Earth is again addressed by satellite operators by the use of batteries during these two periods of semiannual darkness. Since GEO satellites have nearly 300 days a year of total illumination, the issue of eclipse is much greater for LEO systems where a satellite can be in eclipse over a third of the time. Satellites in LEO constellations often use

Table 1 Advantages and disadvantages of the GEO orbit

Advantages and disadvantages of GEO satellites for communications services (Pelton 2001)	
Advantages	Disadvantages
Three satellites in GEO orbit provide essentially global coverage except for the polar regions. This means that global coverage can be achieved at a lesser cost than MEO satellites (10–18 satellites for global coverage) or LEO satellites (48–60 satellites for global coverage). This is because the closer a satellite is to the Earth the less the satellite is able to “see” of the Earth below. Even one GEO launch can create a full-service capability for a region, while with MEO and LEO satellites a full constellation must be in place to create a fully functional system	The satellites in this orbit are almost one-tenth of the way to the Moon and thus there is a very large path loss between the satellite and ground antennas. Since path loss (i.e., diminished signal) is a function of the square of the distance, the satellite is away from the Earth. This is a substantial factor in satellite design and the ability to “close a link” between a GEO satellite and the Earth
A satellite in GEO orbit allows continuous connection with high gain Earth stations without constant tracking of the satellites. This allows for a simpler and less expensive antenna design. Or it requires the ground antennas for LEO or MEO satellites to be much lower gain devices that are essentially “omni-devices” (i.e., ones that can receive a signal from all different directions)	The great distance the spacecraft orbits away from Earth creates delay or latency in the transmission. This latency is on the order of a quarter of a second for the entire pathway from the Earth to the Satellite and the return. This creates problems for telephone communications and in Internet connections. At a low elevation angle from the antenna to the spacecraft coupled with a low elevation angle for the return transmission, the path can be over 75,000 miles or 120,000 km
A GEO satellite with large high-gain antennas can create spot beams continuously pointed to desired geographic locations on the Earth	Inter Satellite Links (ISLs) between GEO satellites are much harder to establish and require much higher capability, power, etc., than is the case with LEO or MEO satellites
A GEO satellite is relatively easy to maintain in orbit and can sustain in-orbit lifetimes of 15–20 years which is longer than medium Earth orbit (MEO) and especially longer than low Earth orbit (LEO) satellites	To use a GEO satellite for mobile satellite service requires a very high power and huge aperture multibeam satellite to connect with a simple hand-held user terminal with reasonable reliability. (This is largely a function of significant path loss)
A GEO satellite can easily have its orbit raised out of GEO orbit at the end of life. This is much easier to accomplish than spending a great deal of fuel (40 % of all fuel) to de-orbit a MEO or to de-orbit a LEO satellite that creates various types of risks to other satellites	Each satellite tends to be larger, more complex, has longer production schedules, is more expensive to launch and insure, and allows less economies of scale than MEO or LEO satellites. This is usually more than offset by the economies achieved by the need to launch many fewer of this type satellite to complete a full system

the time when they are over polar regions or over oceans where traffic demand is low for time to recharge batteries (Table 1).

Medium Earth Orbit (MEO)

A medium Earth orbit (MEO) satellite constellation can be configured in many different ways to achieve global coverage. The main constraint that impacts the planning of a MEO system is to launch the system so that the satellites are essentially flying above the Van Allen belts. As noted earlier, the radiation in the Van Allen belts contains very high-speed particles such as high-energy neutrons. This radiation can do damage to satellite electronics and even with spacecraft shielding of the electronics and glass coating on solar cells, the lifetime of the spacecraft will be significantly shortened if it must fly within the Van Allen belts.

Although a number of communication satellite systems have been proposed that would utilize global MEO constellations, such as the original ICO and the Odyssey systems, these were never deployed. There have also been concepts for using MEO systems for high capacity broadband systems for the Ka-band but such a system for a variety of reasons was never deployed. Currently the system known as O3b (for the "other three billion" people) contemplates using a MEO system to support a high-speed Internet service to broadband wireless users in developing countries with a focus on the equatorial countries of the world.

It is possible to operate just one or two satellites in MEO orbit for store and forward services or machine-to-machine (M2M) connectivity.

In many ways, a MEO constellation provides a compromise between the advantages of a GEO system on one hand and a LEO constellation on the other (Table 2).

Table 2 Advantages and disadvantages of the MEO orbit

Advantages and disadvantages of MEO satellites for communications services (Pelton 2001, pp. 228–230)	
Advantages	Disadvantages
A MEO constellation provides global coverage and with significantly less path loss and transmission delay than a GEO system	More satellites than a GEO system must be purchased and launched to create a global network
A MEO constellation can achieve global coverage with the launch of as few as ten satellites and the tracking, telemetry, and command (TT&C) system needed to support the system is much less than a LEO constellation	Full system must be deployed and completely checked out to operate network, unlike a GEO satellite that can operate as a complete network by itself with wide regional coverage
Inter Satellite Links (ISLs) and mission control are much easier to accomplish than with a LEO system, but more difficult than with a GEO system	There have been very few MEO constellations deployed for communications services and thus there is limited experience with the operation of these systems; optimizing the construction of satellites for MEO operation; or knowledge about special design aspects such as radiation shielding, etc.
These orbits provide a very good trade-off between total number of satellites, complexity of system control and TT&C requirements, requirement for satellite on-orbit "spares," wide area geographic coverage, power of spot beams and geographic coverage, and reasonable path loss and transmission delay	De-orbiting a MEO satellite requires a great amount of thruster fuel and this recreates a cost disadvantage and adversely affects the lifetime of the satellites and of the overall constellation

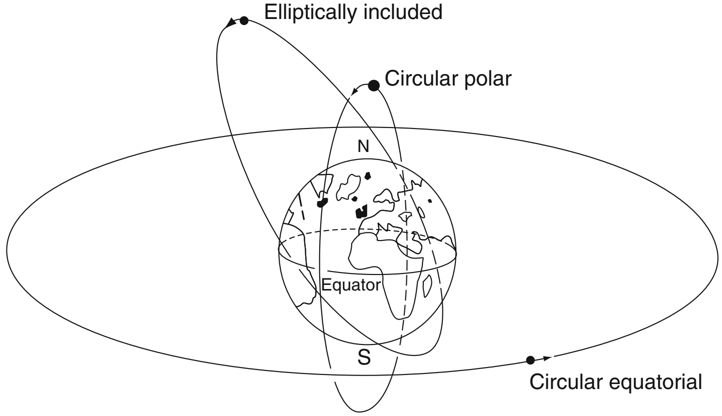


Fig. 1 A “cartoon” depiction of a LEO, MEO, and GEO orbit (Graphic courtesy of the Author)

Low Earth Orbit (LEO)

The orbit of choice for a number of land mobile satellite systems in the past decade or so has been the LEO Constellation. The Iridium and Globalstar systems have deployed and operate LEO Constellations for mobile communications services for nearly 15 years. Further a significant number of store and forward satellite networks have been launched over the years. These have included, among others, the commercial Orbcomm system, the Surrey Space Centre and Utah State University satellites and the Oscar (Amateur Radio) small satellites that have been launched going back many years. LEO constellations represent the opposite extreme from GEO satellites with their strengths being GEO satellites weakness and vice versa.

Indeed the advantage of being close to the Earth and thus allowing transmissions to experience less path loss is also a disadvantage because many more satellites are needed to achieve global coverage. Figure 1 shows, in cartoon fashion, the geometrical profiles of GEO, MEO, and LEO orbits with respect to the Earth. This figure is not to scale since a GEO orbit can, in fact, be some 40 times further out in space than a LEO orbit and this cannot be easily shown to exact dimensions.

The following table presents the relative pros and cons of a LEO satellite constellation (Table 3).

Various Types of Communications Satellite Constellations

There are literally an infinite number of constellation designs that can be devised for low and medium Earth orbit satellite systems. In order to design the optimized constellation, there are a number of key threshold questions that a satellite system operator will typically address. Selection of a particular parameter to be optimized in a satellite system orbital configuration will likely dictate the number of satellites

Table 3 Advantages and disadvantages of the LEO orbit

Advantages and disadvantages of LEO satellites for communications services (Pelton 2001, pp. 228–230)

Advantages	Disadvantages
Low Earth orbit (LEO) satellites are up to 40 times closer to the Earth's surface and thus experience up to 1,600 times less transmission path loss	There is a need for a large number of satellites to complete a global constellation network (i.e., 50 satellites and up). This increases systems costs for a global system because there are many more operational and spare satellites, many more launches, and a more complex TT& C network for system control
Low Earth orbit satellites experience up to 40 times less latency or transmission delay than GEO satellites. This is simply because a LEO satellite orbit is 20–40 times closer to the Earth's surface than a GEO satellite	The system requires more difficult overall system controls, complex billing and authentication systems, and network implementation, including more active spares and system restoration procedures. This can in part be overcome by installing Inter Satellite Links (ISLs) on all satellites, but this also increases costs and satellite complexity
LEO satellites, because they fly more directly overhead and cover the lower and higher latitudes more effectively than GEO satellites, typically will have lower "masking angles" to user receivers and particularly provide more effective coverage at upper latitudes and can even provide service to the polar regions	One cannot use high gain ground antennas constantly pointed toward the satellites because the spacecraft is rapidly moving across the sky with only a few minutes of visibility before moving below the horizon and thus needing to be replaced by another satellite in the constellation
LEO constellations are particularly well suited to mobile satellite services because of the lower path loss, lower masking angles, concentrated beam coverage, modest transmission delay, and the desire to provide users with lightweight, compact, and low-cost antennas with small, relatively low-gain antennas	The satellites are being attracted much more strongly by the Earth's gravitational field and there is more fuel needed for station keeping and thus the lifetime of the LEO networks are less. The operational lifetime of LEO satellites are typically about 7 years. GEO satellite lifetimes can be 12–18 years
Orbital designs for LEO constellations can be adjusted to concentrate coverage at lower latitudes (from 0° to 70° North and South). (Unfortunately, necessary coverage of all longitudes provides coverage of the Atlantic, Pacific, and Indian Oceans where there are limited customers.)	There is a much higher probability of LEO satellites being hit and partially or completely disabled by space junk because satellites and space debris are more closely spaced
Detailed computer programs can be designed to make LEO systems "smarter" and "dynamically flexible." This means that satellites can be programmed to increase or decrease power or performance in specific beams in specific locations. Increased power can be used for ringtones or message-waiting signals	LEO constellations, because they often cover the entire Earth's surface with cellular-like beams and most often are utilized for mobile satellite services (and thus associated frequency bands), tend to have a more significant problem of coordination with radio astronomy services. Also user transceivers can interfere with one another

deployed, determine the maximum and minimum elevation angles for the overhead satellites, and indicate the feasibility of intersatellite links.

The Iridium land mobile satellite system constellation design was heavily dependent on the concept that this global network would provide intersatellite links or cross-links among and between all four of the closest satellites in the network. This led to the decision to have the satellites in polar orbits (nearly 90° inclination to the Equator) so that the system would be highly symmetrical and cross-links easily established among the two closest satellites North and South and the two closest satellites East and West. Other elements in the trade-offs in the constellation design were satellite power versus number of satellites in the constellation and orbital elevation versus typical elevation angles to the nearest overhead satellite in the constellation.

In contrast, the Globalstar satellite constellation decided not to have ISLs or cross-links and decided instead to have all LEO orbits to have less than 70° inclination North and South so as to concentrate satellite "overhead coverage" to the populated regions of the world and to avoid the polar regions. This approach provided a better look angle for everyone below 65° elevation North and South and simplified the number of TT&C stations that had to be put in place for system control.

The Orbcomm store and forward (i.e., machine-to-machine [M2M]) system chose a low Earth orbit constellation design that was able to minimize satellite size, power, and manufacturing and launch costs. Yet the Orbcomm system contained enough spacecraft in the constellation to complete global data messaging within a very few minutes.

The once proposed and now defunct Teledesic "Mega LEO" satellite system for broadband Internet services opted for a design with an exceeding large number of satellites to be deployed (originally over 800 satellites plus a huge number of spares). This design was conceived so as to insure a very high elevation angle to support instant high data rate broadband communications via very narrow and high gain pencil spot beams. These features of many satellites in a low Earth constellation to support minimal transmission delay, very high elevation (or so-called masking) angles, plus high power transmissions from all orbital spacecraft were unique aspects for this proposed system. This was because, unlike the systems for mobile satellite services that are designed to operate in the radio frequency range around 2 GHz, the Teledesic system was intended to operate in the Extremely High Frequency (EHF) or Ka-band frequencies (with a 30 GHz uplink and a 20 GHz downlink). Satellites that operate in these frequency bands require a direct or uninterrupted line-of-sight connection between the satellite and ground antenna systems to complete a transmission link.

There are certain important similarities between the Teledesic system and the more recently conceived "other three billion" (O3b) satellite system whose designers have opted for a very high-powered medium Earth orbit constellation design with its spacecraft orbiting some 8,000 km above the Earth. The original constellation will consist of eight satellites but will expand to eventually include up to 20 satellites to populate the full MEO constellation.[3]

Molniya, Highly Elliptical Orbits (HEOs), Extremely Elliptical Orbits (EEOs), and Loopus Orbits

There are a family of orbits that are variously described as Molniya orbits, highly elliptical orbits (HEOs), extremely elliptical orbits (EEOs), and “Loopus” orbits. These orbits all have in common the following elements – a very high apogee and a low perigee orbit. In the most extreme configuration, the shape of these orbits can be thought of as being “cigar shaped.” The advantage of this type of orbit is that it can have a very long effective “hang” time especially about high latitude countries such as Russia where this type of orbit was first used. In particular, this Russian system employed the Molniya orbit. This orbit had a 12 h period with 8 h of the orbit being above the horizon in the Northern latitudes above the Russia subcontinent. This meant that three satellites placed in three separate Molniya orbits could be deployed like the petals on a flower to provide continuous service to the entire country throughout a 24 h day. Also the very high elliptical orbit with the very, very high apogee meant that the satellite did not “seem to move” as it ascended and then descended along a very narrow track in the sky.

In recent years as the geosynchronous orbit became more and more populated with satellites providing various communications services, the concept of using HEOs or EEOs once again became an attractive idea. The Sirius Radio broadcasting satellite system was initially deployed in this type of extremely elliptical orbit and several broadcasting satellite systems have been proposed for this type of orbit. A very particular type of HEO or EEO orbit is the so-called Loopus orbit that is depicted in the graphic below. Long duration visibility is available in Northern latitudes for this special orbit during positions 1, 3, 5, 7, 8, and 9. This type of orbit can be utilized for fixed and broadcast applications and Earth stations require only limited pointing capabilities (Fig. 2).

The main application for these types of orbits is thus essentially for broadcast types of services. This renewed interest in these types of orbits is fostered by the fact that it is no longer easily available to obtain new orbital slots within the GEO (or Clarke orbit). The long periods over which a satellite “appears” to be in the same location can thus serve to emulate a satellite in GEO orbit. Nevertheless, there is a need to have at least three and probably four satellites populating this type of orbit since they typically will only maintain this “apparent” location during their “highest apogee phase” for a period of 6–8 h.

---

[3] “Agreement signed with Arianespace for Initial O3b Satellite Launches” O3b Networks http://www.o3bnetworks.com/Media_Centre/press_release_details.aspx?id=60.

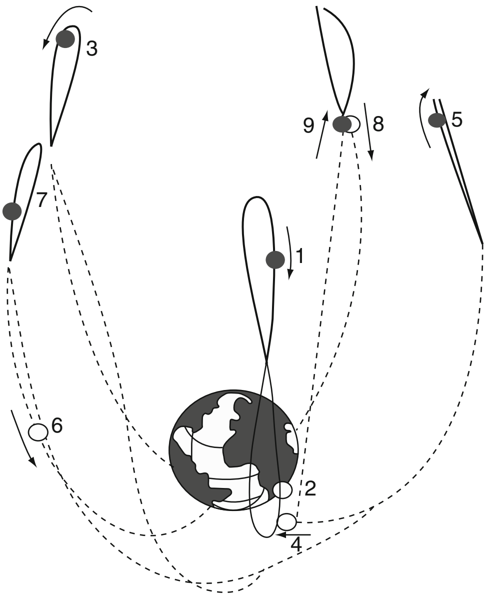


Fig. 2 The "Loopus orbit" shown in its movements relative to the Earth's rotation during a 24 h period (Graphic courtesy of the Author)

String of Pearls Orbit

Another orbital concept that has been considered by a variety of different system planners over time is the so-called string of pearls orbit. The concept involves the deploying of a number of satellites, such as six to eight spacecraft, in a medium Earth orbit around the equator. The concept here is that the satellites would be equipped with zonal beams that would cover more than one-sixth of the Earth's circumference (for the six satellite configuration) or more than one-eighth of the Earth's circumference (for the eight satellite configuration) so that as one satellite moved below the horizon a new zonal beam from the "ascending spacecraft" would provide an equivalent coverage. One would need to have a significant overlap of coverage to provide a seamless handoff between the "departing" and "arriving" satellite so that the handoff would be entirely seamless and so that the ground antenna systems would not need to track a particular satellite and so that the quality of the signal would not be significantly degraded during the handoff process.

The value of this particular orbital configuration is that there can be continuous coverage to the entire equatorial region (i.e., 3–4,000 km above and below the equatorial belt) where some 2.5–3 billion of the world's population is concentrated.

Part of the time, a particular satellite might be over Brazil, or Columbia, or Ecuador, or Peru, or the Congo, or Kenya, or Uganda, or Indonesia, or India, or Southern China, or Laos, or Thailand. Much of the time, however, each and every one of the six to eight satellites would be over heavily populated areas. This is simply a result of the world's land mass geography. Such a satellite network would be well suited to providing domestic services to the various countries of the equatorial region. It would not, however, be well suited to providing international services since it provides limited interconnectivity between equatorial countries and even less with the rest of the world. There are also practical difficulties with how such a system would be financed and a logical system devised for paying for the derived services actually used. Some countries such as Brazil, India, Indonesia, and China might use such a system heavily, but other countries such as Sri Lanka or Guyana might find it useful to a much smaller degree than much larger nations. Brazil had once thought of deploying such a system to meet its own domestic needs and offer the facility as a "gift" to other equatorial countries of the world.

Quasi-Zenith or Figure-8 Orbit

Another orbit that is now being utilized for mobile satellite communications is that which can be variously described as "Quasi-Zenith" or the "Figure-8 orbit." This is essentially a GEO orbit that is inclined 45° to the equatorial plane and then populated by three or more satellites so that a country near 45° latitude such as Japan (near 45° North latitude) or Australia (near 45° South latitude) always has one satellite "overhead" with a steep look angle to the subsatellite point. Japan was one of the countries to first identify this type of orbit to utilize for mobile satellite communications. Japan has a number of cities with very tall skyscrapers such as Tokyo, Osaka, or Yokohama, and thus satellite communications to user transceivers can be easily blocked by buildings that rise high into the sky. The Quasi-Zenith or Figure-8 orbit provides one satellite overhead with something like an 80° look angle down into the cities and thus a much better look angle than a GEO satellite. One of the unique features of this orbit is that satellites create a pattern that appears like the figure 8 with the top of the orbit being at 45° North and the bottom being at 45° South. Thus, a system deployed over the Pacific Ocean would create excellent coverage for Japan in the North while equally providing coverage with the other satellites in the orbit for Australia in the South. Japan has designed an experimental Quasi-Zenith satellite, named Michibiki (see Fig. 3). The Michibiki experimental satellite was designed by the National Institute for Communications and Information Technologies, fabricated by the Mitsubishi Electric Corporation (MELCO), and launched by the Japanese Space Agency (JAXA). The purpose of the experimental satellite is to test out this type of orbit for use for space navigation as well as for mobile communications satellite coverage.[4]

If an operational satellite system were to be deployed, then some form of commercial arrangement would logically be created to provide coverage not only in the Northern latitudes but in the South as well. In this instance, it could be used not only for mobile satellite communications but for space navigation as well.

---

[4] The Quasi-Zenith Satellite System is a project of the National Institute for Information and Communications Technologies (NICT) in cooperation with the Japanese Space Exploration Agency (JAXA) http://www.spacecom.nict.go.jp/control/efsat/index-e.htm.

Fig. 3 Michibiki experimental satellite to test Quasi-Zenith orbit for mobile satellite communications services and space navigation (Graphic courtesy of JAXA)

Because of the high altitude represented by the QZSS orbit, it can indeed easily also support a space navigation function. It is for this reason Japanese experimenters decided to equip these satellites with atomic clocks and the ability to transmit space navigation beacons. Satellites equipped in this manner can thus be effectively used to augment the existing Global Positioning Satellite network and thus to provide augmented space-based reference points to allow more accurate data for navigation and mapping. Because the GPS availability enhancement signals transmitted from Quasi-Zenith satellites are compatible with modernized GPS signals, interoperability is ensured. The Michibiki satellite not only has a highly accurate atomic clock but will also be able to transmit the L1C/A, L1C, L2C, and L5 signals that are compatible with the GPS space navigation system (The Quasi-Zenith Satellite System is a project of the National Institute for Information and Communications Technologies (NICT) in cooperation with the Japanese Space Exploration Agency (JAXA) http://www.spacecom.nict.go.jp/control/efsat/index-e.htm).

Supersynchronous

The above discussed orbits are the main ones used for satellite communications around the world today. The satellites are easier to track, command, and control the closer they are to Earth. The same is true of space debris that is concentrated in these

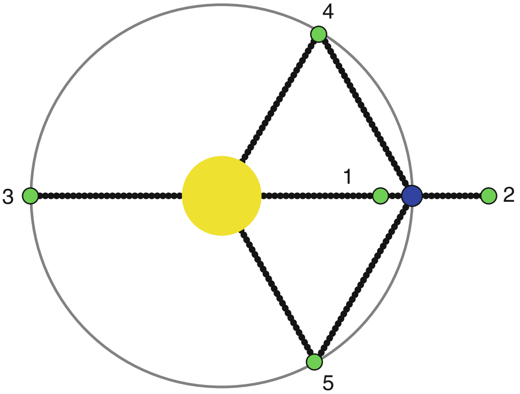


Fig. 4 The five Lagrangian points as shown in relation to the Sun's and the Earth's orbit

various orbits as well. It would be possible to deploy a satellite in orbits that are farther away from the Earth than Geosynchronous orbits for communications or other purposes. These reasons that one might do this would be to avoid detection, such as for military or defense-related purposes, or to establish a relay point for communications to the Moon, Mars, asteroids, or scientific satellites. One of the often discussed such locations are the so-called Lagrangian Points that exist in space as discovered in 1772 by Joseph Louis Lagrange, the Italian–French mathematician. These are relatively "stable" locations where a satellite once located in these positions are trapped in these orbital positions by the competing gravitational effects of the Earth and the Sun (and to a lesser extent the Moon and the other planets). Thus, there are five such points as shown in the following diagram. It has been suggested the L-1 Lagrangian Point could be used as a suitable point for observation of the Earth's atmosphere (i.e., the Triana Project by NASA to observe the Ozone Hole and other atmospheric phenomena), a space colony, or a translunar communications link between the Earth and the Moon. It has also been suggested that other Lagrangian points such as 3, 4, and/or 5 might be used as communications relay positions for broadband communications satellites to provide links between Mars and Earth (Fig. 4).

These points are highly desirable for very long-term satellite positioning since once a satellite or space colony reaches one of these locations it will remain there indefinitely – trapped by the gravitation of the Sun and Earth.

Earth Station and User Terminal Design for Different Orbits

The different orbits described above require different types of ground antennas to operate. The different types of satellite Earth station antennas and terminals are addressed in detail in later chapters. There are some basic concepts that are important to note with regard to antenna designs as they relate to different types of orbits used for satellite communications and particularly with regard to how antenna designs can be optimized. The GEO orbit allows high performance or high-gain antennas to be exactly pointed toward a satellite with a minimum of “tracking.” Thus a large (or even a small) dish (i.e., a parabolically-shaped) antenna can continuously point toward a GEO satellite over head. This is because a Geo orbiting satellite seems to hovers above the same point 24 hours a day; since these type of satellites revolve in synch with the Earth’s rotation. This is in contrast to satellites in MEO and LEO orbits that require tracking antennas on the ground to follow these type satellites as they move from horizon to horizon. Even in the case where a GEO satellite is building up inclination North and South of the equator, a relatively simple mechanical system can be added to the antenna steering system to move in tandem with these small migrations North and South during a 24 h period that is highly predictable. Ground antennas working to GEO satellites for fixed satellite services (FSS) have a higher sensitivity because they can point a focused beam constantly at a “stable satellite.” This means that satellites providing FSS can have smaller antennas in space and lower power because the ground antennas have the ability to send a more focused beam to the satellite and receive a more focused beam from the satellite without a high-speed tracking. This is in contrast to the lower gain end user antennas that typically work to medium Earth orbit or low Earth orbit satellite constellations. These lower gain user antennas are likely to be omni antennas that can capture a signal from any direction or squinted beam omni antennas that can capture signal from anywhere above the Earth’s horizon. Thus, these are ground antenna systems that are designed to capture signals from a satellite that moves rapidly across the horizon.

As in most cases, there are exceptions to the rule. There are especially desired and more expensive higher gain antennas designed to support tracking, telemetry, control, and monitoring functions that have high-speed tracking capabilities that can support the operation of LEO and MEO orbit satellites. These antennas have large aperture dishes but that are also able to track even a low Earth orbit (LEO) satellite that can cross from horizon to horizon in a few seconds. Such ground antenna installations are quite expensive and thus are built and operated only to support the safe operation of satellite networks. These antennas are too complex and expensive to be utilized by actual satellite system users. The broad band consumer users, particularly those equipped with hand-held transceivers or microterminals to support mobile satellite services (MSS), have simple omni or near omni antennas or quite small antennas with limited tracking capabilities.

The other exception comes when one seeks to use a GEO orbit satellite to support MSS type operations. The higher gain antennas for FSS markets or for direct broadcast services work quite well when the satellite is stable and the ground station antenna is stable. If you attempt to support mobile satellite services from a GEO orbit, there is immediately a problem: the satellite is stable, but the user terminals are typically moving. The users and their ground antennas could be moving through a forest, an underpass, or through the middle of a city. In this case, you do not have the satellite antenna constantly pointed toward a higher gain dish antenna on the ground. This means that you are forced to design much high-powered and larger antennas on the GEO satellite to compensate for the much smaller and lower performance "omni" or "near omni" type antenna that the user on the ground must utilize.

The key to designing and engineering a successful satellite network involves what is called "systems engineering" or "system optimization." One must have a sufficient level of power and a focused degree of transmitted radio frequency signal to "close a link" between a satellite and a ground receiver (on the ground, the sea, or in the air). The calculation of the antenna gain (or size) and the power (on the ground and on the satellite) is called a "link budget." Additional power and antenna gain above the minimum needed to complete a "link budget" is called "link margin."

If you were to design a system – a "cable in the sky," where there is only one single high-capacity pathway – let us say between New York and London, then you could afford, within your system engineering, to have two very large, high-gain antennas that send their signals back and forth between a simple satellite with relatively low power and small spacecraft antennas. For such a cable in the sky this might result in the lowest overall systems cost.

If one takes the opposite extreme and wanted to design a system to send a direct broadcast television signal to every home in Europe, the system optimization process would be dramatically different. Now, one would want a very high-powered satellite indeed to send a signal to millions of very small, low-cost, and easy-to-install antennas on the ground. This in many ways accurately describes the process of system engineering that has characterized the development of satellites over the last 30 years. This is sometimes called "technology inversion." This is the evolution from very small and low-powered satellites that worked to quite large, powerful, and expensive Earth stations to the reverse situation where there is a very large and powerful satellite that distributes signals to low-cost user terminals.

This means that within the process of "technology inversion" the number of users on the satellites on the ground has grown from dozens, to hundreds, to thousands, to now millions. The investment of a large amount of money in increasingly powerful satellites with very sophisticated and large in-orbit antenna systems makes economic sense if this allows the overall cost of the entire system to go down. If one takes the example of a direct broadcast satellite, the logic goes as follows. Even if it costs many hundreds of millions of dollars to build and launch a high-power DBS satellite, this still becomes economic if it can reduce the individual cost of "millions" of consumer antennas on the ground to hundreds of dollars rather than thousands or tens of thousands of dollars. This is because there is just one satellite or one satellite and spare, but there are a huge number of user antennas on the ground. If one can reduce their costs by just $100 and there are ten million users, the cost for the ground part of the overall system goes down by $1 billion.

Satellite system engineers actually spend a lot of time trying to figure out how much money will be spent on the satellites manufacture, launch, and satellite operations on one hand versus how much money will be spent on the other hand by the consumers on ground antenna systems. In most satellite systems today, the bulk of money will be spent on consumer-based antennas to receive television signals, mobile satellite communications services, or high-speed broadband data services or telephone circuits. This is simply because there will be so many consumer users – typically numbering in the hundreds of thousands or even millions. Systems engineers then try to design an "optimized" total system that represents the overall lowest cost system. Sometimes they get it quite right and the satellite network is successful and attracts the projected number of users and the lowest possible cost system. In other cases they get in wrong and the satellite, the ground antennas, or some other aspect of the system is badly designed for the market and the system fails and goes bankrupt. Examples of where the projected number of users on the ground failed to be achieved were the Globalstar, Orbcomm, and Iridium satellites that subsequently entered into bankruptcy.

Relative Economics of Different Satellite Orbits

Most satellite planners and systems engineers start out by considering the service that is desired to be provided and the type of orbital configuration that can best provide the desired service at the lowest net overall cost and with the highest level of reliability and service quality. As can be seen in the graphic below, one can cover more than a third of the Earth from GEO (or Clarke orbit). One can reasonably cover the Earth with about eight satellites in medium Earth orbit (MEO), but when the satellites get very close to the Earth, in low Earth orbit, it turns out that one needs 40 or more satellites to provide total coverage of the globe. One only needs to shine a flashlight on a round balloon or basketball at varying distances to see why the satellite coverage capability varies as one nears the Earth (Fig. 5).

The problem is that the trade-offs between and among satellite and satellite launch costs, ground antenna costs, service quality, and tracking, telemetry, command, and monitoring costs and other costs such as marketing, advertising, billing, and regulatory services are not easy to project before a system is designed and deployed. Sometimes, when the service is almost entirely new, the ability to project market and consumer acceptance can be dramatically off – as was the case of the three satellite systems first conceived to provide land mobile satellite services around the world. These three initial systems, namely, Iridium, Globalstar, and ICO all ended up in bankruptcy without it clearly being established as to exactly what went wrong. Possible explanations include that the new market was over estimated, the cost of the system too high, the wrong type of orbital configuration was chosen, or the satellite service design or ground antenna unit for the consumer were not well

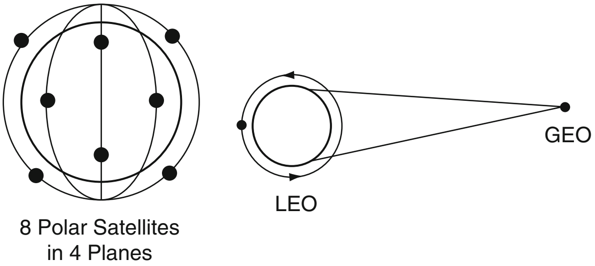


Fig. 5 Different Earth coverage by satellites at different orbital heights (Graphic courtesy of the Author)

matched to the market demand. What is clear is that GEO-based satellite systems that require only a few satellites to begin operations and collect revenues are often lower risk business propositions.

Some analysts suggest that it is a “very steep climb” to seek to deploy a large-scale LEO constellation system at the outset of a new service. In short, a LEO satellite system requires a long lead time and a very large investment to build and launch many dozens of satellites. This large expenditure becomes particularly difficult as a start-up business because there is no established revenue stream. In short, there is enormous challenge in designing, building, deploying, and testing a large LEO constellation with no incoming revenues or established market base. Certainly the bankruptcies of the Iridium System with 72 satellites plus spares, the Globalstar system of 48 satellite plus spares, and the Orbcomm satellite network of 48 satellites plus spares indeed all seem to constitute a strong caution against deploying LEO systems as a totally new start-up business. The Teledesic satellite system that originally envisioned deploying some 840 satellites plus 80 spares was the ultimate example of new LEO system that required the launch of way too much hardware prior to the realization of any revenues against a huge capital debt. In the case of ICO and Teledesic, these projects folded before actual satellite launches began.

What is clear from an economic sense is that a GEO satellite system can be initiated with a single satellite in orbit. Indeed one can lease one or more transponder from an existing GEO satellite system and increase capacity as markets and revenues grow. In short, GEO systems allow the strongest case for organic growth of satellite services for localized, national, or regional services and in many cases the ground antenna systems can be shifted from leased satellite capacity to a dedicated satellite network as market demand grows. A medium Earth orbit (MEO) system can be started with far fewer satellites and perhaps as few as six to eight satellites, although many MEO constellations do require a larger number of satellites. From an economic standpoint, the MEO constellation, in many ways, represents a compromise between a GEO and a LEO system in terms of number of satellites to manufacture and launch, and complexity of TT&C operations. The actual design and implementation of a communications satellite system, however, involves far more than the orbital configuration and the desired national, regional, or global coverage that a satellite system provides to a specific market.

These factors must take into account service requirements such as transmission latency, reliability, quality/bit error rate, coverage, and look angles; design, performance, and cost of user antennas/terminals; operational cost and complexity (including TTC&M design and costs); as well as overall cost efficiencies, capital financing, regulatory constraints, and strategic business case.[5] Nevertheless, one of the key starting elements in any satellite system design will typically be the orbital configuration to be utilized. This thought process will often start with the feasibility of obtaining access to one or more GEO satellite locations or the lease of capacity on an existing satellite network. MEO or LEO constellation designs thus represent a “step

---

[5] Op cit. (Pelton 2001, pp. 1–31).

beyond" in terms of pursuing a business plan that will typically involve an element of greater technical, financial, and regulatory risk. These risks will in many cases be considered to be acceptable in exchange for improvements in high service quality (i.e., lower transmission delay and lower path loss); ability to attain access to orbits and allocated frequencies that may be available; or improvements in user antenna compactness, complexity, and cost.

Conclusion

The ability to attain access to allocated radio frequencies to operate a satellite system continues to be an ever more challenging activity. The difficulty grows as more and more satellites are launched into Earth orbit and very few new opportunities exist for satellite system operators without engaging of closer and closer spacing of GEO satellites, or the use of the quite demanding frequencies in the millimeter wave bands, or possibly opting to deploy satellites into orbital configurations beyond the most "conventional choice" of the GEO orbit. The challenges of opting for other orbital configurations have actually spurred the trend toward closer and closer spacing of satellites in the GEO orbit and the implementation of GEO satellites that utilize the higher frequency bands. Today, the problem of frequency coordination has become even more difficult than ever before. This is simply because there are more and more satellites that are operating at higher and higher power levels and spaced ever more closely together.

In addition, the problem of orbital debris has increasingly emerged as a problem for LEO, MEO, and GEO, and polar orbits and indeed for the general sustainability of all space efforts near Earth in the future. Efforts to coordinate among the various operators of satellite networks to minimize the possibility of collisions among spacecraft are also being intensified through coordinative efforts. This has resulted in organizations reducing frequency interference (now known as simply the Satellite Interference Reduction Group (sIRG)) and also led to the creation of the Space Date Association that has created a coordinated global data base that monitors the orbits of various satellite systems such as those of Intelsat, SES, Inmarsat, and Eutelsat so as to allow avoidance techniques to be followed in case of impending satellite collisions. It is hoped that this initiative will expand to include more and more operators and will include more and more orbits.

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ History of Satellite Communications
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Satellite Communications Video Markets: Dynamics and Trends
- ▶ Space Telecommunications Services and Applications

References

C. Charles, F. Ciovanni, F. Lauren, G. James, M. Mikhai, M.K. Chris, R. Michael, S. Isabelle, The Universe and Us Chapter 5, in *The Farthest Shore: A 21st Century Guide to Space*, ed. by J.N. Pelton, A. Bukley (Apogee Books, Burlington, 2009), p. 157

J.N. Pelton, *Global Communications Satellite Policy: Intelsat, Politics and Functionalism* (Lomond Books, Mt. Airy, 1974), p. 48

J.N. Pelton, *Research Report: Satellite Communications – The Transition to Mass Consumer Markets, Technologies and Networks* (International Engineering Consortium, Chicago, 2001), pp. 228–230

J.N. Pelton, *Basics of Satellite Communications* (International Engineering Consortium, Chicago, 2006), p. 73

J.N. Pelton, S. Madry, Satellites in service to humanity, in *The Farthest Shore: A 21st Century Guide to Space*, ed. by J.N. Pelton, A. Bukley (Apogee Books, Burlington, 2009), p. 220

M. Williamson, *The Communications Satellite* (IOP Publishing, Bristol, 1990), pp. 76–83

Fixed Satellite Communications: Market Dynamics and Trends

Peter Marshall and Joseph N. Pelton

Contents

Abstract

The history of fixed satellite services (FSS) systems, in terms of technological and institutional development, has been previously provided in chapter “▶ History of Satellite Communications” of this handbook to a very large extent. Thus, this chapter addresses the market trends related to FSS systems and also discusses how a variety of new types of satellite services has evolved out of the initial FSS networks over time.

The market dynamics and trends of FSS systems are particularly addressed in terms of four main factors: (1) the competitive impact of high-efficiency fiber-

---

P. Marshall
Royal Television Society, London, UK
e-mail: pmsatellites@aol.com; pmsatellites@btinternet.com

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: peltonjoe@gmail.com; joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_6

optic terrestrial and submarine cable communications networks; (2) the conversion of FSS systems from analogue to digital services that allowed FSS systems to be more cost-efficient and use spectrum more efficiently as well as migrate to spectrum in higher bands more effectively; (3) the move of FSS systems toward deployment of smaller and lower cost ground systems (variously called VSATs, VSAAs, USATs, and microterminals) that allowed services to migrate closer to the "edge" of telecommunication user networks (i.e., satellite services directly to end user facilities); and (4) a shift in regulatory policy that allows FSS systems to compete directly for services that has generally served to reduce cost and spur innovations in services and applications.

These four trends have combined to contribute to what has been previously described in chapter "▶ History of Satellite Communications" as "technology inversion." This "technology inversion" has thus seen FSS systems in space become larger, more complex, longer-lived, and more powerful as ground systems have become more user-friendly, lower in cost, and are designed to interface directly with users at localized office facilities or even small office/home office (SoHo) VSATs or microterminals. These technological, regulatory, and market-based trends have shaped the FSS networks and related market dynamics. All four of these trends have dramatically reshaped the nature of FSS services for both commercial markets and defense-related satellite networks around the world.

The historical trend in FSS markets has been the initial development of global networks since global connectivity was the highest value market and the most underserved by terrestrial telecommunications networks available in the 1960s. Over time, satellite technology matured and the economical viability of regional and domestic satellite systems evolved in the years that followed. Today there are some 300 FSS satellites, essentially all in GEO orbit where these systems provide a complex combination of global, regional, and domestic satellite services. Although broadcast satellite services have outstripped FSS in terms of market value and sales, the FSS is still a very large and growing multibillion dollar industry.

Keywords

Analogue to digital conversion • Bit error rate • C-band • Digital satellite services • Domestic satellite systems • Fixed satellite services • Frequency bands of satellite service • International Telecommunication Union (ITU) • Internet protocol over satellite (IPoS) • Ka-band • Ku-band • Microterminal • Quality of service (QoS) • Regional satellite systems • Satellite markets • Spectrum allocations • Spectrum efficiency • Submarine cable systems • Ultrasmall aperture terminal (USAT) • Very small aperture antenna (VSAA) • Very small aperture terminal (VSAT)

Introduction

This chapter notes how this first type of communications satellite service was defined by the International Telecommunication Union as fixed satellite service (FSS). With the maturation of satellite technology over the years that followed, the development of lower cost and easier to use ground systems, together with regulatory shifts, allowed the further development of direct broadcast satellite services, mobile satellite services, and even store and forward data relay or machine-to-machine type services. FSS services, as the oldest and most mature of the satellite services, is the father and in some cases the grandfather of all the various satellite communication services that have followed since the start of commercial services in the 1960s. Both mobile satellite services, which evolved in the 1970s, and direct broadcast satellite services that date from the 1980s have benefited from the initial technology first developed for commercial FSS systems (Chartrand 2004).

The development of these additional services as well as defense-related satellite services is discussed in detail in chapters "▶ Satellite Communications Video Markets: Dynamics and Trends," "▶ Mobile Satellite Communications Markets: Dynamics and Trends," "▶ Store-and-Forward and Data Relay Satellite Communications Services," "▶ Broadband High-Throughput Satellites," "▶ Distributed Internet-Optimized Services via Satellite Constellations," and "▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications."

The key market dynamics for FSS are discussed in this chapter in terms of six predominant trends that can be concisely stated as: (1) evolution of service capabilities and related competition from terrestrial communications systems; (2) digital satellite communications and the move to higher frequency bands; (3) decentralization of FSS services as small ground systems move to the "edge" of global networks; (4) regulatory shifts with regard to FSS systems to make them openly competitive around the world; (5) the shift of FSS systems from primarily serving global markets to more and more satellite networks serving regional and domestic markets; and (6) key new trends in satellite system design that are rapidly changing the traditional forms of communications satellite system services and economics. These key interrelated trends are discussed and analyzed in terms of their impact on the FSS markets.

The pattern for FSS markets was for networks designed for global services to evolve first because this was the highest value type service. Regional and domestic FSS systems followed thereafter. This was a logical consequence as satellite technology matured and market demand allowed these regional and domestic systems to become economic around the world, particularly as lower cost satellite antennas and terminals became available. The development of military and defense-related traffic has represented yet another dimension of the market for FSS networks around the world. These market trends and dynamics are addressed separately in chapter "▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications."

Evolution of FSS Services and Competition from Terrestrial Communications Systems

The evolution of fixed satellite services (FSS) in the earliest days of satellite communications was largely the history of the Intelsat satellite system in the period from 1965 through the early 1970s. The first Intelsat satellite, known as "Early

Bird," was essentially a "cable in the sky" that could only connect point-to-point service. Then came the Intelsat II series which was able to provide multidestination service and connect several points at once. This satellite was designed and built with US government funding to support the US manned space program Gemini so that ships at sea could maintain communications with the crew in the space capsule. The Intelsat III series that was launched in the 1968/1969 time period were the first commercial satellites to provide a full range of satellite services similar to today's satellites in terms of providing multidestination services to many points with the capability to provide voice, data, color television, and high-quality radio channels.

It was this Intelsat III series that in July 1969 was able to provide global coverage of the Moon landing by Apollo 11. It was only a few weeks before the Moon landing that true global connectivity via satellite was established. As of 1970, satellite communications had become the predominant form of international communications as this technology provided broader band and lower cost connectivity than the coaxial submarine cables of the day. Further, multidestination satellites were able to connect any country in the world to a globally interconnected network by constructing and operating only a few Earth stations. As the first Director General of the Intelsat and former head of Entel Chile once said:

Communications satellites changed almost everything for our country. For the cost of one Boeing 707 airplane, we could build and operate a satellite earth station that could allow Chile to be fully connected to the rest of the world (Interview with Santiago Astrain 1974).

The cost of international telephone calls from remote areas of the world could exceed $50 a minute prior to the advent of satellite communications. However, since the arrival of global satellite systems and ever more efficient submarine cable systems, the cost of an international call has dropped to a level that is little different from the cost of a long distance call within a country. Prior to the advent of satellite communications, the global delivery of live television was simply not possible. Today over 18,000 video channels are available worldwide via satellite connections (Pelton 2006).

For over a century, there has been an ongoing rivalry between terrestrial submarine cables and wireless communications systems to provide better, lower cost, and more reliable communications for overseas communications. In the middle of the nineteenth century, telegraph submarine cable systems began to provide limited international communications service. These cables had limited capacity and were subject to disruptions and failures due to storms, trolling fish vessels, and other factors. The invention of shortwave radio provided a way to provide overseas voice and data services at lower cost and with greater throughput capability. However, shortwave radio was subject to disruptions as the result of space weather interference with the ionosphere. The invention of coaxial cable systems capable of carrying voice traffic in the 1940s and 1950s moved international voice and data traffic back toward terrestrial technology. The resulting submarine cable systems, even with 3 KHz telephone channel spacing and the so-called time assignment speech interpolation (TASI), still had very limited capacities of only 72 voice circuits in the mid- to late 1950s. The advent of satellites such as the Intelsat I with 240 voice circuits in

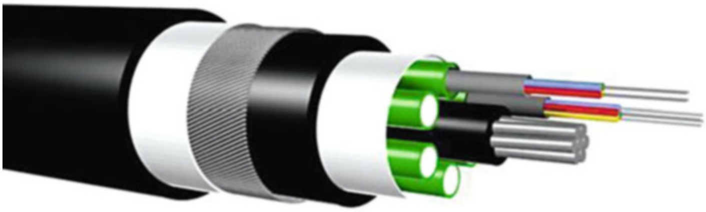


Fig. 1 Fiber-optic submarine cables have become predominant for the heaviest transoceanic routes

1965 and then Intelsat III with 1,200 voice circuits plus two color television channels sharply shifted international telecommunications traffic to satellite connections. Satellite circuits were lower in cost and allowed much more voice and data traffic to be provided between the continents and enabled international television transmissions to be provided, both technically and economically (Pelton and Alper 1986).

Beginning in the 1980s with the advent of new fiber-optic submarine cable technology, the international telecommunications market shifted focus once again. Fiber-optic submarine cables became more and more cost-effective, broadband, and higher quality and thus quickly began to reclaim international telecommunications services, at least on the heaviest transoceanic links (see Fig. 1 above). This shift from satellite telephone and data back to submarine cables, particularly for trans-Atlantic and trans-Pacific Ocean traffic through the 1990s and up to the present time, was hastened by several factors:

- *Quality of service*: Transmission via fiber-optic submarine cables, as measured in bit error rate, could be very low and typically could be in the range of only 10^{-10} or even 10^{-12}. This was an unprecedented level of transmission quality. Further transmission times were typically less than 100 ms as opposed to the 250 ms of transmission delay associated with geosynchronous satellite transmission. This shorter latency or transmission time made fiber the preferred choice for telephone service.
- *Cost of service*: The very heaviest routes, such as between the United States and Europe, could be considerably lower than the costs associated with international satellite connections. Satellites remained cost competitive for more remote locations with thinner routes of traffic or to locations not served directly via fiber-optic networks. Satellites also remained cost competitive for television distribution services.
- *Structure of service provision and ownership*: Submarine cable services were provided as if they were actually owned and capitalized by telecommunications service providers under what were called "indefeasible rights of use" (IRUs) that made provision of service more cost-effective and profitable under current regulatory policies then in effect.

The cost efficiencies of both fiber-optic submarine cables and satellites have continued to plummet as both of these technologies have matured. The improvement of the technology, the extension of the lifetimes of these systems and more have now been so dramatic that the "capital cost" of a single voice circuit might be as low as $5 per voice circuit on a submarine cable and under $50 a voice circuit on an advanced communications satellite.

The economics are such that other costs associated with international telecommunications, such as marketing and sales, advertising, billing, and operations, now tend to be predominant over the creation of the international link itself. Thus, issues such as quality of service, lack of transmission delay, redundancy of service links, network design and complexity, and the ability to establish links to particular locations with great speed often tend to dominate the decision as to whether or not a link is established via satellite or submarine cable. In general, it can be said that most heavy route traffic between countries or even within countries today are carried by fiber-optic networks. Satellite communications networks thus tend to carry medium to thin route voice or data traffic to supplement fiber-optic networks and a variety of different types of television services where distribution to widely distributed audiences of business networks may be involved.[1]

The need to create integrated global telecommunications networks to serve the "enterprise networks" of multinational enterprises, national governments, international organizations, and military systems has seen a growing trend toward forming combined and seamless networks. These combine fiber and coaxial fiber networks, broadband terrestrial wireless networks, and satellite systems under unified ownership. This is, in part, the result of the growth of Internet, intranets, virtual private networks, and digital networks that provide broadband to support voice, data, video, and audio services on demand. The digital satellite revolution and the provision of voice and other services over IP are discussed immediately below.

Digital Satellite Communications and the Move to Higher Frequency Bands

The provision of satellite services for the initial two decades was essentially via analogue-based services. Analogue services and multiplexing systems using frequency division multiple access (FDMA) were inefficient in several ways. The amount of information that was sent via satellite was inefficient in terms of information transmitted per Hertz (or information sent per cycle per second). Also, there were just a limited number of carriers of set size for everything from small routes to very large routes. The information throughput density was progressively lower for smaller and smaller carriers for thin routes of traffic because of the need to separate the carriers with guard bands and because the carriers were only efficient when completely filled with actual active voice traffic. Once a carrier was filled with traffic, however, there was a need to jump to a larger fixed carrier size to accommodate growth. In all of these ways, the analogue satellite service was inefficient. In the

---

[1] Op cit, Chartrand, pp. 9–20.

1980s and 1990s, there was a digital revolution in satellite communications and most space traffic was converted from analogue transmission using FDMA multiplexing to either time division multiple access (TDMA), code division multiple access (CDMA), or a special system developed for very thin routes of traffic known as the SPADE system that allowed single channels to be used on the satellite on a demand-assigned basis.

Some ways by which digital satellite communications can be considered superior to analogue satellite service include the following:

- Greater ability to operate at higher transmission speeds
- Improved quality of service – especially in a high noise environment
- Greater compatibility with terrestrial digital switches – that now predominate
- Greater compatibility with digital fiber-optic systems
- Easier to allow accommodation of encryption/decryption systems
- Easier accommodation of digital signal compression techniques
- Easier accommodation to onboard digital switching and onboard signal processing to overcome rain attenuation and other forms of interference
- Greater compatibility with all other forms of digital transmission services – coaxial cable, fiber, mobile cellular (4G, LTE, 5G), etc., (Lewis 1988)

In terms of market efficiency, the conversion to digital satellite services allows very high new efficiencies to be achieved. A 72 MHz transponder using analogue technology for high-quality television was typically able to derive two color television channels of reasonably high signal to noise (S/N) quality while operating to very highly sensitive Earth stations of 18 m or larger. In contrast, using digital TDMA or other digital multiplexing technology and MPEG compression, on the order of 14-18 digital television channels could be derived from a 72 MHz transponder while also using smaller antennas to uplink the video signals. The improved throughput for voice channels and data transmission was not as dramatic as was the case for digital television, but there were nevertheless considerable gains.

The gains in efficiencies were approximately four to six times depending on a variety of factors such as the volume of traffic, the size of Earth station antennas, etc. These gains created market disruptions during the transition because the dramatic gains in efficiencies offered by digital services could not easily be reflected in pricing policies without creating a shortfall in revenues.

Also, because the ownership of the satellites and the space segment was divided from the ownership of the Earth stations within the structure of the Intelsat organization, there was a division of interests involved in terms of seeking the benefits from digital satellite services. The owners of the ground stations, especially those with low volumes of traffic, questioned why they should invest in the new digital equipment after having invested in analogue equipment only a short time before. Their position was that the benefits, which would flow from digital efficiencies, went primarily to the largest users and owners of the space segment and not to the smallest users – particularly if they continued to use analogue equipment. Many of the smallest users of the space systems, especially the developing countries, thus had the least incentive to convert to the more efficient digital equipment.

The resulting decision that ensued from this dilemma of what might be called conflicting interests of conversion to the more efficient digital technology was a compromise within the Intelsat Board of Governors. This compromise decision was to phase in the "efficiency pricing" for digital services over a series of years. In short, the plan was to phase in the new pricing for digital services and not to seek to reflect immediately all of the gains achieved by rapid, high-efficiency digital throughput all at once, but to gradually reflect the digital efficiency as TDMA systems were introduced. This was known within the Intelsat organization (the organization that dominated international satellite communications up through the 1980s and was the first to introduce commercial digital services) as the decision in the "spirit of Chang Mai." This was so-named because the Intelsat Board meeting that reached this compromise decision was held in Chang Mai, Thailand, where the local markets were known for their intensive bargaining over price.

In the years that followed, digital conversion continued apace in international, regional, and domestic satellite systems. In many of these systems, networks began with digital systems and thus there was not a question of analogue to digital conversion or the need for a transitional pricing scheme as the switchover occurred.

The competitive processes that were set in motion with the divestiture of AT&T in the United States in 1984 and the liberalization of telecommunications competition within Europe and Japan in the following years helped to speed the conversion to the more efficient digital technology in the form of TDMA and SPADE and subsequently CDMA and spread spectrum services. (This relationship between and among the technology, the market dynamics, and the regulatory process are discussed later in this chapter.)

Ironically, the greater efficiencies of digital satellite services and the reduced cost of service led to a rapid surge in demand. International satellite communications and international submarine transmission capabilities in the 1960s and 1970s were miniscule in comparison to national telecommunications networks. The dramatic decrease in cost for telecommunications and IT systems that occurred in the 1980s, further driven by competition and the spread of multinational enterprises, led to a dramatic increase in demand for international communications. Thus the digital satellite revolution that was thought would create excess satellite system capacity had almost the reverse effect. The net result was that the communications satellite spectrum that had been the mainstay of the satellite industry in the 1960s, 1970s, and 1980s was almost saturated even with the efficiencies that digital communication satellite services engendered. In key locations for geosynchronous satellites, providing for relay over the Atlantic, Pacific, and Indian Oceans, the C-band spectrum was fully consumed.

The new growth of regional and domestic satellite networks further compounded the problem of limited available spectrum for geosynchronous FSS services. The result was to push forward to exploit higher frequencies and also to seek more efficient designs for FSS satellites to allow more reuse of available frequencies. Both solutions were needed to keep up with rapidly growing market demand.

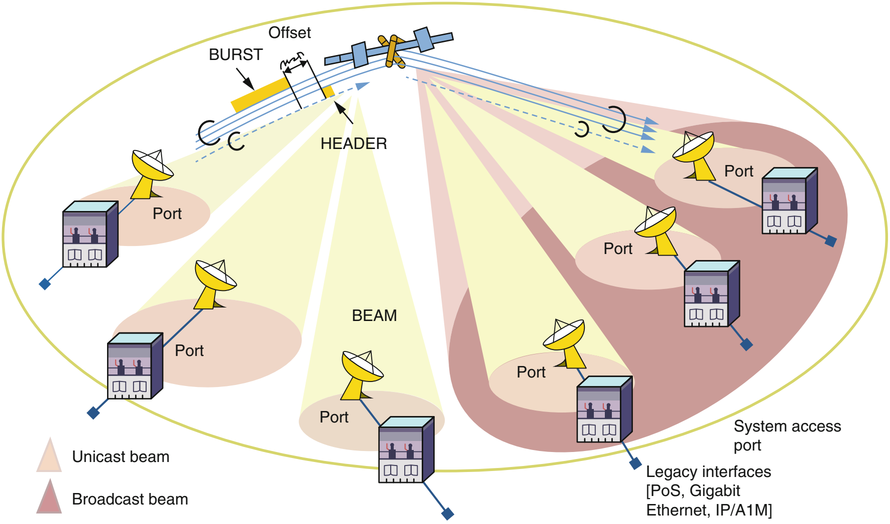


Fig. 2 Multibeam ARTES satellite showing digital spot beams and onboard beam interconnection between beams (Illustration courtesy of ESA)

Technical innovation led to the creation of more efficient designs with FSS satellites deploying many more spot beams that allowed frequency reuse. Spot beams that were sufficiently isolated from one another allowed the same frequencies to be used over again, just as was being done with terrestrial cellular systems. Digital switches on board the satellites allowed these various spot beams to be interconnected. A signal could thus go up to the satellite in a spot beam at one location and then be switched to another spot beam for the downlink connection.

If these spot beams illuminated different parts of the Earth's surface, then the same spectrum could be used without interference. This type of multibeam satellite that uses high-speed onboard spot beam digital switching thus can provide interconnection as illustrated in Fig. 2. This illustration shows the ARTES satellite, which has been developed by the European Space Agency to provide flexible interconnection of many different VSAT ports. This design allows the efficiency of multiple reuse of the same spectrum, and the high-efficiency spot beams can support more rapid throughput within the high-powered beams. These higher-powered spot beams allow smaller aperture antennas to operate to the satellite and also allow for more margin against rain attenuation.

The migration of more and more traffic from the "C"-band spectrum (i.e., 6 GHz uplink and 4 GHz downlink) to the "Ku"-band spectrum (i.e., 14 GHz uplink and 12 GHz downlink) thus accommodated new growth associated with more and more regional and domestic systems and more demand for international services. The Ku-band was in many ways well suited to spot beam operation since the higher frequencies and thus smaller wavelengths were suited to creating higher and higher gain spot beam antennas that could be smaller yet have higher gain just because of the physics of radio waves. A Ku-band antenna could be four times smaller in aperture size but has the same gain as a C-band antenna.

The transition to higher frequencies was not without its difficulties. The new and more demanding higher frequency transmission equipment (on the ground and in space) was more difficult to engineer and build and was thus more expensive. Further, rain attenuation problems that were minimal at C-band increased as one moved up the microwave band to the higher frequencies. The closer the wavelength of radio waves approaches the size of raindrops, the greater the problem of heavy rain acting as a sort of lens to distort the pathway of radio wave transmissions to and from the satellite. Thus, more power margin had to be added to overcome these rain attenuation problems at the higher frequencies.

Most recently, the demand for additional satellite capacity has driven satellite services toward even more powerful and narrow spot beams interconnected by digital switching technology to allow even more frequency reuse. Market demand has also supported the move upward to the still higher “Ka-band” frequencies (i.e., 30 GHz uplink and 20 GHz downlink). The rain attenuation issues associated with “Ku-band” are even more present with “Ka-band” frequencies and the much higher frequency equipment is even more difficult to design and build. Thus, the cost of the Ka-band equipment is still higher than the Ku-band equipment. There is also a need for greater power margins to protect against heavy rainfall (i.e., rain attenuation).

One might ask why not accommodate traffic growth and new market demand by simply allocating new frequencies in lower bands? The problem is that the demand for terrestrial mobile wireless communications has outstripped all other demands for over a decade. There is no realistic hope of new satellite communications allocations for FSS requirements in lower frequency bands. The likelihood of new allocations for FSS services in the microwave band for instance is almost none. This is particularly true since broadcast satellite services (BSS) and mobile satellite services (MSS) are also seeking new allocations as well. The BSS systems, because they provide direct-to-home services to millions of customers, and MSS systems, because also serve millions of customers directly at locations on land, the sea, and the air, are likely to receive priority for obtaining new frequency allocations over FSS systems because of considerations related to rain attenuation and consumer costs.

The bottom line, as noted in more detail above, is that digital services are more efficient than analogue systems in being able to overcome noise and interference. They are certainly better suited to rapid switching of digital traffic between numerous spot beams on the satellite. This factor alone has been critical to the growth of both FSS and MSS satellite systems. Digital satellite systems have also been critical to the effective use of small VSAT and microterminals on the ground. The efficiency of digital satellite services and the resulting reduction in the cost of services stimulated the rapid growth of global, regional, and domestic demand and has also seen a shift of space-networked FSS offerings to ever higher frequencies. Thus FSS offerings are now in the C-band, Ku-band, and Ka-band and there could conceivably be use in future years in even higher bands such as the so-called Q, V, and W bands in the millimeter wave frequencies.

Decentralization of FSS Services as Small Ground Systems Move to the "Edge" of Global Networks

The early days of satellites were controlled by the large telecommunications monopolies that saw fixed satellite services as a means to interconnect national communications with overseas countries because of the limited capacities of the submarine cables of the day. In this early satellite market, large national Earth stations connected to satellites of still limited capacity and therefore it was the national telecommunications terrestrial networks that controlled all international traffic. The subsequent emergence of national satellite systems and national television satellite distribution changed not only the market structure, but also spurred the rise of new satellite systems to compete with national terrestrial networks.

Once this trend started, it created increasing pressure to design smaller and more cost-efficient satellite Earth stations that could bring traffic connectivity closer and closer to the headquarters of large businesses, to satellite broadcasters, and to cable television networks. It likewise created the demand to design and build very low-cost, small, receive-only satellite ground stations for consumers to get television and radio programming. This trend started with the early national satellite systems in Canada, the Soviet Union, the United States, and Indonesia and then spread to dozens of countries around the world. In turn, this spawned what might be called the VSAT (or the very small aperture terminal) revolution. Instead of hundreds of Earth stations to connect the countries of the world, there were, over time, hundreds of thousands of transmit and receive small satellite antennas located at businesses and eventually many millions of television receive-only (TVRO) terminals.

This trend started during the analogue era of satellite communications but mushroomed with the dawn of the era of digital satellite communications. Digital transmission, with its more efficient use of limited satellite bandwidth and allocated frequencies, made the system efficiencies of satellite communications that connected much smaller antennas ever more attractive. Instead of 30 or even 10 m Earth stations, there were now two-way transmit and receive VSATs that were 3 m or smaller in diameter. As satellites became larger and more powerful, the system economics and evolving technology encouraged the building of even smaller microterminals which were also cost-efficient. Thus, there was a series of technological advances such as 3-axis stabilized satellites with higher gain antennas; more powerful satellites; the deployment of satellites in new higher frequency bands such as Ku-band and Ka-band; the conversion to digital satellite services; and onboard intelligence, switching, and processing. These advances not only allowed higher capacity satellites but also satellites capable of working to even smaller and more cost-effective ground antenna systems. The NASA program in the United States to develop the Advanced Communications Technology Satellite (ACTS) that demonstrated the use of Ka-band frequencies and onboard processing helped to move this process along during the late 1980s (Fig. 3).

The most remarkable aspect of the new technology made possible by experimental satellites such as the ACTS satellite in the United States, the ETS VI satellite in

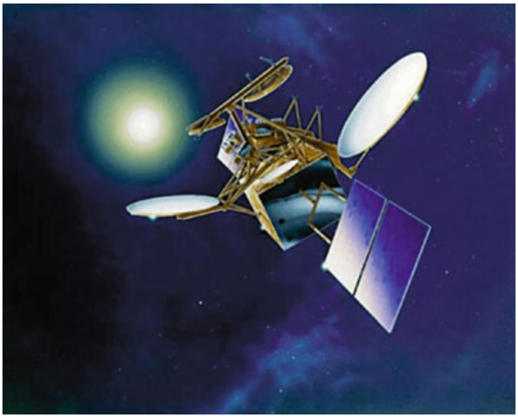


Fig. 3 The advanced communications technology satellite developed by NASA in the 1980s to promote new digital capabilities and Ka-band utilization (Graphics courtesy of NASA)

Fig. 4 The ACTS smallest user terminal was only 60 cm in diameter (Graphics courtesy of NASA)

Japan, and the ARTES satellite in Europe was actually realized on the ground. These new satellites demonstrated very broadband capabilities that could be accomplished to small and compact ground antenna systems. The ACTS ultrasmall aperture terminal (USAT) pictured below had only a 60-cm (about 2 ft) aperture yet could receive data rates of 45 Mbits/s with a lower upstream return rate of 1.5 Mbits/s (Fig. 4).

This research program hastened the conversion of the FSS industry to digital video broadcasting services. The digital broadband distribution function could send high-speed data to support television, voice, or data services, which could be used to download new computer software, validate a credit card, or update a global corporation's inventory at thousands of outlet stores. In the age of the Internet, this has perhaps been the most significant stage in the evolutionary process for today's FSS digital networking services. The latest stage in this evolution has been the increasing shift by businesses and private users to employ Voice over IP (VoIP) services regardless of whether the data stream might be going over satellite, fiber, coax, or microwave relay.

The international standards to allow this digital broadcast service to be interactive with downstream rates have now fully evolved. This digital broadcast service is often in the 36–72 Mbits/s range downstream, with thinner stream response uplink rates that originate from 1 m microterminals. This shift to digital broadcast services have thus served to move FSS services closer and closer to end users. Large multinational enterprises with enterprise networks can thus use such digital broadcast satellite networks to connect efficiently with thousands of node points. For example, large oil companies can use these networks to link with all their service stations and automobile companies can link to all their dealerships. Global department stores, insurance companies, banks, and airlines can also connect with great flexibility to thousands of locations worldwide.

The two most predominant standards that allowed the development of this type of asymmetrical global satellite digital networking (i.e., a heavy stream of data out from corporate headquarters and thin route return data service) are known as: (1) Digital Video Broadcasting with Return Channel Service (DVB-RCS) and (2) Digital Over Cable System Interface Standard (DOSCIS). In the case of DOSCIS, this service was first developed for cable television networks on terrestrial systems, but then adapted to use on satellite networks as well.

This new type of digital broadband satellite service has truly allowed satellites to support global networks with thousands or even tens of thousands of nodes very cost-effectively.

The shift to large-scale digital networks via satellite has, however, presented a great challenge to the fixed satellite service (FSS) industry. The problem is that most large-scale digital networks today operate using the Internet Protocol. However, the original IP interface connections were established on the basis of terrestrial networks where the issue of satellite transmission delay and the IP Security (IP Sec) procedures did not take into account the particular characteristics of satellite transmission. These two issues initially made it very difficult to use satellite-based digital networks using TCP/IP (Transmission Control Protocol/Internet Protocol) efficiently. Satellite transmission delay was mistakenly interpreted as network congestion and led to slow recovery procedures. In time, the clock for detecting congestion was reset to take into account satellite transmission times and the so-called spoofing methods compensated for geosynchronous satellite-related transmission times. Also the problem of IP Sec procedures, that stripped off key routing header information needed for effective satellite transmission, has also been largely rectified by the

Internet Expert Task Force (IETF) Request For Comment (RFC) processes. The result is the now widely adopted Internet Protocol over Satellite (IPoS) transmission standard. Thus, today large-scale digital satellite networks using IP-based interface standards can operate with much higher efficiency and are typically within 80 % of the efficiencies achievable on terrestrial networks (Kadowaki 2005).

Regulatory Shifts Concerning FSS Systems to Make Them Openly Competitive

The regulatory environment in which telecommunications and IT services are provided on a global basis has shifted dramatically since the 1980s. It was in this decade – especially in 1983–84 with the divestiture of AT&T and its loss of near monopoly status in the U.S. that the satellite market began to shift rapidly. This was when “liberalization” or competitive services began to replace the so-called rate-based regulation of monopoly carriers. This occurred first in the United States, then Europe and Japan, and then around the world.

The inital step in this process actually began in the United States in the 1970s when the MCI Corporation challenged the monopoly status of AT&T in courts by claiming anticompetitive actions. It was also in the early 1970s during the Nixon administration that the US Justice Department opened an investigative process against both AT&T and IBM, charging that there was evidence of anticompetitive practices by both firms. In time, the proceeding against IBM was dropped but the action against AT&T continued. In fact, there were two different but related proceedings. There was the MCI suit against AT&T seeking damages for anticompetitive practices that was ultimately successful. And then there were the antitrust charges brought by the Justice Department which continued through the Ford and Carter administrations until the very waning days of the Carter administration.

At that time, Federal Judge Harold Greene adjourned the proceedings on January 16, 1981 to let the Justice Department and AT&T to see if they could reach a final negotiated settlement. After months of negotiations that went many months into the Reagan Administration, a negotiated final settlement was reached between the Justice Department and AT&T and approved by Judge Harold Greene.

Under the terms of this negotiated final agreement, planning was undertaken to begin the restructuring of AT&T with the divestiture of AT&T actually occurring as of January 1, 1984. Under this negotiated final agreement, the divested AT&T would continue its long distance and international services but it would give up ownership of its various local Bell Operating Companies, which were reconstituted as a series of new regional corporations. AT&T, as of 1984, faced competition for its telecommunications services for long distance and overseas services and it also faced competition in the design and installation of telecommunications facilities. In order for AT&T and its Bell regional operating companies to reliably interconnect, the FCC established rules called “open network architecture” (ONA). This allowed these various systems in the United States to interconnect to common digital standards (MacAvoy and Robinson 1983/1984).

The negotiated agreement changed the entire regulatory structure for telecommunications in the United States. In the past situation, the Federal Communications Commission regulated AT&T by explicitly approving new facilities for telecommunications services that went into an “official rate base.” This rate base of approved facilities allowed AT&T to make a certain amount of profit or rate of return on this “officially approved” investment. Critics of this arrangement included those who were heavy users of telecommunications such as banks, insurance companies, airlines, etc. They argued that this “rate base” system for regulating monopolies created the wrong incentives and that it led to wasteful investment in unnecessary telecommunications facilities (switches, microwave relays, coaxial cables, satellites, Earth stations, etc.) and thus stymied innovation and cost efficiency. Under the new FCC regulatory regime, US telecommunications providers were given incentives to make higher profits if they could lower investment costs and lower their prices to business and consumers.

In Europe, the newly formed European Union was beginning to wrestle with a different but somewhat parallel problem. Its objective was to create an integrated telecommunications system that could allow all of the networks within Europe to be compatible with one another and connect seamlessly as if it were one system. The concept for digital communications under development at that time, called Integrated Services Digital Network (ISDN), allowed largely compatible digital networking and served to provide part of the solution. The major breakthrough was to adopt what they called “Open Network Provisioning” (ONP). The bottom line in Europe was that ONP not only allowed national networks to interconnect seamlessly, but it set the stage for national telecommunications to start competitive telecommunications networks in neighboring countries.

In Japan, there was also interest in the competitive approach to regulation of telecommunications and they sent observers to the United States to monitor the divestiture of AT&T. The result was that the Japanese Diet (the legislative branch for Japan) passed two new telecommunications laws – one dealing with domestic telecommunications and other dealing with international telecommunications. These laws authorized competition for telecommunications services in Japan but restricted ownership of competitive networks to Japanese-owned entities.

Thus from 1984 through 1992 there was a major shift in many of the so-called developed economies to “liberalize” telecommunications regulation and create a regulatory process under various types of open network standards to allow the efficient interconnection of competitive networks.

The situation for satellite communications was complicated in that the Intelsat Agreements that acted very much like a treaty among all member countries and territories (almost 200 in number) specified that there should be a single global satellite network with a mission to provide services at low cost to developing nations. These Intelsat Agreements had been set up under US initiatives starting from the Kennedy Presidential administration. The United States was caught in a difficult situation. The single Intelsat Global Satellite System had been the brainchild of the United States and the Communications Satellite Corporation (Comsat) that had been created by the 1962 Communications Satellite Act by the US Congress.

The United States was the predominant member and owner of the Intelsat system and from 1965 to 1975 Comsat had been the system manager.[2]

In 1983, several filings were made to the FCC to build and deploy new satellite systems that would provide international links in competition with Intelsat. This left US policy makers caught up in a dilemma. Article XIV of the Intelsat Agreement specified that any member country of Intelsat that wished to deploy and operate a separate satellite system must technically coordinate it with Intelsat and if it wished to carry international traffic then it must "economically coordinate" with Intelsat under Article XIV(d) of the Agreements to show that such removal of international traffic did not create "economic harm" to Intelsat.

This economic coordination was successfully carried out by the "Eutelsat" organization for regional traffic essentially within Europe and involved traffic that Intelsat was for the most part not carrying. The Reagan administration favored competition and believed that competitive satellite systems would serve to reduce prices to businesses and consumers. It nevertheless proceeded slowly. It authorized several competitive systems to proceed, but on the basis that the competition would only be to serve large corporations on dedicated "enterprise networks" and not to be competitive for publicly switched telephone traffic. In time, the emergence of regional satellite networks such as Eutelsat, Arabsat, and proposals for an Africasat that proved to be economically viable, as well as a growing number of domestic satellites, created a groundswell of opinion within governments around the world to abandon the concept of monopoly satellite systems owned by governments. There were meetings of the Intelsat Assembly of Parties that allowed the Agreements that had been negotiated originally in 1963–1965 and adopted in definitive form in 1983–1986 to be abandoned with the result that Intelsat was "privatized" and part of the monopoly system spun off as the New Skies organization of The Hague, The Netherlands.

This shift to "privatize" Intelsat and take away its intergovernmental status affected not only Intelsat. The Inmarsat organization for mobile satellite communications and Eutelsat for European and other international services proceeded toward privatization as well. In fact, Inmarsat, of London, United Kingdom was the first to complete the privatization process. Today everything concerning Inmarsat has been "privatized" except for a small unit to assist with public safety for ships and aircraft and a unit to provide assistance for developing countries to obtain satellite services (GAO Telecommunications 2004).

There are several ironic results with regard to the global privatization process for satellite services and the opening of international satellite services to competition. The prime competitor to Intelsat in the 1980s was the so-called Panamsat organization. In the aftermath of privatization, Intelsat has now totally acquired Panamsat through merger arrangements. Thus, the competitor that played a prime role in forcing the privatization of Intelsat has essentially disappeared while Intelsat is as large as ever with ownership and investment in some 80 satellites and is earning the

---

[2] Op cit, Pelton and Alper.

largest amount of revenues ever in its history. The entity named New Skies that was spun off to compete with Intelsat has been acquired by the SES Global organization of Luxembourg and thus has also essentially disappeared as well. Privatization, followed by a number of acquisitions and mergers in the past decade, has seen the reemergence of just a handful of dominant carriers.

The good news for consumers is that this global competitive process has largely seemed to accomplish the goal of lowering the cost of television, data, and voice services via satellite. The price of international connections via both fiber-optic cable and communications satellites are at an all time low. The very largest carriers, namely Intelsat, SES Global, and Eutelsat, have also tended to form alliances with regional carriers in many instances (Pelton 2005). The Appendices to this Handbook indicate the various international and regional systems and the many alliances and partnerships that now exist around the world in the field of satellite communications.

Evolution of FSS Markets from Global Networks to Regional and Domestic Satellite Systems

As described earlier, the first major FSS system was the Intelsat global network that was established to provide international connectivity in 1964 with the first satellite going up in 1965. Intelsat first provided connectivity across the Atlantic Ocean and then followed with connections across the Pacific Ocean. Global connectivity across all three major oceans was not established by Intelsat until 1969, just before the Moon landing.

The success of these early international satellite services stimulated all other uses. The enthusiasm to employ satellites for regional and domestic national services thus also grew apace. By the early 1970s, Intelsat began to lease capacity to countries for domestic services. Even in the late 1960s, dedicated national satellite systems began to be deployed. The Soviet Union and Canada led the way and then the United States adopted an “open skies” policy. This new policy adopted during the Nixon Presidency urged the development of national satellite systems. Shortly thereafter, multiple national satellite networks began to emerge in the United States for fixed satellite and especially for satellite television distribution services. In time, other nations and regions allowed multiple satellite systems to be financed and deployed as well even though the US market remains the most dynamic in this respect.

The United States shifted quickly toward more competitive telecommunications markets and the so-called liberalization process also ensued in Europe, Japan, and elsewhere around the world, particularly within the OECD. This process has continued under a competitive process backed by the World Trade Organization (as discussed in the previous section) and these factors all served to spawn more and more satellite systems at the international, regional, and national levels. This openly competitive process, however, has also led to consolidation. Mergers, competitive failures, and/or outright acquisitions of other satellite systems have also served to narrow the range of competition. Today, there appears to be a narrowing range of global networks as Intelsat has acquired its chief competitor Panamsat and

SES Global has acquired New Skies and bought an interest in many other regional systems. Today Intelsat, SES Global, and Eutelsat are the most prominent globally ranging systems, although there are also many vibrant regional systems and of course an even larger number of domestic systems.

In the appendices to this Handbook, there is an extensive listing of the various national, regional, and international communications satellite systems that exist around the world today. The shift in FSS markets in the past nearly 50 years have been dramatic in terms of the range and volume of services. The Early Bird or Intelsat I satellite that started commercial satellite services had but 240 voice circuits of capacity using analogue technology and had both very low power and low gain antenna. Today's satellites using digital technology and deploying as many as 100 transponders (such as on the Intelsat 8 satellite) can have the capacity of millions of voice circuits or over a thousand video channels. The remarkable thing is that not only do the satellites now have tremendously larger throughput capacities, but the ground antennas are no longer huge, multiton structures but can be only 1 m or less dishes. Despite their small size, these dishes – thanks to digital video broadcast standards – can still support fast, broadband data rates. As of 2012, upward of 18,000 video channels are available via FSS networks for television distribution around the world on a 24 h a day and 7 days a week basis. These FSS networks can be used in very flexible ways to support corporate enterprise networks, data networking, and multicasting, as well as a flow of traffic that can dynamically shift from voice, data, audio, video, or videoconferencing depending on consumer demand.

New Trends in Satellite System Design

Finally there are several important new trends that are creating major shifts in the economics and the overall market dynamics of global satellite communications. The first of these trends that has made a large impact on service costs is the deployment of so-called High Throughput Satellites (HTS). New satellites such as ViaSat 1 & 2, Intelsat Epic, and Hughes' Jupiter are providing major increases in satellite throughput capabilities.

These new types of high throughput satellites represent at least as much as a tenfold increase in data throughput over conventional FSS satellites. This has led to an impulse jump in available satellite capacity that is only increasing. This will impact satellite pricing and destabilize markets in the 2016–2020 time period. The ViaSat 1 and 2 high throughput satellites with a throughput capability of 140 gigabit/second are clearly already changing the pricing structure for video and broadband satellite services in the North American markets. The launch of the Intelsat Epic and the Hughes Jupiter also serves to accelerate the downward movement of transponder pricing. (See Fig. 5).

Another new trend is the deployment of satellites in medium and low earth orbit that are optimized to provide Internet-based services – particularly for underserved areas such as countries with developing economies in the equatorial regions of the world. To date, the 03b (i.e., "Other three billion") satellite system that is deployed in

Fig. 5 The ViaSat-2 that will soon accompany the ViaSat-1in orbit

medium earth orbit is already in service. Planned services such as One Web, Leo Sat and Commstellation would deploy perhaps hundreds of satellite in new types of global LEO constellations for Internet-optimized broadband services. This could well represent "game changing" and "disruptive" technology" in the global satellite business. If Space X, in even more extreme fashion, were to deploy LEO constellations with perhaps thousands of small satellites in such a network this could lead to new economies but also heightened concerns about orbital debris.

Further, it is possible that new capabilities to refuel and provide on-orbit servicing, particularly to high throughput satellites, could further change the economics of the industry. The capabililty to do on-orbit servicing could ultimately help provide relief to orbital debris build-up as well. All of these new trends are addressed in subsequent chapters.

Conclusion

The FSS satellite systems that started the satellite communications in the mid-1960s nearly 50 years ago were the "grand-daddies" of the satellite industry. As the technology matured and the range of services that satellite could deliver expanded, new types of satellite services were developed and systems were adapted to this growing market in a diversity of ways. Today, FSS has spun off direct broadcast satellite systems (known in ITU as BSS networks), mobile satellite systems (known as MSS networks), store and forward data relay (or machine-to-machine networks), and specialized defense and strategically oriented satellite networks. These latter two types of satellite networks actually use different spectrum bands. Even within the mainstream FSS services there are global, regional, and domestic networks and even within these there are networks that specialize in data networking, video distribution, or emphasize highly connective “mesh networks” versus those that use a star (or hub and spoke) architecture. This market specialization tends to affect the technical design of the satellites, the user antennas and terminals, and the interface standards. These specializations can at times complicate the ease and quality of interconnectivity with terrestrial or even other satellite networks. The evolution of IP-based standards, however, continues to serve as the key “glue” that allows all forms of global communications and IT systems to connect together as seamlessly as possible.

The long-term progress made in satellite communications seems likely to continue, but there remain key challenges for the future. The challenges that are discussed throughout the chapters of the Handbook and that consider satellite communications and related spacecraft and launcher needs include:

- Expanding or at least preserving satellite communications spectrum allocations and the need for effective migration to the use of higher frequencies in the millimeter wave band in overcoming precipitation attenuation issues in these new bands.
- Access to adequate GEO orbital positions and minimizing intersystem interference. Closely related to this issue is the effective management and deployment of LEO satellite constellations so as to minimize interference and coordinate between GEO, MEO, and LEO systems.
- Coping with the problem of orbital debris.
- Technical standards to achieve seamless connectivity between FSS and terrestrial networks and even other types of satellite networks – especially related to completely fluid IP interfacing.
- Coping with the issue of satellites constantly changing role as a complement to terrestrial networks, as a potential restorer of terrestrial networks, and at times a direct competitor. (The satellite use of CDMA and TDMA multiplexing vs. fiber-optic systems using DWDM creates an ongoing compatibility issue beyond that of satellite latency and IP Sec-related disruptions.)
- Developing improved satellites, lower cost and more compact ground antenna, and lower cost launch systems to keep the cost of satellite networking moving to even more competitive levels.

The remarkable growth of computer and IT systems and fiber-optic networks worldwide has been so dramatic that they have at times overshadowed the rapid expansion of satellite technologies and markets. Few industries in the history of humankind have expanded more than a 1,000-fold in less than a half century, but the satellite industry in general and the FSS networks around the world have exceeded even this rate of expansion. Now, something approaching 20,000 satellite television channels have replaced the single low-quality black-and-white television channel that Intelsat I was able to achieve in 1965. Instead of satellites with hundreds or thousands of equivalent voice circuits, there are today satellites equivalent of millions of voice circuits. Just one of these massive satellite networks could transmit the equivalent of the Encyclopedia Britannica in a few seconds and the equivalent of the Library of Congress in a matter of hours. New capabilities such as intersatellite links, onboard processing, active rain attenuation response capabilities, extremely high-gain multibeam antennas, exploitation of additional spectrum in the Ka-band frequency bands, and new digital interface standards will allow satellites to improve their performance to even higher levels during the twenty-first century to keep pace with new user and institutional demand for communications and IT services.

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Broadband High-Throughput Satellites
- ▶ Distributed Internet-Optimized Services via Satellite Constellations
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Satellite Communications Video Markets: Dynamics and Trends
- ▶ Store-and-Forward and Data Relay Satellite Communications Services

References

M.R. Chartrand, *Satellite Communications for the Nonspecialist* (SPIE, Bellingham, 2004), pp. 27–42

GAO Telecommunications, *GAO Telecommunications Report to Congressional Requesters: Intelsat Privatization and the Implementation of the ORBIT Act*. US GAO 04–891. (GAO Telecommunications, Washington, DC, 2004). www.gao.gov/new.items/d04891.pdf

Interview with Santiago by Joseph N. Pelton on September 25, 1974 at Intelsat Headquarters, Washington, D.C

N. Kadowaki, Internet and the new broadband satellite capabilities, in *Satellite Communications in the 21st Century: Trends and Technologies*, ed. by T. Iida, J.N. Pelton, E. Ashford (AIAA, Reston, 2005), pp. 62–72

G.E. Lewis, *Communications Services via Satellite* (BSP Professional Books, London, 1988), pp. 82–83

P. MacAvoy, K. Robinson, Winning by losing: the AT&T settlement and its impact on telecommunications. Yale J. Regul. **1**, 1–42 (1983/1984). http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/

N. Pelton, *Future Trends in Satellite Communications Markets and Services* (IEC, Chicago, 2005), pp. 1–20, and 109ff

J.N. Pelton, *Basics of Satellite Communications* (IEC, Chicago, 2006), pp. 41–72

J.N. Pelton, J. Alper (eds.), *The INTELSAT Global Satellite System* (AIAA, New York, 1986)

Satellite Communications Video Markets: Dynamics and Trends

Peter Marshall

Contents

Abstract

The advent of satellite communications brought a new era to the TV industry. In the early years, however, the use of satellites was quite costly and limited by the modest capacity of the first commercial communications satellites. However, the evolution of satellite technology and the development of satellite aggregators, such as Brightstar, Wold Communications, Bonneville, IDB, Keystone, and Globecast, led to a sharp reduction in the cost of satellite television. The development of full-time, annualized satellite transponder charges – as opposed to

---

P. Marshall (✉)
Royal Television Society, London, UK
e-mail: pmsatellites@btinternet.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_85

per-minute fees – was also critical to the development of much lower satellite television fees.

The evolution of digital television services was another important breakthrough. Instead of the one or two television channels per transponder with analog systems, it became possible to derive up to 18 channels per transponder. Digital transmission speeded up the evolution of domestic television satellite systems and played a key role in the growth of direct-to-home satellite broadcasting.

The development of satellite-based video systems was not seamless and encountered periods of major market difficulties. One of the most prominent market development issues was the failure of early direct broadcast satellite systems (or BSS in the terminology of the ITU).

However, DBS (or BSS) satellite markets are now well established in international, regional, and domestic markets. They are not only highly successful but represent by far the largest single satellite market in terms of revenues. Most recently, there has been a rapid growth of high-definition television (HDTV) service via satellite.

These satellite services compete with coaxial cable and fiber-optic-based CATV systems.

The economics of satellite television are quite different from terrestrial networks because once a direct broadcast satellite system is launched and operational there is very little incremental cost beyond the consumer terminal needed to receive service. The advent of new digital interface standards known as digital video broadcast with return channel service (DVB-RCS) and Digital Over Cable Service Interface Standard (DOCSIS) have allowed the rapid development of digital television over satellite and cable systems. DOCSIS is now widely used for both satellite and cable television systems. These digital standards, together with high-power fixed satellite systems and broadcast satellite systems – that by definition have high power – allow not only the distribution of a large number of video channels to consumers but also low-cost distribution of high-speed digital data service to both home consumers and businesses. These digital satellite video systems can be – and indeed are – used to provide broader band digital services to the “edge” of digital networks at low cost. Thus DBS and higher powered FSS satellite systems are now being used to provide commercial broadband data services to business as well as broader band digital services to remotely located consumers.

Keywords

Three-dimensional television (3DTV) • American advanced television systems committee (ATSC) • Broadcast satellite services (BSS) • Broadcast satellite services for radio (BSSR) • Cable news network (CNN) Digital audio broadcast service (DABS) Direct access radio service (DARS) Direct broadcast service (DBS) • Direct-to-home (DTH) DirecTV Dish • Eutelsat • Federal communications commission (FCC) • Fixed satellite services (FSS) • High-definition television (HDTV) • Intelsat • Olympic Games (IOC) • Relay experimental satellite • SES • Sirius Society of Motion Picture and Television Engineers (SMPTE) • Sky

Television • Syncom experimental satellites • Telstar experimental satellite • Ultra high-definition television (UHD) World cup soccer (FIFA) XM Radio

Introduction

The idea of using artificial satellites for broadcasting services was central to the original concept of why a telecommunications satellite system might be deployed in the first place. Today the prime usage of communications satellites is for video services. This can be for video and audio distribution to support CATV and national television networks (i.e., a so-called fixed satellite service application) or to support direct broadcasting services (a so-called broadcasting satellite service) that is provided directly to home consumers or business users. This prime application is reflected in terms of the delivery via satellite of well over 15,000 conventional satellite channels around the world and over 5,000 high-definition television (HDTV) channels and a growing number of direct-to-home and vehicular audio satellite channels. This predominance of the satellite video market is even more apparent in terms of total revenues derived from satellite television and audio services, which exceed 70 % of all satellite communications-related income.

Television and audio distribution and broadcasting services have grown steadily from the outset of commercial satellite communications offerings in 1965 and have diversified in their nature in the past few decades. Although the advent of fiber-optic cable services have tempered the growth of fixed satellite services for telephony and data services, satellite usage for audio and television has continued to expand around the world. This, in large part, is due to the fact that satellite networks (particularly those in geosynchronous orbit) have the ability to cover such very broad areas and thus allow an extremely cost-effective one-to-many service.

The advent of digital television, high-definition television, and the spread of television across the world have strengthened the satellite television market. The innovative new applications related to broadcasting satellite service for radio (BSSR) and digital audio broadcasting services (DABS) have created a new type of satellite market that not only distributes high quality and diverse programming for radio but has evolved a range of two-way applications, such as those related to antitheft services and communications with emergency services in cases ranging from natural disasters to traffic accidents.

The Early History of Satellite Television

The history of satellite communications and the role that Sir Arthur Clarke played in that history were addressed in an earlier chapter of the Handbook. However, there are some key elements of that history that particularly apply to satellite broadcasting services that are important to recall here.

It was Arthur C. Clarke who explained the potential for using satellites for broadcasting purposes first and did it most clearly. This “first” explanation was provided in the British journal *Wireless World* in February 1945, even prior to his better-known October 1945 article. This famous article in the fall of 1945 explained the technical aspects and potential uses of satellites in geosynchronous orbit. But before he presented the technical explanation, he indicated the most powerful application – the “why” of launching an artificial satellite. In a letter to the editor entitled “Peacetime Uses for V2” he wrote:

> *An “artificial satellite” at the correct distance from the earth would make one revolution every 24 hours; i.e., it would remain stationary above the same spot and would be within optical range of nearly half the earth’s surface. Three repeater stations 120 degrees apart could give television and microwave coverage to the entire planet. I’m afraid this isn’t going to be of the slightest use to our post-war planners but I think it is the ultimate solution to the problem*. (Clarke 1945a)

The V2 referred to in the headline was the German rocket-propelled launcher with its high explosive warhead, developed by Werner von Braun and his team of scientists. Hundreds of these weapons were launched against the UK in 1944 and 1945 as instruments of war, but Arthur C. Clarke was the first to present clear technical ideas as to how such systems could be instruments of peace (Fig. 1).

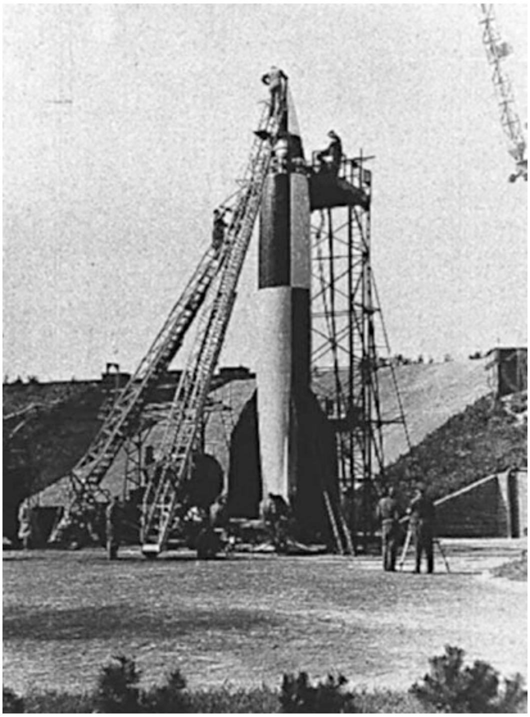


Fig. 1 A German V2 rocket being prepared for launch (Photo courtesy of NASA)

It was through Clarke's writings that the thought of "broadcast towers in the sky" became a reality. In less than two decades, the concept went from being science fiction to becoming scientific reality. By the mid-1960s, distribution of global television news "live via satellite" would become accomplished fact (Clarke 1945b).

The Telstar and Relay satellites in 1962 demonstrated that it was possible to send and receive radio signals from space and indeed to transmit live television pictures via this new telecommunications media. In contrast, the capacities of submarine cables of the 1960s were too limited in transmission capacity to send "live television" programming. Slowing down the transmission of a television signal and then recreating it as a "full motion transmission" at the other end was too expensive for commercial use (Pelton 1974).

The experimental Telstar and Relay satellites, however, were in a low earth orbit. This meant that they were "visible" at ground earth stations for only an 18–19 min window out of every 90-min orbit. Nevertheless, great excitement was created by the first television signals to be sent and received between the AT&T earth station in Andover, Maine, and the British Telecom earth station at Goonhilly Downs in the southwest tip of England.

The successful 1963 launch of the first geosynchronous telecommunications satellite, the Syncom 2, after the launch failure of Syncom 1, created even more excitement. This satellite raised the solid prospect of continuous satellite transmission over an ocean. The three Syncom satellites were designed and built by Hughes Aircraft Company and launched by NASA (Pelton 1974, p. 48).

Television engineers were eager to experiment with the relay of television via the Syncom 2 satellite that was continuously available 24 hours a day. These first experiments were conducted with the cooperation of the Columbia Broadcasting System (CBS) in the USA and the British Broadcasting Corporation (BBC) in the UK. The first on-air transmission was a program broadcast in the UK which consisted of the BBC anchorman Richard Dimbleby introducing grainy and flickering black-and-white pictures of familiar scenes from the USA, including Mount Rushmore, the Mormon Tabernacle Choir from Salt Lake City, and Wall Street, New York (Fig. 2).

By October 1964, Syncom 3 was in orbit above the Pacific Ocean, and the Japanese government learned that it would be in a position to transmit live coverage of the Tokyo Olympic Games to the USA. The Japanese government lobbied for extensive live broadcasting to American viewers. However, NBC Sports had already acquired the exclusive rights from NHK, the host broadcaster in Japan, and were not prepared to change their program plans. They had already sold the commercial advertising time on the assumption that its coverage would be produced on videotape and flown to Seattle for transmission over landlines for delayed distribution across the nation. After considerable pressure from the US government, NBC agreed to a compromise for the opening ceremonies to be transmitted live by satellite. And so, at 1:00 a.m. New York time on a Saturday morning, American viewers saw live black-and-white coverage of Japan's Emperor Hirohito and Empress Nagako presiding over the event and the parades of participants from the Meiji Olympic Stadium (Marshall and Wold 2004).

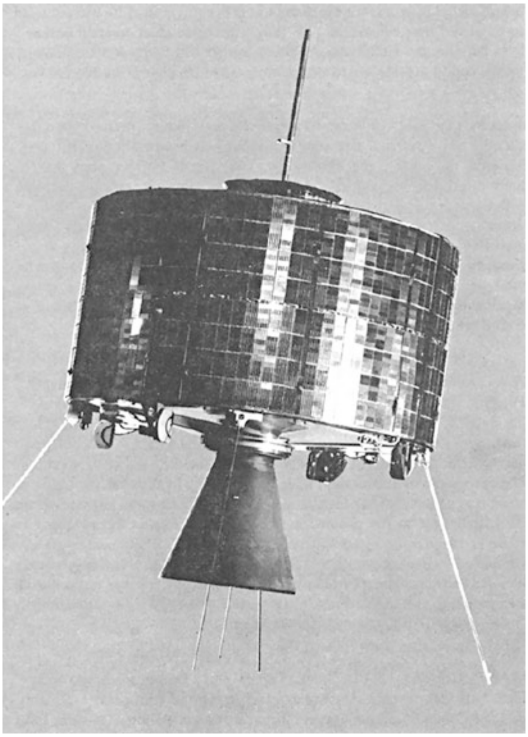


Fig. 2 The Syncom-2 satellite (Courtesy of NASA)

The New York Times TV critic, Jack Gould, wrote: “Live television coverage of this morning’s opening ceremony of the Olympic Games in Tokyo was of superlative quality, a triumph of electronic technology that was almost breathtaking in its implications for global communications” (New York Times 1964).

Although NBC’s viewers were denied any further live coverage of the Olympic events, the opportunity was not entirely lost because Canadian and Mexican broadcasters arranged to receive the satellite signals for some of the events in California and then retransmit them across the borders North and South. And so, although the outcome was limited by NBC’s commercial pressures, this was still a historic occasion. The world had suddenly shrunk and the era of satellite television broadcasting had truly begun.

In the following year, on April 6, 1965, the world’s first commercial communications satellite, Early Bird was launched into geosynchronous orbit and placed into commercial service. This satellite, manufactured by the Hughes Aircraft Company under contract, was an upgraded version of the first geosynchronous Syncom satellites. The Early Bird satellite (officially known as Intelsat I (F-1)) could transmit 240 voice circuits or one low-quality black-and-white TV channel. However, it was not able to transmit television services and telephone or data traffic at the same time, since even a low-quality television signal commanded the full transmission capacity of the satellite.

The Intelsat I (F-1) remained in continuous, full-time service for nearly 4 years. During this period, there were occasional opportunities for the major broadcasters on both sides of the Atlantic to book time on the satellite for special events such as the first launches of NASA's series of Apollo spacecraft. The Intelsat I (F-1) and the Intelsat II satellites, launched largely to support NASA launch requirements associated with the Gemini program, were too limited in capacity to support a high-quality television transmission.

Then in 1968 and 1969 came the Intelsat III series of satellites. These satellites, with a much higher gain antenna and more power, had a capacity of two high-quality color television channels, plus 1,200 voice circuits. The Intelsat III series got off to a difficult start in that the "spinning antenna" that allowed continuous transmission toward the Earth initially "froze" and would not spin. The next Intelsat III satellite was planned to become known as "Olympico" and was slated to provide live coverage for the Mexico Olympics, but it turned out to be a launch failure. Fortunately, the following series of launches were successful. The redesigned bearing and power transfer assembly solved the antenna "freezing" problem and spun around smoothly to allow the "despun" antenna to constantly point toward Earth. Thus eventually these Intelsat III satellites provided coverage for the Atlantic Ocean and the Pacific Ocean and finally the Indian Ocean regions, thereby completing the implementation of a truly global system in June 1969.

It was this network of Intelsat III satellites that enabled a worldwide audience of over 500 million people to see the Moon landing of Apollo 11 on the lunar surface and later to see the first space walk by astronaut Neil Armstrong in July 1969. The signal was sent from the Lunar Excursion Module (LEM) to a radio astronomy telescope in Australia. The signal was then transmitted to the Australian Intelsat Earth Station and from there the signal was relayed around the world.

Events such as the Moon landing and the 1968 Mexico City Olympics created an ongoing demand for live satellite television coverage. But in spite of these momentous occasions, it was still another 20 years or more before the full potential of satellite broadcasting began to be realized and daily television live via satellite relays became truly routine. Years of technological progress and especially digital television transmission were necessary for satellite television to reach its potential. But in the 1970s and 1980s there were other issues in the political, regulatory, and economic arena that certainly served to slow the growth of satellite television – at the national, regional, and global level.

The Evolution of Global Satellite Television Regulation and Tariffs

Political, regulatory, and cost factors limited the growth of satellite television in the 1970s and 1980s with extremely high tariffs serving to act as a significant brake on live global television coverage. Ultimately, it was the spread of competitive telecommunications services in many parts of the world in the 1980s that stimulated satellite television growth on a global basis.

During the first two decades of commercial satellite television distribution around the world, from the mid-1960s to the mid-1980s, it was only the most well-capitalized TV networks of the USA, Canada, Europe, Japan, and Australia which had the financial strength and high production value budgets to be able to afford the use of satellite transmission facilities. But even the most affluent network broadcasting companies around the world, such as ABC, CBS, NBC, BBC, FR3, RAI, NHK, CBC, etc., were limited in their ability to use satellite transmission except for special, high-budget occasions such as the Olympic Games, World Cup soccer tournaments, and news events of global importance such as armed conflicts, royal weddings, or US presidential elections.

A prime barrier to the expanded use of satellites by the television industry was the high tariffs imposed on the use of Intelsat's capacity which restricted the growth of international transmissions. At that time, video service was "sold" on a per-minute tariff basis with a 10 min minimum. Even more importantly, the structure of global telecommunications was set up on the basis of each country having a single monopoly telecommunications provider. Intelsat was owned by these telecommunications monopolies (or signatories) who then bought Intelsat's services at what was to them a "wholesale rate." They each calculated the cost of their earth station operations and what seemed to them a reasonable profit and then charged their broadcasting organizations a retail rate that was equivalent to thousands of dollars an hour for satellite television coverage. This rate could, of course, be significantly higher in countries with limited satellite traffic, and thus the broadcasters ended up facing a very large bill for anything other than a very short transmission.

The signatory tariffs on top of the Intelsat tariffs were thus the product of an international government-owned monopoly structure where there was no great incentive to lower television broadcasting rates that kept pace with technological innovation. This was also partially an artifact of the "focus" of the Intelsat organization and its signatories. Intelsat was originally created to improve global telephony and telecommunications services and most signatories were ministries of posts and telecommunications. Thus to most of them, global television or radio transmissions were seen as not only a "sideline" but also a way to help lower the cost of international voice and data services (or in effect to subsidize what was seen by these telecommunications entities as the most vital service) (Fig. 3).

The total tariffs imposed on the global broadcasting industry were therefore a combination of the short-term (per minute) rates charged by Intelsat to the national carriers (usually government-owned ministries) for the use of satellite bandwidth, plus the 30 %, or sometimes even greater, markup added by the national carrier, and then additional charges these same organizations billed for using their terrestrial earth stations and the landlines to the studios of the broadcasters.

Also, there were certain governments around the world who saw potential threats to national sovereignty or security in the use of the Intelsat system for broadcasting to the rest of the world television news coverage concerning local events. Countries with authoritarian rule and with closely managed news saw the free exchange of satellite news stories by a "free world press" as being against their own best interests.

Fig. 3 The INTELSAT building in Washington, DC (Photo courtesy of Intelsat)

Such governments were therefore not displeased to see high tariffs acting as a barrier for international satellite television.

Despite persistent lobbying over many years by the broadcasters and their regional organizations such as the European Broadcasting Union (EBU), the North American Network Broadcasters Association (NANBA), and the Asian Broadcasting Union (ABU), tariffs remained quite high and discounted rates for volume usage were not available.

Most of the Intelsat signatories were reluctant to recognize the simple economic fact that lower tariffs would release a pent-up demand for satellite transmission services. In short, it would eventually be demonstrated that when prices were cut, there was price "sensitivity" associated with international television utilization so that when prices finally were reduced overall revenues dramatically increased.

Change began to occur in the 1970s. At the urging of the Nixon White House, a new US "open skies" policy was put into place under the auspices of the Federal Communications Commission. This policy opened up the design, manufacture, and deployment of US domestic satellite communications systems to competitive applications. This action was to set the stage for competitive international satellite systems, but not until the 1980s.

Meanwhile, there was the launch of the Intelsat IV satellites in the early 1970s, each with a capacity of 4,000 telephone circuits plus two color television channels. This greatly expanded satellite capacity suggested the need for change in Intelsat's charging policies. The officials within the Communications Satellite Corporation (COMSAT), which served as the manager for Intelsat, set up an Intelsat IV Charging Policy Group. This group explored how satellite services might be sold in new ways. The emergence of new digital technologies in the 1970s further stimulated thoughts about new ways to sell satellite services. In the Intelsat IV era the Intelsat Board agreed, among other innovations, that Intelsat would sell spare capacity to signatories who wished to lease satellite transponders to establish domestic telephone, data, and television distribution services. In the early 1970s, Algeria became the first country to lease such capacity for this purpose. They used the capacity during the day for voice and data services, but from 5:00 p.m. they switched over to distributing television programming to regional capitals across the country. In a few weeks, the centuries-old tradition that the bazaars closed at sunset changed with the shops closing at 5:00 p.m. because everyone went home to watch television.

It was with this first Intelsat transponder lease of capacity to Algeria that the idea was born of leasing capacity on a bulk basis, rather than only selling services on a per minute or per voice circuit charge. Dozens of other Intelsat member countries followed suit to sign domestic leases for telecommunications and television services. This was later followed by agreement by the Intelsat Board to lease capacity to the Australian television broadcaster Kerry Packer to provide regional television distribution in the Pacific region.

The true breaking of the logjam created by high per-minute rates for television came with the move to end monopoly telecommunications services and to introduce competitive opportunities in the USA, Europe, and Japan. The year 1983 proved to be pivotal.

In the USA, a US Justice Department suit was brought against the AT&T Corporation regarding its "near monopoly" role and various alleged abuses against new telecommunications competitors such as the MCI Corporation. Judge Harold Greene "settled the suit" by obtaining consent from AT&T to his final rule making and order. This order led to the divestiture of AT&T into various parts that could broaden telecommunications competition. This Federal court ruling, that provided new levels of competition at all levels of telecommunications throughout the USA, also served to provide a surge of filings for new domestic satellite systems under the "open skies" policy. This in turn also hastened the development of satellite distribution to cable networks and further set the stage for competition.

Finally, in 1983, in a move that was a surprise to the world telecommunications community, a newly organized company in the USA named Orion filed with the FCC for permission to launch and deploy a satellite system. This was not for domestic service, but instead this new entity proposed to own and operate an international satellite system that would compete with Intelsat for international telephone, data, and television services. In a short period of time, other new companies filed to do the same. Panamsat sought permission to compete directly with Intelsat on international routes and indicated that they could provide services at lower prices.

The idea of other international satellite systems coexisting with Intelsat had of course already been established. Since 1971, the Intersputnik system had maintained services for Soviet Bloc countries. Next, there had been the creation of Inmarsat in 1979 to provide international maritime and aeronautical satellite services. Further, the creation in 1977 of Eutelsat, an intergovernmental organization in Europe to provide regional satellite services, clearly had set the stage for this additional step toward international competition from the private sector. Despite these precursor steps, the direct attempts to bring commercial competition to the "single global satellite network" were entirely unexpected by Intelsat officials since the USA had heretofore been the organization's strongest supporter and President Kennedy and his administration had championed the idea of a single global system operated for the good of all. Even during the Nixon administration, when the domestic "open skies" policy had been announced, the US government had supported the idea of a continued single global system during the negotiation of the permanent Intelsat arrangements in 1969.

What followed these 1983 filings with the FCC by Orion and Panamsat for separate international systems was years of careful international consultations. In these consultations, the US government, and the Reagan administration in particular, continued to support the Intelsat Organization's standing intersystem coordination provisions as contained in Article XIV of the Intelsat Agreement. Yet, the Reagan administration also quietly worked to find a way to promote international telecommunications competition. From 1980 to 1988 the Reagan presidency thus encouraged telecommunications competition. It actively sought a process whereby the Intelsat Agreements could be amended to allow competition. It took more than a decade and many congressional representatives pushing for competition, but ultimately (in 2001) Intelsat was "privatized" and capitalized as another commercial entity that would compete for its traffic like any other commercial company. No longer would this new entity be an international organization, organized under an international treaty, but simply would compete on the basis of the strength of its technology and commercial offerings. This was, in fact, a sea change in the way international satellite services were to be provided. In fact, it was Inmarsat, located in London, UK which led the way and was the first of the "monopoly" international satellite carriers to become "privatized." It quickly converted to become a commercial carrier owned by private equity firms. Eutelsat, located in Paris, France was also privatized and became owned by shareholders rather than by governmental ministries.

Internationally, banking, financial agencies, insurance companies, airlines, and other major users of global telecommunications services actively supported this process. With the privatization of Intelsat, Inmarsat, and Eutelsat, multinational enterprises and businesses were ultimately able to realize substantial savings on their telecommunications bills. TV news agencies and media networks who actively used satellites to service their many clients around the world, not surprisingly worked in concert with this "liberalization process" to open up competition and bring down the cost of international satellite television services along with the lower cost of global telecommunications.

The difficult aspect of this history to assess is what impact technology made on this process, beyond the regulatory and policy shifts that allowed competition for international satellite services. On one hand, new fiber-optic submarine cables began to be deployed in the 1980s. Certainly, this new form of technological competition would have served to drive down the cost of satellite communications services without the regulatory and policy changes that led to the privatization of Intelsat, Inmarsat, and Eutelsat. Further digital satellite communications also began to be introduced in the 1980s and into the 1990s. In the analog era, only one or possibly two television channels could be derived from a 36 MHz transponder. But by the 1990s up to 18 television channels could be derived from a 72 MHz transponder. In short, almost ten times more television channels could be transmitted via a single satellite transponder as a result of the transition from analog to digital technology. This transition to digital technology, and its breakthrough efficiencies, clearly drove down costs as well. The latest in digital compression technologies, as represented by the Motion Picture Expert Group (MPEG) standards such as MPEG 2, MPEG 4, and MPEG 6, enabled the most efficient throughput of video channels, through fiber-optic cables, coaxial cables, or satellite transmissions. These have certainly served to boost the growth of video services around the world and helped to greatly reduce the price of television distribution and television broadcast.

The Need to Understand the Key Concepts of Television "Contribution," Television "Distribution," and Television "Direct Broadcast"

To help understand the evolution of satellite broadcasting, it is useful to explain at this stage the three separate and distinct terms used in the field of satellite communications – "television contribution," "television distribution," and "television broadcast."

Contribution is the use of a satellite link to transmit TV coverage from where an event is occurring to the broadcasting center or news agency for the television program production. This is usually a point-to-point, unilateral transmission with the technical arrangements being made by the broadcaster or agency concerned in order to get the original television signal to their production center. This process is often referred to as "satellite news gathering." The ubiquitous characteristics and widespread coverage provided by satellite footprints means that it is technically possible to uplink a signal from virtually any point on the earth's surface. In the 1980s, the only limitations were accessibility for the necessary uplink vehicles with antennas of 3 m diameter – although there were many examples of intrepid teams taking equipment to the jungles of South America and the Antarctic.

As technology developed – including the gradual transition from the mid-1980s to the 1990s from analog to digital signals – the size of the transportable uplink equipment was reduced, first to "flyaway" units which could be shipped in boxes as regular airfreight and then to portable one-man backpacks. At the same time, the arrival of each new generation of satellite news gathering (SNG) equipment was matched by a reduction in satellite transmission costs and improvements in technical quality. Thereby, miniaturization and mobility has now made possible a whole new era of "instant news" coverage, with the latest being the use of videophones and even video over Internet.

Another aspect of this trend toward miniaturization has been the use of multiple miniature cameras for special events (with or without satellite transmission) bringing a new dimension to sporting events and major outside broadcasts. Cameras can now be located discretely in race cars, on ocean-going yachts, and even on sports fields and sometimes on players themselves.

New systems, reduced costs, and ingenuity have all contributed to a transformation in the possibilities available for "contribution" by satellite in the past 10 years or so and the trend will undoubtedly continue.

Distribution describes the use of satellites to transmit TV programming or program channels from a broadcasting center to viewers at home, either by "direct broadcast" to an individual rooftop antenna (as will be discussed shortly) or to a cable-TV operator for onward transmission to local subscribers or to an over-the-air local TV broadcaster. There is the possibility of using "distribution service satellites" for the so-called direct-to-the home satellite service in such a way as to "approximate" a direct broadcast to the consumer service. This can involve so-called backyard dishes to receive signals (with or without authorized decoders) in rural and remote areas. Or it can involve very high-powered FSS satellites that provide a service that "mimics" direct broadcast satellite services to very small dishes, but in fact is delivered in the FSS frequency bands. The classic case of the latter type of service is the SES-Astra system in Europe. Thus distribution covers both indirect television distribution via cable television providers or over-the-air television broadcasters and direct broadcasting directly to end users (including direct-to-home television services that do not technically use direct broadcast satellites but still come directly from the FSS-type satellite to the end user).

Satellite television distribution, therefore, is provided in the form of a point-to-multipoint transmission, where the signal from a single uplink can then be received by suitable antennas at any point within the "footprint" of the satellite. These broadcast signals may be "free to air" or they can be encrypted to allow access only by paying subscribers. Clearly, there is a "gray area" between where "distribution" ends and "direct broadcasting" (or broadcasting satellite service) begins. The distinction is based as much as anything on the frequency band utilized rather than the functional operation of the service.

During this period of development, there were an increasing numbers of instances where the broadcasters linked "contribution" and "distribution" to effectively provide live transmissions from outside locations directly into their news and other programs. This has now become commonplace practice around the world.

Direct broadcasting by satellite (DBS) or the broadcasting satellite service (*BSS*): The first implementation of DBS services developed in the 1980s, at much the same time that international satellite communications was being privatized. The plan was to utilize a new type of high-powered satellites with a specific new global allocation of downlink frequencies and orbital locations to provide this type of service. In some instances, this was seen as a new opportunity to meet national ambitions to have a new truly national television service. However, such systems were high-cost ventures with limited channel capacity (since only high cost and limited capacity analog transmission technologies were available at that time). The early operators in the USA, Europe, and Japan had limited success. These ventures essentially failed due to the fact there was limited programming available at reasonable cost. Also, the satellites were not sufficiently powerful and analog technology could not provide sufficient channels to compete with terrestrial cable television systems. In addition, the home receivers were too large and proved to be difficult and expensive to install (Farr 2008).

To the viewers at home, however, these technical improvements were not significant. It was the range and quality of the programming made available by satellite – plus the low cost of the receivers and service – which created a new mass market. The DirecTV satellite system launched by Hughes and now spun off as a separate corporation, together with Echostar, has today established their leadership in the direct broadcast television service in the USA. Meanwhile, Rupert Murdoch's Sky broadcasting ventures in Europe, Australia, and Asia began to develop new television services which competed seriously with the traditional terrestrial broadcasters by acquiring rights for major sporting events and building up significant audience numbers for their pay-per-view or subscription services. In Europe, Sky together with other operators utilized capacity of the highly competitive direct-to-home television services provided by SES Astra out of Luxembourg that offers a "quasi-DBS service" via very high-powered FSS satellites. Further, Eutelsat now operates a series of Hotbird direct distribution television services across the countries of Europe.

The result is that satellite TV is now received directly in over 35 % of TV homes in Europe, but with variations from country to country. In smaller countries such as Belgium and Holland with widespread cable networks, the figure is around 10 %. However, in the UK, Germany, and Spain the figure for Sky-TV's channels is significantly higher and growing each year.

In Japan, BSS services have been quite successful. The first experimental broadcasting satellite, called BSE or Yuri, was launched in 1978. The major national broadcaster, NHK, started experimental broadcasting using the BS-2a satellite in May 1984. This provided just three channels of programming to 40–60 cm (13–20 in.) home antennas. However, two of the three transponders failed within a few months and regular transmissions did not begin until the launch of BS-2b in 1989. Another Japanese company, JSB, started broadcasting via the BS3 satellite in 1991, and by 1996 the total number of households receiving satellite TV exceeded ten million. Today, in addition to the NHK service, there are in Japan direct broadcast satellite systems operated by BSAT, JCSAT, WOWOW Broadcasting, and SKY PerfecTV.

In Australia, satellite television has proved to be more feasible than cable TV due to the long distances between cities. Foxtel operates both cable and satellite services to all major cities and its main competitor is Optus Vision. Meanwhile, rural areas are served by Austar.

Another country with a widespread area to cover is Russia, where satellite TV began in the days of the Soviet Union using the Moskva system via the Gorizont and

Express satellites. Today, satellite broadcasting is based on the more powerful Gals, Express, and Yamal satellites, while Eutelsat also provides program channels to Russian homes.

The Middle East region also has a high penetration of homes receiving satellite TV. MBC broadcasts from Cairo via Arabsat and one of its competitors is One TV, based in Dubai and mainly serving the expatriate community with Western programming. The first digital DTH network was Orbit Satellite TV, transmitted from Rome, and there are now many other channels available to viewers in the region from the Arabsat, Asiasat, Eutelsat, and Panamsat systems. In Israel, satellite TV is distributed via the national Amos system.

In all regions of the world, as the technology developed for satellite television "contribution," "indirect distribution," and "direct broadcast" services, the various methods have been increasingly combined on a single satellite. Whether it is a significant news story, even an earthquake in a remote part of the world, or a major sporting occasion, the incoming program material is frequently simultaneously retransmitted to a broadcaster's viewing audience. Thus, the distinction between "contribution" and "distribution" and "direct broadcast" has largely become invisible to the viewer at home.

The Special Role of CNN in Global Satellite Television Development

A major driver in the development of TV by satellite was the launch of CNN, the 24-h Cable News Network which began as a continuous channel distributed by satellite to cable-TV stations around the USA in 1980. The founder, Ted Turner, boasted that once he had switched on the first transmission, it would never stop! And so far, that has proved to be true.

From its Atlanta, Georgia, headquarters, CNN started to build a national and then international newsgathering network and although it was at first derided by the established US networks, it soon became serious competition. In time, the character of CNN began to change to include more in-depth reporting, news features, and interviews – including long-running series with commentators such as Larry King and Wolf Blitzer. However, in 1993, they launched a companion channel called CNN Headline News to maintain the continuous news style.

Major news events such as NASA's Challenger disaster, the Gulf War, the 9/11 terrorist attacks, and successive US election campaigns helped CNN to build growing viewer loyalty. Meanwhile, the Turner broadcasting group, TBS Inc. has grown at a phenomenal pace and now also includes the following networks and businesses:

TBS, Turner Network Television (TNT), Cartoon Network, Turner Classic Movies (TCM), truTV, Adult Swim, Boomerang, TNT Europe, Cartoon Network Europe, TNT Latin America, Cartoon Network Latin America, TNT & Cartoon Network/Asia Pacific, Cartoon Network Japan, Cable News Network (CNN), HLN, CNN International, CNN en Español, CNN Airport Network, CNN en Español Radio, CNN.com, and CNN Newsource. Today, CNN claims that its services reach nearly one billion people around the globe, but it also has to compete with other 24-h news channels transmitted by organizations such as the BBC, Rupert Murdoch's Sky, and the Qatar-based Arabic operator, Al Jazeera.

The international expansion of CNN is another story – and one of the early targets was China. This involved an agreement where CNN would give free satellite terminals to the Chinese leaders so that they could see the world's news. In return, China would ask only $50,000 from CNN for the broadcasting rights for the first year. However, CNN soon found that the Intelsat satellite tariffs to deliver a single television channel in the 1980s would be a prohibitive sum in excess of $20 million a year when per-minute charges were computed.

At about the same time, another TV mogul, Kerry Packer in Australia was urging his national member on the Intelsat Board of Governors to lease a full-time television circuit to spread his Channel-7 programs around the Pacific region. This led to spirited discussions among Intelsat's international members and after a few months, a new rate emerged enabling the annual lease of a TV channel on a full-time basis – subject to the availability of suitable capacity. It took even longer to agree a formula which imposed no restrictions on the number of downlinks within the same coverage area.

Daily News by Satellite

The daily flow of news by satellite outside of CNN and the USA took much longer to achieve. Throughout the 1970s and 1980s, most of the world's TV stations and networks, large and small, relied on companies who were geared to move television news and entertainment around the world efficiently. These organizations included Visnews based in London, UPITN based in London and New York, and the syndication service of America's CBS network. These organizations had camera crews located in all corners of the globe and they relied on airfreight to ship their news stories back to their headquarters to be edited, scripted, and copied, and then redistributed again by air shipments to hundreds of waiting TV stations around the world. Originally, this was a 16 mm film – first black and white and then color – until the first technology shift in this business came with the arrival of video cameras and videotape.

For television news editors and their viewers in every part of the world, these air shipments of film and then videotape provided the only source of foreign material for inclusion in daily news bulletins. This material included not only the top news events of the day but also sports events, fashion shows, and even those humorous items such as "the skateboarding dog" so often used at the end of news programs.

The news agencies provided an increasingly efficient service and as the world TV market expanded, they became large and established businesses. But their product inevitably reached the viewing public 2 or even 3 days after the event in many countries and news commentary writers had to evolve creative ways to prepare scripts which incorporated days-old video into their bulletins without loss of immediacy. But there was a solution waiting – in outer space.

It was at the London-based agency, Visnews, where in the late 1970s it was recognized that satellites could provide a whole new future for news coverage and syndication. By then, it was already technically possible to collect video news pictures from almost anywhere in the world by satellite and then to redistribute it from a single uplink to any TV station within the coverage footprint of a satellite. And yet, as CNN was discovering in China at about the same time, the costs of video transmissions via Intelsat and its members made this uneconomic. The charges for video on a per-minute basis were far beyond the news budgets of even the larger TV stations around the world.

News is a 24/7 business and the first objective was to find a way to distribute the daily packages of news stories to overseas TV stations, 7 days a week and 365 days a year. Executives at Visnews began to explore the possibilities for changing the structure of Intelsat tariffs charged by their national members. But it was not until 1982 that a breakthrough was finally achieved. It came in Australia, where the five competing national TV networks (all of them customers of Visnews) agreed that they would share the costs of a daily transmission from London. Protracted negotiations took place in London between Visnews and BT (British Telecom), and also in Sydney between Visnews, the Australian broadcasters, and OTC (their national telecommunications carrier). Eventually, BT and OTC agreed a basis for securing a block of satellite capacity from Intelsat on the Indian Ocean satellite for a regular daily 10-min transmission from the UK to Australia.

This pioneering breakthrough effectively ended the constraints of regulation and tariffs which had blocked the evolution of video transmission by satellite for many years – and Australian viewers entered a new era of same-day pictures on their evening news shows.

The next step was to exploit the fact that the same daily transmissions from London via the Indian Ocean satellite could also be received in other Asian countries and negotiations began with the TV networks in Japan and their national telecommunications carrier KDD. This was the biggest market in the region and as in Australia, the competing TV stations agreed to share the cost of a downlink from the daily news feed from Visnews in London. And so, by 1982 the international distribution of TV news by satellite began to take off.

Transatlantic services were a bigger challenge. Because of the size of the competing broadcasters in the USA and Europe, they were unlikely to reach a cooperative agreement and Visnews took a different approach by forming a new company to lease a full-time transponder from Intelsat and then sell spot capacity to the broadcasters. This company was named Brightstar, a joint venture between Visnews of London and Western Union in the USA. (Note: A few years later, Visnews was acquired by the Reuters news agency.) Then as digital technology began to replace analog, the American news agency, Associated Press entered the market aggressively with the launch of its AP-TV service.

Meanwhile, other companies such as IDB and Wold International in Los Angeles and Bonneville in Salt Lake City were operating satellite delivery services to broadcasters in the US domestic market. In 1989, Wold and Bonneville merged to create Keystone Communications which went on to expand into international markets. Next, Keystone acquired the video distribution business of IDB and in 1996, the company was acquired by France Telecom and became Globecast, which remains the largest supplier of transmission service to the global broadcast market and leases over 100 satellite television channels around the world.

How Television via Satellites Influenced Global Politics

One of the most momentous events of recent years was the end of the Cold War and in particular the fall of the Berlin Wall in November 1989 (Marshall and Wold 2004). These major political events owed much to how the people of Eastern Europe, through TV and radio, became aware of the freedoms enjoyed in the West.

A West German commentator put it this way: “Totalitarianism could not survive in the East when the people’s antennas were pointing to the West” (Shane 1994). The East German leader had recognized the danger when he said: “The enemy stands on the rooftops” and ordered the communist youth brigade to clamber on to houses and remove the offending antennas. But there were just too many antennas by then and they gave up the unequal task of trying to stop people watching West German TV. This was the period of Mikhail Gorbachev’s *glasnost* which brought an era of increased political freedom in the Soviet Union. According to author Scott Shane in his 1994 book *Dismantling Utopia: How Information Ended the Soviet Union*, the KGB’s information blockade turned out to be “more like a tennis net than an iron curtain” (Shane 1994).

Meanwhile, in Moscow the headquarters of the Central Committee set up an antenna to receive news of world events from CNN’s satellite transmissions, and when this became known, many of the general public began to receive the programs too through the use of illegal and home-made antennas. Radio also played its part in the downfall of communism and the Polish leader Lech Walesa wrote later: “If it were not for independent broadcasting, the world would look quite different today. Without Western broadcasting, totalitarian regimes would have survived much longer” (Nelson and Walesa 1997).

As satellite broadcasting resources expanded, “instant news” became a factor in shaping public opinion. In the later stages of the Vietnam war, Americans came to trust the coverage shown by Walter Cronkite of CBS News and became less willing to accept the official statements from the US military. At that time, the film coverage from Vietnam had to be flown to Hong Kong or Tokyo for satellite transmission across the Pacific, but it still made a graphic impact on the TV audiences and this was an important factor in the change of US policy and the eventual withdrawal.

Another early example of how instant television coverage could make a large political impact came in 1984. This was the case of the Shatila Refugee Camp massacre in Lebanon. Several Western TV cameramen arrived on the scene at almost the same time as the main Israeli forces. This was just a day after an advance Israeli force had descended on the camp and had essentially engaged in a massacre. The Israeli government was still considering how to explain and minimize the political impact of this horrible event when the first video images were being fed by satellite to Visnews in London. The undeniable images of mass death were widely shared by broadcasters in Europe and the USA. This resulted in a global outcry that took the Israeli government by surprise before top leaders were even aware of what had transpired.

The Chinese government was certainly not prepared in 1989 for the live global video coverage of the events in Tiananmen Square. By coincidence (or perhaps not), international TV crews and satellite links were already in Beijing for a visit by the Soviet leader, Mikhail Gorbachev, and there was little the authorities could do to restrict their activities. The use of tanks against unarmed students and the symbolic picture of a lone student facing down the advancing military units had immediate impact around the world.

By the first Gulf War in 1991, the technology had moved on and the images of invasion, and the bombs and missiles over Kuwait and Baghdad were also seen live via satellite. So too were the military incursions into Grenada and Somalia.

More recently, on September 11, 2001, the terrorist attacks on the World Trade Center and the Pentagon were seen "live" by satellite around the world, thereby influencing public opinion on a global scale. What then followed, first in Afghanistan, then in the United Nations, and in the invasion of Iraq in March 2003, were examples of ways in which both public diplomacy and warfare are now carried out in the full glare of instant media coverage.

Sports Programming on Communication Satellites

Sports coverage on television – and especially live sport – invariably works to attract big audiences. It is therefore to be expected that some of the biggest audiences in the world have come from satellite coverage of the Olympic Games and World Cup soccer tournaments.

It was the experimental Relay-1 satellite in 1964 that provided limited television coverage of the Winter Olympics at Innsbruck in Austria. The European Broadcasting Union's production people were able to send limited coverage to the USA. Although the EBU's programming only lasted 20 min at a time due to the satellite's low Earth orbit, it still gave audiences a taste of what was to come.

A few months later, in October 1964 came the Tokyo Olympic Games described earlier. Since 1964 the global audience for the Olympics has grown to staggering levels. According to Nielsen Media Research, 4.7 billion viewers worldwide tuned in to at least some of the television coverage from the Beijing Olympics in 2008 – one-fifth greater than the 3.9 billion who had watched the previous 2004 Olympic Games in Athens.

In the intervening years, the IOC (International Olympics Committee) has been able to exploit the growing global audience, and today it has become more than a billion-dollar enterprise. More than half of the IOC revenues come from the TV rights, which are the subject of fierce negotiations among the major broadcasters.

As satellite systems expanded the market, viewership increased exponentially, and the escalation of broadcasting rights has rocketed. For example, the rights to the

Winter Games in Innsbruck in 1964 were sold for $936,000. For the Atlanta Summer Olympics in 1996, NBC paid $456 million and by the Sydney Olympics in 2000, the figure had nearly doubled to $707 million.

Then for the Winter Olympics in 2006 and the Summer Olympics in 2008, NBC was willing to pay a combined total of $1.5 billion. And this has been followed by a staggering sum of $2.2 billion for the 2010 Winter games in Vancouver together with the 2012 Summer games in London. [1]

For the 2016 Olympics in Rio, the IOC negotiated similar sums for global TV rights, and with growing competition from broadcasters, a deal was announced in 2015 with the rights for the Summer and Winter Olympics in 2020 and 2022, respectively, being awarded to the US Discovery channel for 1.3-billion Euros, for transmission through its associated broadcasters around the world.

Estimates of the worldwide viewership were 600 million for the Mexico City Games in 1968; then 900 million for the Los Angeles Olympics in 1984. But the arrival of global satellite distribution swelled the audience to an estimated 3.5 billion for the 1992 Games in Barcelona. Every 4 years the number of viewers continued to rise and in 2008, the Beijing Olympic Games attracted the largest global TV audience ever. The Nielson Company estimated that between August 8 and 24, some 70 % of the world's population tuned in to watch the TV transmissions.[2]

Over the same period, the worldwide enthusiasm for soccer produced comparable viewing figures for each successive World Cup tournament. This event was first televised in 1954 and now competes with the Olympics as the most widely viewed and followed sporting event in the world. The cumulative audience for all of the matches played in the 2006 tournament in Germany was estimated to be over 26 billion spread across 214 countries and territories. For the World Cup final, it was estimated that 715 million individuals watched the match (a ninth of the entire population of the planet).[3]

The Growth of "Direct-to-Home" Satellite Television

With "national" satellite systems in place, it was not surprising that a number of scientists, engineers, and businessmen around the world (and especially in North America and Europe) began to ask why not send the satellite signal directly to the home and bypass the terrestrial networks. There were regulatory as well as technical obstacles to be overcome, but many consumers, particularly those in rural and remote areas without access to cable television or over-the-air broadcast, began buying back yard television receive only (TVRO) dishes to receive satellite TV intended for cable stations. In many cases, they also bought "descramblers" and

---

[1] http://en.wikipedia.org/wiki/Olympic_Games

[2] http://blog.nielsen.com/nielsenwire/media_entertainment/beijing-olympics-draw-largest-ever-global-tv-audience

[3] http://www.FIFA.com

began to watch so-called premium channels such as HBO and Cinemax. After a few years, there were literally millions of these “pirate” backyard dishes in the USA, Caribbean, and elsewhere, and it became obvious that so-called direct broadcast satellite television services would be a viable business.

Indeed today direct broadcasting represents the largest commercial satellite market, as measured in the size of its global revenues. Because of this history, whereby the initial service evolved from fixed satellite service (FSS) satellites providing service to backyard dishes before the arrival of new types of more powerful direct broadcast satellites designed for direct-to-home services (DTH), there is a confusion as to where one service ends and the other begins. Today, with a more pragmatic approach, the designation DTH is generally used to cover both types of services, whether for backyard dishes served by FSS satellites that operate in several downlink frequency bands and broadcast satellite services (BSS) spacecraft that operate in another downlink band.

However, BSS remains the formal terminology used by the International Telecommunication Union (ITU) for a direct-to-the-consumer television service. This BSS offering is considered by the ITU to be separate from the FSS satellites, and it is different in that higher transmit powers are authorized, different frequencies are allocated to the service, and the user terminals are typically much smaller and compact and cost less money than the backyard dishes. BSS terminals are typically 30 cm to 1 m in size (12–39 in.), while backyard dishes are 3–7 m (10–29.5 ft) in size.

The “true” BSS direct broadcast service is one in which many television channels are uplinked to a broadcast satellite in geosynchronous orbit and then downlinked at very high power via a highly concentrated beam to a country or region so that the signal can be directly received by quite small home or office-mounted satellite antennas. Under the ITU allocation of frequency bands for this type of service, different spectra are used in different parts of the world. These high-power downlink transmission bands are as follows: The spectrum 11.7–12.2 GHz is allocated in what is known as ITU Region 1 (this region includes Europe, the Middle East, Africa, and parts of Russia). The downlink spectrum of 12.2–12.7 GHz is allocated for ITU Region 2 (this region includes all of the Americas). Finally, the spectrum band of 10.7–12.75 GHz is allocated for downlinking BSS services in Region 3 which includes Asia and Australasia.

Many countries, especially developing countries and those without satellite communications capacity, thought that the initial process by which frequencies were allocated for FSS services at special and extraordinary sessions of the World Administrative Radio Conferences in 1959 and 1963 were biased in favor of the most advanced countries. Thus, when sessions were held to allocate frequencies for this new broadcast satellite service in the 1970s, a number of countries supported the idea that allocations of frequencies for this service should also include allotments for countries that did not yet have satellite capabilities against their future needs.

In the late 1970s and early 1980s, the European Space Agency was interested in launching an experimental direct broadcast satellite that was given various names

L-Sat, H-Sat, and eventually Olympus. This project was complicated by a procurement process in which the French and German governments decided that they would not sign on to fund their “voluntary allocations” associated with this project and instead decided to proceed with their own joint project known as TV-Sat in Germany and the FR-3sat in France.

As these various “official projects” proceeded to develop direct broadcast satellites in accord with ITU allocations, the Luxembourg-based company, now known as SES, decided that it would use a high-powered fixed satellite service (FSS) satellite, operating in the FSS Ku-band, to broadcast over Europe to provide what became known as direct-to-home (DTH) service. Consumers could have small antennas installed at their home and receive a wide range of television programs – far wider than that offered by national terrestrial broadcasters – and the consumers cared little about ITU allocations or service definitions. In short SES, via its Astra satellites, stole a march on these European official BSS alternatives. In the UK, the company known as British Satellite Broadcasting made expensive investments but when they ran into financial difficulties, they merged with Rupert Murdoch’s Sky Television, operating on the Astra satellite system. This company became known as BSkyB and now transmits the multichannel Sky services targeting much of Europe and with nearly ten million subscribers in the UK alone.

Satellite Radio Broadcasting

Radio broadcasting is now an important additional aspect of the satellite industry. Commercial satellites from the earliest days were able to use broader band channels to send high-quality audio, music, and radio shows from one location to another. As the transition was made from analog to digital satellite transmission, the idea that satellite radio systems might be deployed came into much clearer focus. The motivations for this type of service came from many different perspectives. One motivation was from “official” national radio broadcasting systems that might be characterized as sending favorable “propaganda” on behalf of one country or another. During the Cold War years, very large amounts of money was spent to establish high-power terrestrial broadcast systems to send music, entertainment, and news to locations across “Cold War boundaries.”

It was not surprising that many satellite planners thought that a global or regional satellite system, able to broadcast to small compact shortwave radio receivers, could be a technically more efficient and cost-effective method for sending this information to millions of listeners. From another perspective, broadcasters, educators, and news people in the developing world recognized that there were more radios than television sets in some of the poorest countries. They believed that a satellite radio broadcasting services might be an effective way to reach a new and broader audience in these parts of the world.

In the more economically advanced areas, entrepreneurs envisioned that a satellite radio broadcasting system might be a way to reach a broad new audience in their automobiles and even home listeners and office workers who wanted high-quality news, entertainment, and sports on a commercial-free basis by paying just a small monthly subscription fee. For these various reasons the ITU allocated frequencies for a satellite broadcast service that is variously known as digital audio broadcast service (DABS), direct access radio service (DARS), and broadcast satellite services for radio (BSSR).

Once a frequency allocation for satellite radio broadcasting was finally agreed, a number of companies actively pursued this business. At the lead was a company called Worldspace, headed by a charismatic Ethiopian visionary, who had actually been a leading advocate of this new radio service and who had led the fight for new satellite frequency allocations within the ITU processes. Worldspace proceeded with the immediate design, manufacture, and launch of its Worldstar satellites in partnership with the French firm of Matra Marconi. The first of these launches put Worldstar 1 into geostationary orbit for new radio broadcasting services to cover all of Africa, the Middle East, and parts of Europe.

The business model for Worldspace was to offer a lease of one or more individual radio channels to broadcasters who would provide their own programming to be transmitted on the satellite broadcasting downlinks. These channels could be used not only for radio news, sports, and entertainment, but at nighttime (or even daytime in some places) they could be utilized to provide educational programming. Indeed these digital radio channels could even be used to download short video educational programs which required a period of hours to send a “slower and narrower signal,” using the smaller “digital pipe.” The cost of the radio sets (including the attendant national import tariffs that often doubled the actual cost of the satellite radio receivers) ranged from about $100 to $200 (US) and resulted in a largely unsuccessful business model. As a result, Worldspace encountered major financial difficulties when deployed in Africa.

Meanwhile, in the USA, two domestic systems known as Sirius (using highly elliptical orbits) and XM Radio (using geosynchronous satellites) were licensed by the FCC and both companies managed to successfully deploy radio broadcast satellite systems. However, the very high cost of building and deploying very large aperture satellites for this type of service led to financial difficulties for both these satellite radio services. Their services were marketed largely to automobile owners on a subscription basis, offered on a 24/7 basis with a very wide range of radio programming, emergency communications, and antitheft services. XM was offered via General Motors automobiles and Sirius was offered via Chrysler and Ford. When the financial difficulties mounted these two systems eventually merged, with XM Radio essentially acquiring Sirius.

In Europe there are also plans for a direct radio broadcasting satellite service to consumers, but the financial problems appear to have delayed the deployment of such a European system. High-quality audio distribution to radio stations, however, is offered with Eutelsat providing service to 1,100 radio stations and SES Astria providing a similar amount.

High-Definition Television (HDTV) via Satellite

One of the major developments over the past 20 years has been the introduction of HDTV, which provides much greater video resolution than standard TV – i.e., one or two million pixels per frame. The early experiments in the USA, Japan, and Europe used analog systems, but these were rapidly overtaken by the introduction of digital broadcasting.

The most successful analog experiment was carried out by NHK in Japan with their MUSE system, but bandwidth restrictions meant that only one channel could be carried by satellite. Although it provided public transmissions until 2007, it was superceded by a growing range of digital HDTV channels from 2000 onward.

In Europe, there were also analog experiments in the 1990s with the 1,250-line HD-MAC system, but this did not lead to a public broadcast system and it was abandoned when digital systems were developed.

In the USA, the Federal Communications Commission (FCC) brought together a group of TV companies, together with the Massachusetts Institute of Technology (MIT) and created the Digital HDTV Grand Alliance in the 1993. They began field testing in the following year, and the first public broadcast took place from the WRAL station in Raleigh, North Carolina, in July 1996.[4]

This led to the formation of the American Advanced Television Systems Committee (ATSC) which organized the live coverage of astronaut John Glenn's return space mission in October 1998, which was transmitted in HDTV to specially equipped theaters and science centers across the country.

Following that inaugural transmission, the TV networks gradually introduced HDTV technology and as more and more channels became available, the sale of home receivers expanded. The first major sporting event to be broadcast in HDTV was the Superbowl XXXIV in January 2000.

Satellite TV companies, such as DirecTV and the DISH network, started to carry HDTV programming in 2002 and today, dozens of channels are available on all the major networks as well as via satellite, cable, and Internet systems. A survey in November 2009 by Home Media magazine found that between 33 % and 50 % of Americans have at least one HDTV receiver in their home (Home Media Magazine 2009).

Meanwhile, the European Broadcasting Union (EBU) coordinated the development work among major broadcasters, and the first HDTV broadcast in Europe was the traditional New Year's concert from Vienna on January 1, 2004. The Belgium-based company Alfacam was the first to begin broadcasting its HD1 channel on SES-Astra's 1H satellite with 4 or 5 hours of programming each day. This helped to "jump start" the sale of HDTV receivers, and over the following years, the number of channels gradually increased so that today there are 114 HD channels carried on Astra and Eutelsat satellites.

---

[4] History of WRAL Digital www.**wral**.com/**wral**-tv/story/1069461/

The number of European households receiving HDTV channels increased to six million by the end of 2009 and to around 25 million by 2013 (SES-Astra). Usage has continued to increase and by the end of 2015, it was variously estimated that in both USA and Europe between 60 % and 75 % of TV homes were equipped with HDTV receivers.

Future Trends

Another technological development seeking to provide a new dimension for both terrestrial and satellite TV is 3D (three-dimensional television), but this is still in its early and experimental stages. Following on from the success of "blockbuster" 3D Hollywood movies in the cinema, the TV industry is now trying to catch up.[5] Major broadcasters in many parts of the world have invested in the new equipment necessary to shoot and transmit in this format and have carried out a range of test and promotional transmissions.

However, despite a great deal of competitive activity, most content providers and manufacturers appear to be seeking standardization of 3D home electronics technology across the industry before moving too quickly into program and movie production. Disney, Dream Works, and other Hollywood studios, and technology developers such as Philips, have asked SMPTE (Society of Motion Picture and Television Engineers) [6] and other industry groups for the development of a 3DTV standard in order to avoid a battle of formats and to guarantee consumers that they will be able to view 3D content and to provide them with 3D home solutions for all pockets. [7]

With improvements in digital technology, 3D movies have become more practical to produce and display, putting competitive pressure behind the creation of 3D television standards. There are several techniques for stereoscopic video coding and stereoscopic distribution formatting including anaglyph, quincunx, and 2D plus Delta.

Most currently available receivers equipped to receive 3D signals are very expensive and require the viewers to wear special eyeglasses. Several major companies are planning to enter this market in 2010 or 2011 and one of the first is Panasonic – their Panasonic Viera TC-P50VT200 comes with glasses and has a retail price of approximately US$2,500. In June 2010, they also announced a 152 in. (390 cm) 3D-capable TV (the largest so far) that will go on sale for 50 million yen (US$576,000). However, it is also reported that the Samsung UN46C7000 46-in. 3D TV can now be purchased for US$2,000 or less.[8]

---

[5] http://www.bbc.co.uk/news/10446419

[6] www.televisionbroadcast.com/article/93370

[7] http://www.eetindia.co.in/ART_8800569756_1800010_NT_d9538c56.HTM

[8] http://en.wikipedia.org/wiki/3D_television

Meanwhile, companies including Philips and Toshiba are reported to be developing 3D television sets using autostereoscopy technology which will obviate the need for special glasses to be worn by viewers. Also, the Chinese manufacturer TCL has worked on the same technology to develop a 42-in. (110 cm) LCD 3D TV which is currently available in China. This model uses a lenticular system and currently sells for approximately US$20,000.

TV program transmission of 3D services has been launched in the USA, Europe, and Asia. In fact, the first 3D programming was reportedly broadcast on the Japanese cable channel, BS11 in 2008 with four programs each day. However, the first complete channel is believed to be the Sky3D channel which began broadcasting in South Korea in January 2010. Then in the UK, the Sky Sports 3D channel was launched in April 2010, with coverage of the World Cup soccer in July and the company has announced that its Sky Movies 3D will follow later in the year.[9]

In the USA, the satellite broadcaster DirecTV provided a live demonstration of their 3D feed at the Consumer Electronics Show in Las Vegas in January 2010, and they became the first US service provider to offer a complete 3D channel package in July 2010.[10] Now with four 3D channels (DirecTV Cinema, N3D, N3D On Demand, and ESPN 3D), it is ahead of most cable and satellite providers.

Meanwhile, Cablevision launched a 3D version for subscribers to its MSG channel in March 2010[11] and ESPN transmitted 25 matches in 3D from the World Cup soccer tournament in South Africa – the first major sporting event to be broadcast in 3D (Satnews April 2010).

ESPN is planning to start transmitting a new channel dedicated to sports and with up to 85 live events a year in 3D (Guardian January 6, 2010).

In Australia, Nine Network and SBS also used a major sporting event, this time a Rugby tournament, for a joint 3DTV trial from May to July, 2010. Using a system developed by the Harris Corporation, the events were broadcast in 3D in Australia's five major metropolitan centers – Sydney, Melbourne, Brisbane, Adelaide, and Perth – as well as in Wollongong and Newcastle (Satnews June 21, 2010).

Conclusion

Satellite television and audio distribution and broadcasting have dramatically changed our world over the past half century. Today people expect – and receive – instantaneous coverage of global news and sporting events in what seems almost like live coverage. Satellite technology has evolved rapidly to allow more immediate origination of news programming in the field. Reporters with mobile units are

---

[9] http://www.3dtv-prices.co.uk/sky.html

[10] http://news.cnet.com/8301-17938_105-20009692-1.html#ixzz15XRmc6xS

[11] http://www.msg.com/3d

able to uplink programming instantly from even the most remote parts of the world. Satellite technology has also become more sophisticated in distributing radio and television programming. Satellites are essential to global television and radio programming since it is satellites that provide the most common way to deliver programming, either to the head-end of cable television networks or alternatively to broadcast programming directly to consumers in homes, offices, or even automobiles, auto buses, trains, and airplanes. As the technology has matured and the cost of satellite programming has dropped over time, satellite video and audio has become more and more pervasive around the world and the number of satellite television channels has soared to 5,000 high-definition television (HDTV) plus over 15,000 digital television channels. This represents a total of $100 billion (US) a year in direct-to-home (DTH) and direct broadcast services plus $3 billion in audio broadcast services primarily for vehicular services. These revenues represent some 77 % of all worldwide satellite services revenues. This huge revenue stream represents over a quarter billion satellite television and audio subscribers worldwide. The advent of digital satellite television has particularly allowed the cost of satellite television to drop and the number of channels available to expand rapidly. Today, the latest frontier is the development of live broadcast of 3D and ultra high-definition television (UHD) directly to consumer-receiving units and provision of digital audio broadcast services directly to consumers on the move. The latest challenge is to integrate satellite broadcasting services with broadband terrestrial cellular services (Tauri Group).

Cross-References

- ► History of Satellite Communications
- ► Space Telecommunications Services and Applications
- ► Trends and Future of Satellite Communications

References

A.C. Clarke, Peacetime Uses for V2, *Wireless World*, (London, 1945)

A.C. Clarke, The Space Station, *Wireless World*, (London, 1945)

K. Farr, *Satellite Television: How Did It Start*? Background notes for history of telecommunications course (2008), kfarr.com/2008/01/26/satellite-tv-how-did-it-start/

Guardian.co.uk. 6 Jan 2010

Home Media Magazine, 25 Nov 2009

P. Marshall, R. Wold, The World of satellite TV: news, the Olympics and global entertainment, in *Communications Satellites: Global Change Agents*, ed. by J.N. Pelton, R.J. Oslund, P. Marshall (Lawrence Erlbaum, Mahwah, 2004), p. 256

M. Nelson, L. Walesa, *War of the Black Heavens* (Syracuse University Press, Syracuse, 1997)

New York Times, 10 Oct 1964, p. B1

J.N. Pelton, *Global Communications Satellite Policy: Intelsat, Politics and Functionalism* (Lomond Press, Mt. Airy, 1974), pp. 48–53

Satnews, 21 June 2010, http://www.satnews.com/cgi-bin/story.cgi: number = 90185488

Satnews, 6 Apr 2010, www.satnews.com/cgi-bin/story.cgi? number = 447140674

SES-Astra Presentation, 17 Mar 2010 (unpublished)

S. Shane, *Dismantling Utopia: How Information Ended the Soviet Union* (Ivan R. Dee, Chicago, 1994)

Tauri Group-Satellite Industry Group, *State of the Satellite Industry* (2015), http://www.sia.org/wp-content/uploads/2015/06/Mktg15-SSIR-2015-FINAL-Compressed.pdf

Mobile Satellite Communications Markets: Dynamics and Trends

Ramesh Gupta and Dan Swearingen

Contents

Abstract

The first commercial mobile satellite service (MSS) system was implemented to meet the urgent needs of the maritime community for improved communications. As enhancements occurred in satellite technology and circuit integration resulted in availability of low-cost digital signal processors, MSS systems were deployed

---

R. Gupta (✉)
Ligado Networks, Reston, VA, USA
e-mail: Ramesh@ligado.com

D. Swearingen
Communications Satellite Corporation, Arlington, VA, USA
e-mail: d.swearingen@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_7

to support smaller earth terminals for aeronautical and land mobile applications. The growth of the Internet and improved wireless access has led existing MSS system operators to introduce new capabilities that include improved data and multimedia access. This transformation in the MSS markets has offered several technology challenges, particularly with higher-power spot beam satellite deployment designed to operate with low-cost personal user terminals.

This chapter addresses the history and evolution of the technology starting with the Marisat system in the 1970s. This historical review covers both the space technology and user terminals and the market characteristics of the various types of MSS systems. Some systems have not been successful in the market and have not done well financially for reasons associated with market demand, cost, and reliability of service and technology. Currently, there is a convergence of terrestrial wireless services with mobile satellite services. One approach has been to offer dual frequency band handsets that switch from one band to the other depending upon whether the user is within the terrestrial system's coverage. The other approach is for the MSS system operator to support both satellite and terrestrial coverage in the MSS frequency bands to achieve seamless operation for the user. The demand for very high data rates for better Internet access had also led to a convergence of MSS and FSS system capabilities, where many of the capabilities of an MSS are being offered by FSS system operators and vice versa. Fortunately, continuing innovations are assuring mobile users the availability of reliable communications from anywhere in the world.

Keywords

Aeronautical mobile satellite service (AMSS) • Ancillary terrestrial component (ATC) • Asia Cellular Satellite System (ACeS) • Complementary ground component (CGC) • Geosynchronous earth orbit (GEO) • Globalstar satellite system • Hybrid satellite systems • ICO satellite system • Inmarsat satellite system • International Telecommunication Union (ITU) • Iridium satellite system • Land mobile satellite service (LMSS) • LightSquared low earth orbit (LEO) • Marisat • Maritime mobile satellite service (MMSS) • Medium earth orbit (MEO) • Mobile satellite services (MSSs) • MSATs • SkyTerra-1 • Thuraya satellite systems • Very small aperture terminals (VSATs) • World Radio Conference

Introduction

Mobile satellite service (MSS) refers to a radio-communication service between mobile earth terminals and one or more satellites. In most MSS systems, the mobile to satellite links are cross-connected to feeder link frequencies which operate with fixed gateway earth stations. MSS systems provide voice, data, and Internet services to a wide variety of mobile users including those based on ships, aircraft, and land. One can think of an MSS system as a cellular system with a repeater in the sky servicing multiple cells (coverage beams) which are much wider than terrestrial cells. The first MSS systems evolved shortly after the successful initiation of fixed satellite services (FSSs) which initially used fixed earth stations with large antennas. Ultimately, more powerful FSS satellite technologies evolved which enabled services between very small aperture terminals (VSATs) with 1–2 m diameter antennas.

Mobile satellite communications systems, on the other hand, have enabled travelers to the remote corners of the world to communicate with the rest of the world with small portable or even handheld terminals. Those systems, initially developed for maritime communications, were deployed on land for mobile users because large areas of the world were not served by terrestrial systems. While the coverage of terrestrial wireless systems rapidly increased around the world, mobile satellite systems have still been needed to fill in coverage gaps and to satisfy the growing demand for a wide variety of services. The various land mobile satellite systems optimized for service to terrestrial areas and to supplement ground cellular systems have been assigned by the ITU Global Mobile Satellite System (GMSS) "country codes" as if they were an independent nation. The following are the assigned GMSS country codes with there being two per carrier: ICO +881 0 and +881 1; Ellipso +881 2 and +881 3; spare +881 4 and +881 5; Iridium +881 6 and +881 7; and Globalstar +881 8 and +881 9. Ellipso was never deployed. Since Inmarsat was initially seen as aeronautical and maritime mobile service, it was not allocated a country code. Mobile satellite systems designed to provide only service to one country were not assigned country codes.

There are actually quite a range of mobile satellite services and these include global messaging, aircraft position reporting, remote area connectivity, disaster relief communications, search and rescue communications, and Internet access for portable and mobile terminals (see Fig. 1).

Today a traveler will most likely have a choice of more than one system to choose from, since the coverage areas of numerous systems tend to overlap. There are systems which provide global coverage like Inmarsat (Gallagher 1989), Globalstar, and Iridium, whose coverage overlaps the regional coverage of systems like Thuraya and ACeS as well as the regional/national coverage of systems like TerreStar and Ligado Networks (Whalen and Churan 1992) (USA and Canada), Solidaridad (Mexico), Optus (Australia), INSAT (India), and JCSAT (Japan). Although there are systems optimized only for low data rate communications like ORBCOMM, most systems were developed to support telephony and various data communications capabilities (Reinhart and Taylor 1992). To understand the capabilities of these various systems and the likelihood of their continued availability, it is helpful to review how these systems evolved along with a brief description of their performance characteristics.

This chapter addresses the evolution of MSS systems over the past four decades, during which satellite technologies have evolved toward higher-power satellites with corresponding reduction in the size of mobile terminals.

First Commercial Mobile Satellite System

Mobile satellite communications services were first provided to the maritime community which needed better radio communications resulting in significantly improved safety services on the high seas. Although the fixed satellite services to

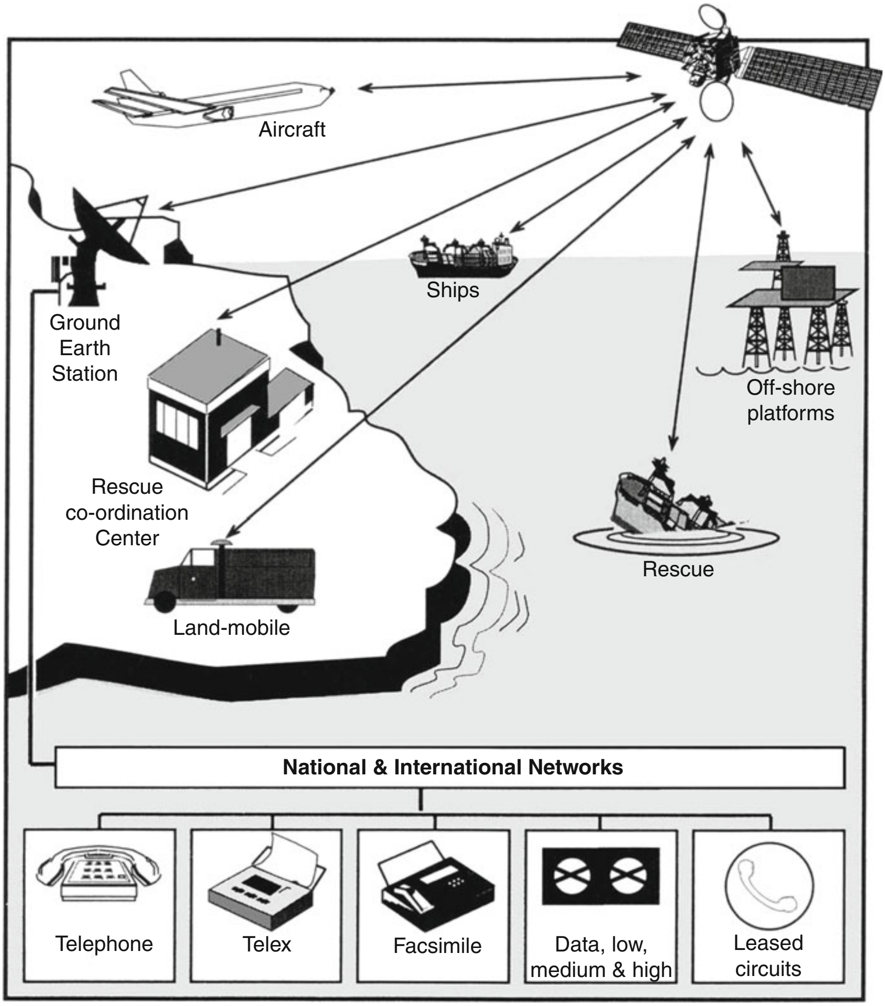


Fig. 1 Traditional mobile satellite services (*MSSs*)

link the continents were firmly established in the 1960s, it was not until the 1970s before a civilian satellite system for ship to shore communications was established by Communications Satellite Corporation (COMSAT). This system, called Marisat, was the predecessor to the Inmarsat (International Maritime Satellite) system that began its operations in early 1982.

Marisat and its successor, Inmarsat, made use of the frequency bands allocated to the maritime mobile satellite service (MMSS) by the 1971 World Radio Conference. These were a pair of 7.5 MHz wide bands at L-Band near 1.5 and 1.6 GHz for the MMSS (one band for ship to shore transmissions and the other for shore to ship transmissions). Those allocations were conserved by cross-strapping with shore to satellite feeder links in the fixed satellite service bands at 6/4 GHz. It should be noted that the 1971 World Radio Conference had also allocated a pair of 15 MHz wide bands to the aeronautical mobile satellite service (AMSS) which were close to the MMSS allocations. At that time the only L-Band allocations for a generic mobile satellite service were a pair of 1 MHz wide bands between the MMSS and AMSS allocations that were restricted to distress and safety communications usage. In effect, because development of a land mobile satellite service (LMSS) was not a high priority in the early years, there were no L-Band allocations for LMSS until some years later.

Early Studies and Programs

Prior to Marisat, high-seas maritime communications depended upon terrestrial radio transmissions in the medium and high frequency bands (MF and HF), which suffered from extreme propagation variability, congestion, and lack of automated station facilities. For this reason, various groups around the world conducted communications satellite technology studies and experiments for aeronautical and maritime applications. The US Navy also sponsored studies and deployed the LES (Lincoln Experimental Satellite) series of experimental UHF band satellites for ship and aircraft use in the late 1960s. The Navy then contracted with TRW Corporation for the construction of five operational ultrahigh frequency (UHF) mobile satellites (i.e., the FleetSatCom series).

The commercial deployment of operational systems for the commercial maritime and aeronautical communities faced serious business start-up challenges. There needed to be enough users equipped with mobile earth stations to recover the satellite investment costs within the 5–7-year lifetimes of the satellites. Furthermore, the maritime and aeronautical communities were somewhat cautious regarding investments in new types of radio equipment.

The maritime community had also begun studies of satellites to improve radio communications for ships under the auspices of the Intergovernmental Maritime Consultative Organization (IMCO). These studies led to consideration of a possible international maritime satellite system (Inmarsat) to be created through intergovernmental agreement and with an organization similar to that of Intelsat. The Inmarsat Convention and Operating Agreements were agreed in 1976, but the new organization did not come into existence until the treaty was ratified in 1979.

In the meantime, the European Space Agency began an experimental satellite program called MAROTS to support an L-Band maritime communications payload. Within Intelsat, consideration was given to the viability of adding a small L-Band maritime payload to its planned Intelsat-5 series of satellites and pricing its usage on an incremental basis. As promising as the studies were, Intelsat initially decided (in 1972) not to complicate its Intelsat-5 program with an additional payload.

Another opportunity for an add-on commercial maritime payload occurred in 1972, when the US Navy sought a pair of “Gapfiller” UHF satellites to service its ships in the Atlantic and Pacific for a couple of years before the Navy’s new

FleetSatCom series could be deployed. COMSAT undertook to explore development of a hybrid satellite that could support both a small L-Band maritime payload for commercial maritime services and a UHF payload to satisfy some, if not all, of the Navy’s stated capacity requirements.

First Commercial Mobile Satellite System: Marisat

In 1973, COMSAT enlisted the help of Hughes (Now Boeing) to develop a small multi-payload satellite design concept with both UHF for the US Navy and L-Band capabilities for civilian users. A fortunate combination of similar coverage requirements (Atlantic and Pacific) and complimentary capacity requirement schedules enabled Marisat to offer cost advantages to both the anchor user (US Navy) and the civilian users (Martin and Keane 1974).

At its launch in 1976, the Marisat configuration provided only a small amount of power for the L-Band transponder in addition to the power needed to support the full UHF capacity for the Navy. Later, as the Navy’s capacity requirement was reduced or disappeared, more power would be switched to the L-Band payload to satisfy growing civilian traffic demands. A third Marisat spacecraft was built to be an on-the-ground spare, but after the first two spacecraft were successfully launched in early 1976, the Navy requested that the third be launched to cover the Indian Ocean region. Consequently Marisat spacecraft were in position to enable global maritime service by the end of 1976.

The Marisat goal was to improve the reliability of telegraphy and telephony services available to ships in the Atlantic and Pacific satellite coverage areas. Because the digital technology of the mid-1970s was still at the early stage of development and relatively expensive (the first microprocessors were just being introduced), the Marisat system architecture was relatively simple (Lipke and Swearingen 1974).

Since the 1971 WARC had allocated only 7.5 MHz of L-Band spectrum for each direction of the ship-satellite link, those allocations were conserved by cross-strapping with shore to satellite feeder links in the fixed satellite service bands at 6/4 GHz. The high receive sensitivity of the 13 m diameter coast earth station antenna enabled the satellite to shore links to be supported with relatively little satellite power. The bulk of the satellite power could, therefore, be dedicated to supporting the satellite to ship links at L-Band.

The shore to ship-satellite repeater translated the 6 GHz carriers uplinked by the coast earth station to 1.5 GHz for reception by the ship earth station. The ship to shore repeater translated the 1.6 GHz carriers transmitted by ship earth stations to 4 GHz for reception by the coast earth stations. In addition to a 13 m diameter antenna system, each coast earth station included telephone and telex channelization systems, an access control subsystem, and switching systems to interface with the terrestrial telephone and telex networks. In order to enable operator assistance to maritime customers, operator positions were also included at the Marisat coast earth stations.

Marisat Channelization

A single channel per RF carrier (SCPC) design was selected for telephony, while a time division multiplexed (TDM) digital RF carrier was used for the telex (50 baud data) service in the shore to ship direction. SCPC was also used in the ship to shore direction, while a time division multiple access (TDMA) was used for telex. The telephone channels used frequency modulation (FM) to ensure a transparent audio path for voiceband data. To lower the perception of background thermal noise during voice conversations, 2-to-1 companding was included in the channel path (similar to the choice made a few years later when the first cellular systems were introduced by the Bell System in the USA).

In the early 1970s, telex had become a nearly ubiquitous mode of automatic telegraphy and was well standardized and widely used for international maritime business data communications. For that reason, telex channelization became the Marisat low-speed data channel standard. Each ship terminal included a 50 baud telex machine to support automatic interworking in both the shore to ship and ship to shore directions. Shore to ship telex channels were transmitted on a 1,200 bps BPSK-modulated RF carrier using a time division multiple (TDM) channel format that supported 22 channels per carrier. In the ship to shore direction, a burst-mode time division multiple access (TDMA) format was selected with up to 22 ships sharing the same frequency by bursting once every 1.8 s with a carrier BPSK modulated at 4,800 bps and containing 12 characters.

Marisat Access Control

Channel assignment and call alert messages were piggybacked onto the shore to ship TDM carrier. Channel requests from the ships were carried in short digital RF bursts on a shared random access frequency called the Request Channel. This basic access control architecture (TDM for outbound signaling and Random Access Request Burst Channel for inbound signaling) turned out to be the essential template for later mobile satellite systems, starting with Inmarsat. The access control design objective was to enable calls to be set up quickly in either direction, in a manner as close as possible to terrestrial telephone and telex calls. For prompt ship response to calls from shoreside parties, the ship terminal was to always be listening to a call announcement channel for possible call alert with its address when not already engaged in a call. In addition to a ship terminal identification code, each assignment would contain codes for the type of call (e.g., telephone or telex) and terminal instructions (e.g., frequency/time slot tuning for the call).

In order to provide rapid ship to shore call setup, historical maritime polling protocols were abandoned in favor of a random access burst signaling channel design. The design provided for a separate ship to shore frequency on which the ship could send a short request message when it wished to place a new call. The message was sent via a single short digitally modulated burst at 4,800 bps. The use of the "Request Channel" frequency would be on a random basis by all ships accessing the satellite (similar to the "Aloha" protocol). The ship to shore request burst contained only minimal information needed by the coast earth station. It included the ship terminal identity code, the type of channel requested, and priority. The coast earth station would validate the ship terminal's eligibility for service and assign a channel frequency (and time slot if telex) via the shore to ship TDM assignment channel.

Marisat Ship Earth Stations

The ship earth station was comprised of an above decks antenna subsystem and a below decks terminal subsystem. The antenna subsystem was enclosed in a protective radome and included a circularly polarized transmit/receive antenna, a frequency diplexer, a transmit power amplifier, a low-noise receiver, and an antenna-pointing/stabilization system. Connected to the above decks unit by signal cables was the below decks equipment. It included the RF to baseband transceivers, access control interface, telephone termination unit, and telex machine.

Practical considerations (e.g., need to locate antenna high on ship for clear sky view) limited the ship terminal antenna size to something on the order of a meter diameter, even though a larger antenna would have required less power from the satellite. For Marisat, the compromise was an antenna size of 1.2 m. Since the ship antenna was directional and would require some sort of automatic pointing, each Marisat could fly in a slightly inclined geosynchronous orbit. This in turn allowed the station-keeping fuel budget to be greatly reduced and more of the satellite mass budget to be allocated to payload. By initially launching into an inclined orbit with a phasing such that the inclination would decrease with time before increasing again, the absolute value of inclination was assured of remaining within 3° over the 5-year specified design lifetime of the satellite.

Marisat Multi-coast Earth Station Interworking

The availability of the third Marisat satellite offered the possibility of commercial L-Band service in the Indian Ocean region. In early 1977, the Japanese company, KDD, decided to undertake the construction and operation of a coast earth station and contracted with COMSAT for the use of the Indian Ocean C/L-Band capacity. In order to insure compatibility, minimum performance requirements for the new station were developed by COMSAT. These were akin to the technical requirements developed earlier for the ship terminals. Acceptance test procedures were also agreed and tests were included to demonstrate compatibility with existing models of ship terminals.

Toward the end of the 1970s, other maritime nations considered implementing a coast earth station (CES) to operate initially via Marisat before Inmarsat capacity became available. Anticipating this possibility, the Marisat designers had included a 4-bit "Coast Earth Station ID" code field within the "Request Message" burst that ship terminals transmitted when requesting a channel. Other provisions that had been made to facilitate possible evolution to multi-CES operation included spare codes within the assignment/alerting partition in the shore to ship TDM carriers (Swearingen and Lipke 1976).

To coordinate the sharing of satellite telephone channel frequencies among coast earth stations, one of the stations was designated as the "Network Coordination Station" (NCS). This station transmitted a "Common TDM" carrier that all ships tuned to when idle. For shore to ship calls, the NCS relayed the call alert from the gateway CES via the Common TDM. If the call was a telephone call, the NCS also added the channel frequency pair to be used for the new call in the alert message.

A ship-originated telex channel request was responded to by the desired CES, which then sent a call assignment message on its own TDM. If the ship sent a telephone channel request, the desired CES would send an incomplete assignment message (without a specific frequency pair) in the signaling partition of its TDM carrier. The NCS would then select a specific frequency pair for the call, complete the assignment message, and relay it via the Common TDM.

The NCS also maintained a list of terminals engaged in calls so that a busy signal could be returned to a shoreside party attempting a call when the ship terminal was already busy in a call. In order to insure that telephone channel frequencies were clear after calls ended, the NCS monitored the frequency slots of recently ended calls to confirm that the RF carriers were gone and sent special clearing signals if a slot still had an RF carrier present.

Inmarsat Start-Up

In 1982, Inmarsat assumed the role of system manager for the existing commercial maritime system, hitherto known as the Marisat system. In 1982, transition was managed smoothly with Inmarsat leasing the L-Band capacity on the three existing Marisat satellites. Using COMSAT's multiple CES interworking design as a basis, Inmarsat adopted its own compatibility standards that enabled existing ship earth stations to continue operation under Inmarsat system management.

Several other satellites with L-Band capacity were nearing completion by the time Inmarsat became operational including ESA's two MARECS satellites and Intelsat's maritime L-Band payloads on its Intelsat-5 series. Within a few years, all of the planned new payloads were deployed and brought into Inmarsat service. The locations of the ultimate first-generation Inmarsat satellite payloads provided coverage from three nominal locations with two payloads near each, one in an operational role and one or more payloads in a standby backup role.

Inmarsat Satellite Capacity Evolution

Because of traffic growth, Inmarsat contracted in 1986 for a new series of Inmarsat-2 satellites with higher power and more bandwidth to be built. The new satellites were successfully deployed in the period 1990–1992 and included bandwidth in aeronautical mobile satellite service (AMSS) bands as well as additional maritime mobile satellite service (MMSS) bandwidth. The L-to-C transponder specifications provided for higher signal gain to support new types of lower-power ship earth stations. The four Inmarsat-2 satellites offered the possibility of introducing a new western Atlantic operating location near 54° W to complement the eastern location near 15° W.

Higher capacity Inmarsat-3 satellites were launched in 1996–1998 which provide spot beam coverage and incorporate the wider frequency band allocations provided to the three mobile satellite services by the 1987 Mobile World Radio Conference. In order to allow for changes in bandwidth requirements, the transponder passband in each direction was segmented in a variety of widths between 450 and 2,300 kHz. The segments were made switchable between beams and capable of being joined to form contiguous channel passbands significantly wider than the widest segment. In order to accommodate diurnal variations in RF power requirements among the various antenna beams, a matrix power amplifier design was used. Each of the five Inmarsat-3 satellites also includes an add-on navigation transponder that serves to relay a GPS overlay signal and to support the GPS Wide Area Augmentation Service for aeronautical applications.

Inmarsat Ship Earth Station Evolution

In order to reduce power and bandwidth consumption per telephone channel and to take advantage of the wider bandwidth and greater sensitivity of the new Inmarsat satellites, new ship earth stations were needed. The new stations needed to be "future-proofed" so as to be compatible with future satellite spot beam configurations and frequency plan changes.

The first new type of ship earth station was the full capability digital *Standard B*. In order to reduce the required RF power, a 16 kbps adaptive predictive coding speech encoding was chosen. The channel design included a rate ¾ forward error correcting code and offset quadrature phase shift keying (O-QPSK) modulation. The feature of the new Inmarsat Standard B design which "future-proofed" the standard was the "System Bulletin Board" included in the broadcast channel. The System Bulletin Board contained essential information regarding the current spot beam configuration and access control frequencies. Changes are notified via a new page of the System Bulletin Board. The "System Bulletin Board" scheme was subsequently adopted and is used as the technique for ensuring ship terminal compatibility with different generations of satellites and various spot beam configurations. In this scheme, the ship terminal reviews the bulletin board page number to determine whether it has already received the current information or needs to be updated. The information in the bulletin includes the frequency of at least one shore to ship signaling channel for each of the spot beams, and the ship to shore frequencies for each beam to be used for requesting service and for acknowledging call announcements. A portable version of Standard B terminal was introduced for land mobile communications.

The next new type of ship earth station introduced in the 1980s was the *Standard C*. This type fills the needs that require only a minimal messaging capability, but a more compact and less costly terminal (e.g., fishing ships and yachts). The Standard C incorporates a nondirectional antenna and a low bit rate messaging channel to support basic store and forward messaging services, as well as distress alerting, data reporting, and shore to ship group calling. The Standard C was well received by the maritime community and adopted by the International Maritime Organization as a minimum communications capability component of the new Global Maritime Distress and Safety System (GMDSS) standards for the high seas. (See Fig. 2 to see a variety of MSS antennas that have decreased in size over time.)

Several aspects of the Standard C access control design are similar to the Standard B access control design. Shore to ship TDM carriers were used to announce message

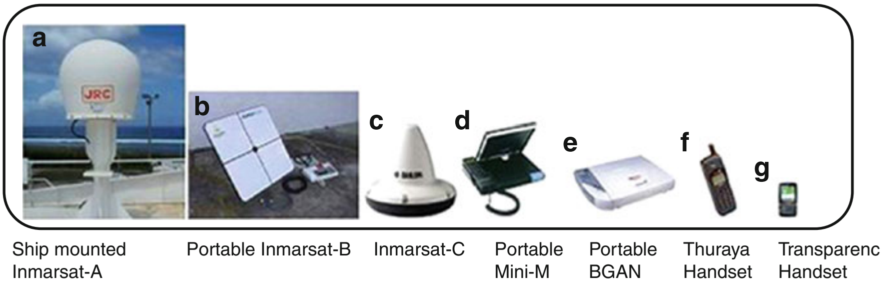


Fig. 2 Evolution of MSS terminals (Graphic courtesy of R. Gupta)

transmissions to the ships and to deliver messages. Initial ship-originated channel requests were transmitted via random access slotted-Aloha channels. Subsequent ship message transmissions were normally transmitted via reserved time slots in a TDMA burst mode similar to the telex transmission modes used for Standards A and B.

Aeronautical Mobile Satellite System Introduction

In the late 1980s with the Inmarsat-2 satellites under construction, a cost-effective aeronautical mobile satellite service became a realistic possibility, provided the aeronautical community was prepared to make use of the Inmarsat-2 satellite capacity. One of the prerequisites was that the aeronautical system design should satisfy requirements established by the Airline Electronic Engineering Committee (AEEC). Consequently, the Inmarsat technical staff participated actively in the working groups of the AEEC to ensure a suitable set of aeronautical system standards.

There was early agreement within the AEEC that data communications would be a "core capability" for air traffic control (ATC) and that an early introduction of such capabilities was important. Furthermore, there was a need for the "core capability" data channels to be compatible with a low gain aircraft antenna. In the forward direction toward the aircraft, the data channels and signaling were to be carried via low data rate (600 bps) time division multiple access (TDM) carriers. In the return direction from the aircraft, data traffic was to be sent on low data rate TDMA channels. Service request signaling message were to be transmitted via slotted-Aloha burst channels. Before the total aeronautical system specifications were completed, an early prototype of the core service was deployed in 1991 for flights over the Pacific. This enabled the airlines to experience the benefits of reliable automatic position reporting for entire flights.

Inmarsat continued to complete the full set of system requirements to support telephony and higher-rate data services. These requirements included a requirement for a directive aircraft antenna with auto-pointing capability. The receive sensitivity (G/T = 10 log [antenna gain/noise temperature]) for the high gain aircraft antenna was to be ≥ -13 dB/K, as compared to the -26 dB/K requirement for the low gain "core capability" aeronautical earth station. With the higher gain, telephony service could be supported in a 21 kbps carrier with a 9.6 kbps speech encoding plus framing and rate ½ error correction coding. Higher-rate data channels were also to be supported via 10.5 kbps carriers.

By the mid-1990s, aeronautical system capabilities were added to several Inmarsat coast earth stations to enable aeronautical service to be offered in all ocean regions. At the same time, satellite communications avionics for various aircraft were developed and were adopted by several airlines. Since the aeronautical system's introduction, it has been adopted not only by airlines but also for general aviation by various government agencies.

Portable Mobile Earth Stations for Telephony and Data

In 1988, Inmarsat decided it should develop a new standard that could operate with a smaller antenna and lower power than the maritime Standard B. Key to its introduction was the selection of a more efficient speech encoder. A competition was held that invited developers from academia and industry to submit their encoding/decoding algorithms for testing against agreed test criteria. The competitive test program was completed in 1990 and resulted in the selection of a very robust 6.4 kbps algorithm known as Improved Multi-band Excitation (IMBE).

The new Standard M combined the 6.4 kbps encoded speech output with some signaling overhead and some additional coding to form a 8 kbps channel, which was then modulated with offset QPSK. With such a low channel rate, the terminal effective radiated power was about 19 dBW under typical conditions and the required receive sensitivity was also greatly reduced so that a much smaller antenna could be used. The typical Standard M terminal was about the size of a standard hardback briefcase and opened up in a similar manner with the lid containing a flat plate multi-patch antenna.

The introduction of the Inmarsat-3 satellites with their improved sensitivity in spot beam coverage area offered an opportunity to make the portable terminals even smaller than the Standard M terminals. In the early 1990s, Inmarsat revisited the choice of speech encoding design and selected Advance Multi-band Excitation (AMBE) which reduced the data rate to 4.8 kbps. Operating through the more sensitive Inmarsat-3 satellite spot beams, the new Mini-M terminal (see Fig. 2) power and size could be significantly smaller than the Standard M. The typical Mini-M terminal is much like a notebook computer with the patch antenna in the raised lid.

By the late 1990s, Inmarsat decided to bring out a service to provide high-speed Internet access via a modified version of its Mini-M terminals. The new GAN terminal standard (Inmarsat M-4) incorporated powerful turbo forward error correction coding and more bandwidth-efficient modulation to enable packet data links to the Internet at speeds up to 64 kbps. This service took advantage of the higher sensitivity of the Inmarsat-3 receive spot beams and the higher effective power levels of the satellite.

The benefits of the improved capabilities of the Inmarsat-3 satellite and the signal processing technologies used for the GAN terminals were put to use in Inmarsat's development of new types of maritime and aeronautical communications standards. The new maritime standards (Fleet 77, Fleet 55, and Fleet 33) offered improved data and voice capabilities via three alternative antenna sizes. The new aeronautical service, Swift 64, offered a compact service optimized for data communications.

Inmarsat-4 Satellites and Broadband Global Area Network (BGAN)

Although the Inmarsat-3 spot beams covered all the high traffic areas for the maritime and aeronautical services, they did not provide totally global coverage. In order to improve and expand the availability of its broadband services, Inmarsat decided to procure three Inmarsat-4 satellites for deployment starting in 2005 (see Fig. 3).

Fig. 3 Inmarsat-4 satellite with multiple spot beams (Graphic courtesy of EADS Astrium)

Each Inmarsat-4 satellite provides 19 regional spot beams covering the entire visible earth disk. In addition, each Inmarsat-4 supports 200 narrow spot beams. With the greater sensitivity of the Inmarsat-4 narrow spot beams, an improved Internet access service, BGAN, could be introduced with data speeds up to 492 kbps and more compact mobile terminals.

The much higher RF power of the Inmarsat-4 satellites and greater sensitivity of the satellite receivers also have enabled Inmarsat to introduce a handheld telephony service. The handheld phone introduced in 2010, the IsatPhone Pro, is similar to a GSM cellular phone. The progression of mobile satellite terminals toward smaller sizes (shown in Fig. 2) illustrates just how much smaller the terminals can be as a result of introducing more powerful satellites (Gupta 2004).

The Inmarsat system has now deployed its latest L-Band satellites, the Inmarsat-5. Further, the latest concept is to integrate Inmarsat BGAN services with the Ka-Band Inmarsat Global Xpress satellites. These latest of all the Inmarsat satellites operate in the Ka-Band spectrum. This network design thus integrates the L-Band services of Intelsat-5 satellites with the much broader spectrum available on the Inmarsat Xpress satellites. Three of the four Inmarsat Xpress satellites have been deployed and the fourth is scheduled for immanent launch. These satellites were manufactured by the Boeing Company.

The Inmarsat-5 satellites operate with a combination of fixed narrow spot beams that enable Inmarsat to deliver higher speeds through more compact terminals. The Inmarsat-5 steerable beams allow for additional capacity to be redirected in real time to where it is needed.

The integrated Inmarsat-5 and Global Xpress networks now allow customers across aviation, maritime, enterprise, and government sectors to have reliable and assured access to high-throughput communications. As part of this new architecture, Inmarsat has worked with Cisco to develop new digital interface architecture. This has led to the development of the Inmarsat Service Enablement Platform (ISEP) and the Inmarsat Gateway. This approach allows Inmarsat to deliver a whole new world of innovative, content-rich applications that are individually tailored to meet users' needs.

As the latest step to fully integrate L-Band mobile communications with the broadband Ka-Band services, Inmarsat has awarded a contract to Airbus to build at least two Inmarsat-6 satellites (Airbus Defense and Space Press Release – Dec 24 2015). These very large and capable spacecraft will constitute a hybrid L-Band and Ka-Band capability within a single spacecraft.

National Mobile Satellite Systems

USSR

Shortly after Marisat was deployed, the Soviet Union undertook to deploy a similar system called VOLNA to serve its maritime communications needs. Like Marisat, it made use of L-Band transponders on geosynchronous satellites stationed at around the earth to provide global service. They included global coverage L-Band transmit and receive antenna beams and used L-Band frequencies below the bands used by Marisat. Also like Marisat, the ship earth stations used antennas approximately one meter in diameter and similar system parameters. Over years, the VOLNA program was expanded to include repeaters that also operated in the aeronautical frequency allocations, although there was very little descriptive material about the associated avionics made available to the western public. The Soviet Union's maritime fleets also made significant use of the Inmarsat system, but little is known about how heavily the VOLNA system was used.

Australia

Australia, with its vast underdeveloped interior land mass, had a serious need to improve radio communications for its land mobile community. To meet that need, the Optus mobile satellite system was developed and its first satellite launched in 1985 with antenna coverage of Australia and New Zealand. The system provided for small transportable terminals similar to the Inmarsat M terminals as well as small vehicular terminals for use while in motion. The Optus system developers made use of speech encoding chosen after a competitive set of laboratory trials and field trials conducted in collaboration with Inmarsat.

Canada and USA

The earliest commercial use of satellites for land mobile communications in the USA made use of existing domestic satellites operating in the Ku fixed satellite service frequency bands. The OmniTRACS system, introduced in 1988, was designed to support the trucking industry in its fleet management. By exploiting the low-power spectral density of low data rate code division multiple access (CDMA), the system developer, QUALCOMM, was able to obtain US regulatory approval for the use of fixed satellite frequency bands for mobile service.

In the late 1980s, a company was established to develop a dedicated mobile satellite system to serve Canada called Telesat Mobile Inc. (TMI) with mobile telephony and data services. Shortly thereafter in 1988, another company called American Mobile Satellite Corporation (AMSC) was formed to provide mobile satellite services to the USA. The two companies collaborated and procured similar geosynchronous satellites (MSAT-1 and MSAT-2) to enable service restoration in case of a failure in one of the satellites. Both systems were designed with wide enough frequency bands to enable not only land mobile but maritime and aeronautical mobile services as well. Unlike the Inmarsat satellites, the MSAT satellites used L-Band frequencies cross-strapped to Ku-Band. The two satellites were successfully deployed in 1995 and 1996 and have been extensively used to support a wide variety of services (e.g., truck fleet management and coastal maritime communications).

Mexico

In 1993, the second-generation domestic Mexican satellite, Solidaridad-1, was launched into a geostationary orbit. This satellite was a multipurpose satellite that included a pair of L-Band repeaters to support mobile services for Mexico. The mobile terminal technology used with the Solidaridad system was similar to that used for the Optus, TMI, and AMSC systems.

Japan

In 1996, MSS services were started in Japan with NSTAR satellites at S-Band (2.6/2.5 GHz) operated by NTT-DoCoMo. The feeder link frequencies were at 6/4 GHz bands. Portable land mobile, maritime, and aeronautical services were offered with 5.6 kbps voice and 4.8 kbps data. Subsequently, JCSAT-9, a hybrid C-, Ku-, and S-Band satellite, was launched in 2006 to augment MSS services at S-Band over a wider maritime coverage in Japan.

Low Earth Orbit (LEO) Satellites

Iridium

In mid-1990, Motorola announced its plans to develop a satellite system using 77 satellites in a low earth orbit to enable handheld terminals to communicate from anywhere in the world. The atomic number for the element Iridium being 77, the new system was given that name. However, the system design was modified so as to use 66 operational satellites, but the system name was retained. The Iridium system called for the 66 satellites deployed in six orbit planes at an altitude of 781 km with an inclination of 86.4°. To minimize the number of required gateway earth stations, inter-satellite links are provided using Ka-Band frequencies to enable mobile stations outside the coverage area of the satellites visible to a gateway to have its signals relayed via one or more inter-satellite links. Unlike the other mobile satellites of the time, the Iridium satellites provide for onboard digital packet processing and packet switching to enable the packets to be relayed via the inter-satellite links when needed. The satellites also use a band in the lower L-Band segment (1,616–1,626.5 MHz) in a burst mode. Technically, the Iridium satellite system was a success with a successful series of multi-satellite launches in 1997–1998. However, the original company with its approximate $5 billion investment was not a success financially and went into bankruptcy. Fortunately for users, the assets were bought by a group of private investors (for approximately $25 million), who provided for continued operation and development of the system.

Iridium is now building its generation NEXT that will consist of 70 satellites that is to be launched during the 2016–2018 time period to greatly expand the constellation's capabilities to provide voice, data (i.e., machine to machine), vehicular tracking, and search and rescue communications. These satellites are being manufactured by Thales Alenia and virtually all are planned to be launched on SpaceX Falcon 9 vehicles although the last two launches are scheduled to be by ISC Kosmotras Dnepr-1 vehicles. Launch failure review has delayed the deployment by a year.

The generation NEXT Iridium network will have hosted payloads that will constitute the Aerion network that will provide augmented air traffic management and control.

Globalstar

A different low earth orbit system with 48 satellites called Globalstar was deployed by Loral Corporation in 1998. Unlike Iridium, Globalstar uses simple bent-pipe repeaters on spacecraft flying at higher orbits (1,400 km vs. Iridium's 781 km) and a lower inclination (52°). The system design employs code division multiple access (CDMA), which allows seamless satellite to satellite handoffs. Access to Globalstar services is dependent upon proximity to a gateway earth station, since both the mobile terminal and a gateway must be within the field of view of the same satellites for access to work. Globalstar uses L-Band frequencies (in the 1,610–1,621.5 MHz band near those used by Iridium) for mobile to satellite links. For satellite to mobile transmissions, it uses a segment at S-Band (2,484–2495.5 MHz). This system, like Iridium, also experienced financial difficulties during the early years of its deployment and went through a period of filing for bankruptcy.

The Globalstar is deploying a second-generation constellation that consists of 32 low earth orbiting (LEO) satellites. The Globalstar satellite is simple; each consists of a communications system of both S- and L-Band antennas that are deployed from a trapezoidal body along with two solar arrays. Each of the second-generation satellites operates at an altitude of 1,414 km (approximately 876 miles). These satellites, as is the case with Iridium, are manufactured by Thales Alenia Space.

The satellites utilize a very simple "bent-pipe" architecture. On any given call, several satellites transmit a caller's signal. The satellites employ code division multiple access (CDMA) technology as was the case with the first-generation satellite services. Mobile calls are sent to a satellite dish at the appropriate gateway where the call is then routed locally through the terrestrial telecommunications system. This system relies on a much larger ground satellite network since it does not deploy complex inter-satellite technology onboard the satellites.

ICO

When Motorola and Loral announced plans for the Iridium and Globalstar systems in the early 1990s, Inmarsat carried out a study program of alternative orbits and technologies to support handheld mobile telephony. The study concluded that an intermediate orbit at an altitude of 10,390 km with 10 satellites at an inclination of 45° would provide superior satellite visibility for mobile users. However, the Inmarsat Council (governing board) would not approve the necessary investment by Inmarsat, so a separately funded corporation was created in 1995 named ICO. ICO was not able to launch its satellites before going into bankruptcy in 1999. The assets were bought by a private investor and a new company was created called New ICO. Only one of the ten satellites was successfully deployed before New ICO decided to abandon its plans for a non-geosynchronous constellation in favor of a geosynchronous satellite, which was launched in 2008.

Little LEO Satellite Systems

Some applications such as position reporting and status monitoring require only low data rates and can be supported by small low-power satellites. In 1993, Orbital Sciences announced the ORBCOMM system which was designed to address those needs with a constellation of small low earth orbiting satellites. The satellites were to enable communications with very low-power VHF transceivers using store and forward onboard message processing by the satellites for delivery to ORBCOMM gateway earth stations.

Geosynchronous Regional Satellite Systems

By the late 1990s, it was technically possible to design a geosynchronous satellite with a very large aperture antenna system capable of supporting handheld mobile satellite communications. Furthermore, because cellular radio coverage continued to have large geographical coverage gaps, there was a need for dual-mode cell phones that could also operate via satellite. Two major projects were initiated, the Asia Cellular Satellite System (ACeS) in 1995 and the Thuraya system in 1997. Both projects had a similar goal, to enable mobile users with GSM roaming subscriptions to have access to coverage via satellite when outside terrestrial GSM network coverage. In other words, the satellite system is to serve as an ancillary satellite component to the terrestrial networks.

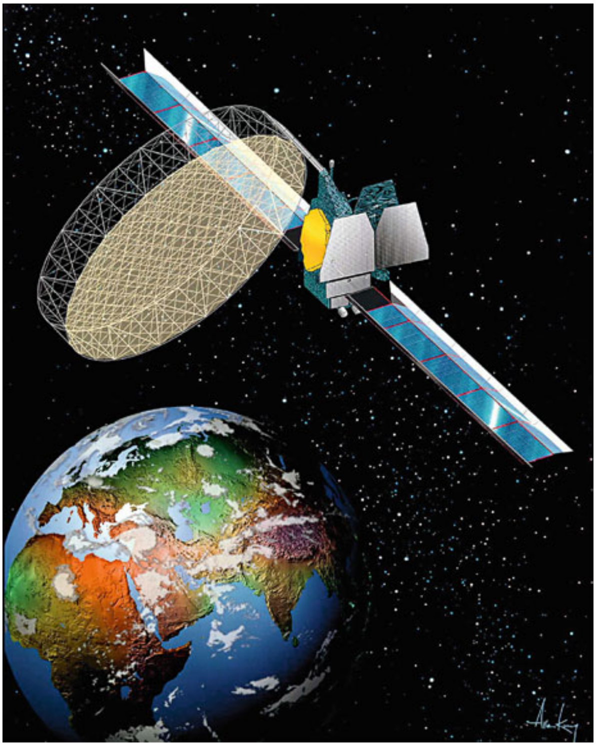
ACeS was designed to cover Indonesia, Malaysia, Thailand, Philippines, Sri Lanka, Vietnam, China, and part of India. The satellite was launched in 2000 and used two 11.9 m L-Band reflectors to enable 140 spot beams to fill the coverage area from an orbit position of 123° E longitude.

Thuraya (Fig. 4) was designed to cover most of Europe; the Middle East; North, Central, and East Africa; Asia; and Australia. The first of the Thuraya satellites was launched in 2000 and used a single 12 × 16 m reflector to enable over 200 spot beams to fill the coverage area from an orbit position of 44° E longitude.

Terrestrial Cellular Service Convergence with MSS

Wireless communications services have emerged as one of the largest growth engines in the telecommunications industry. Second-, third-, and fourth-generation (2G, 3G, and now 4G LTE) cellular wireless services have been leading growth over past decades, with annual rates exceeding 30 %. According to International Telecommunication Union (ITU) statistics released in 2015, there are estimated seven billion mobile cellular subscriptions worldwide. The ITU release explains that this number is misleading in that a very large number of people have duplicate subscriptions for business and personal use. As growth rates in Europe and USA have leveled off, Asia and Africa have become very attractive growth markets, with India and China leading the charge. Today more that half of all cellular subscribers are in Asia. To complement the terrestrial mobile service, a number of satellite-based global personal communications systems (PCS) have been deployed. These included Iridium and Globalstar (low earth orbit (LEO) systems) and ICO Global (medium earth orbit (MEO) systems and Asia Cellular Satellite System (ACeS) and Thuraya (geostationary or GEO) systems). Increasingly Inmarsat networks (especially with

Fig. 4 Thuraya mobile satellite (Graphic courtesy of Boeing Corporation)



the deployment of Inmarsat V and the Inmarsat Xpress networks) have become an important part of global mobile satellite services. All of these systems provided narrowband voice and data (narrowband) services to handheld terminals, although many of these handheld terminals required special antennas. Today Inmarsat and Thuraya can also provide broader band mobile services as well. Some of these systems have encountered major financial problems because of large system start-up costs and slower than required market growth due to a variety of challenges. These have included regulatory barriers and the more rapid than anticipated deployment of terrestrial cellular systems in both the developed world and the developing world (ITU 2015).

Recent decades have witnessed a rapid growth in Internet users and growth in broadband Internet traffic driven by demand for data, video, and multimedia services. According to ITU, the number of Internet users in 2014 surpassed three billion with 60 % of these being in developing countries. The largest growth contributors have been the mobile Internet subscribers who are rapidly taking over the fixed users (Lee et al. 2006). These trends have resulted in deployment of new integrated satellite and terrestrial networks using standard devices with form factors similar to current PCS/Cellular devices. Figure 5 shows the convergence between emerging wireless and mobile satellite services. Examples include deployment of S-Band and

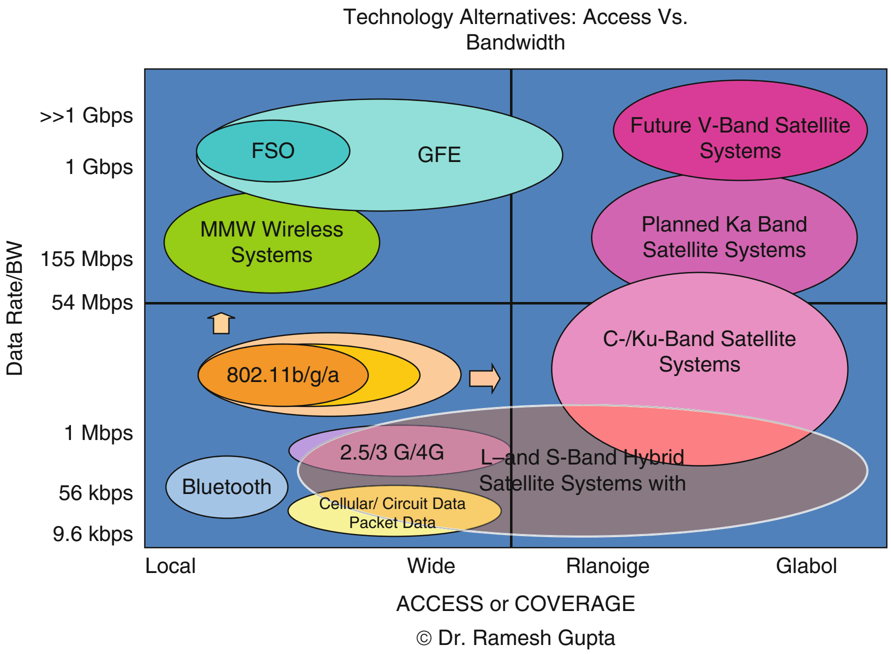


Fig. 5 Satellite and wireless service segments and role of hybrid mobile satellite systems with ancillary terrestrial segment (*ATC*) (*MMW* millimeter wave, *FSO* free space optics, *GFE* gigabit fiber Ethernet) (Graphic courtesy of R. Gupta)

L-Band integrated MSS networks in the USA by ICO Global Communications (DSDB recently acquired by Dish Network), TerreStar, and LightSquared.

Hybrid Transparent MSS Networks

Until recently, the MSS frequency bands were separate from the bands used for terrestrial cellular, so that the mobile user either needed a dual frequency band handset or two separate handsets. However, in the past decade, several system planners have proposed that segments of the MSS frequency bands be used for both terrestrial cellular and satellite communications, so that the handsets might be simplified and the user's service is always through the same service provider (Karabinis and Dutta). The terrestrial cellular network to support this mode of operation is called the ancillary terrestrial component (ATC). Although this will put additional burdens on the existing frequency allocations and require special precautions to protect GPS operations in adjacent bands, conditional approvals for concept have been already obtained in the USA.

Following the release of US regulatory framework guiding the MSS services with ATC, the European Communications Committee (ECC) decided to review the European regulatory framework for MSS in 1,980–2,010 MHz and

2,170–2,200 MHz bands. In the ECC decision of December 2006, the ground-based infrastructure operating in these bands is named complementary ground component (CGC) (ECC 2006).

Both FCC and ECC required that the MSS operators ensure that interference between these transparency ATC and CGC systems will be considered during intersystem coordination. The regulatory agencies believed that these MSS services will be beneficial to satellite operators and manufacturers, will enable innovative services including provision of emergency services, and will thus serve the public interest.

Mobile Satellite Systems with ATC

In the USA, the satellites for all three licensed systems have already been launched between 2008 and 2010. ICO Global Communications (ICO-G/DSDB) satellite built by Space Systems Loral (SS/Loral) was launched on April 14, 2008, and placed at 92.85° W longitude. TerreStar-1, another S-Band satellite built by SS/Loral, was launched on July 1, 2009, and placed at 111° W longitude. Ligado Networks launched its first L-Band satellite (SKYT-1) built by Boeing on November 14, 2010, and placed at 101.3° W longitude. A photograph of the SKYT-1 is shown in Fig. 6. The terrestrial segment development and deployment for all three systems is continuing with development of reference chipset designs and firmware for satellite-adapted versions of the chipset for mobile terminals. Terminal equipment for transparent integrated networks (consisting of MSS and ATC) targets a large consumer market, driving economies of scale for chipset as well as device manufacturing.

As an illustrative example, Ligado Networks next-generation system with MSS and ATC is discussed in this section, which consists of two integrated networks: a space-based network (SBN) consisting of satellites and four gateways and an ancillary terrestrial network (ATN) consisting of several ATCs in high-density population centers. The network is designed so that a single handheld device provides seamless two-way data services and voice through terrestrial as well as space segment. The space segment has sufficient antenna gain and EIRP with narrow beams to establish communication with the user devices. For example, SKYT-1 was designed with a 22 m deployable L-Band reflector which enables the feed element array to form hundreds of spot beams with increased frequency reuse. Network design functions permit seamless transition of device communication from satellite cells (beams) to terrestrial cells, thus achieving transparency to the end user. Satellite provides wide area coverage and is suitable for low-density population areas, where terrestrial infrastructure may not be cost-effective. The terrestrial component on the other hand ensures availability of high-speed broadband services in major population centers at affordable cost. This hybrid system offers the availability advantage for seamless communications in situations like earthquakes or hurricanes when the terrestrial infrastructure may be disabled (e.g., hurricane Katrina in 2005 and earthquake followed by Tsunami in Japan in 2011). The satellite capacity and resources

Fig. 6 Ligado's SkyTerra-1 mobile satellite (Graphic courtesy of Ligado Networks)

(Bandwidth and Power) can be reallocated in case terrestrial infrastructure is disabled by natural or man-made disasters. The recent decision by the US FCC in February 2012 to suspend Ligado Networks (Previously LightSquared) ability to operate as a high-speed wireless terrestrial network in the USA in the requested frequency band that they had intended to operate – due to potential interference with the GPS satellite navigation service – has had a major consequence for the company. In fact LightSquared was forced to declare bankruptcy. Additionally, in December 2015 the company emerged from bankruptcy, and in February, 2016 rebranded as Ligado Networks. This action was triggered in part by the FCC action. The ICO/DBSD North America satellite facilities and TerreStar have been bought out of bankruptcy by DirectTV, and these satellites are being used today to support broadcast satellite services.

A simplified illustration of a next-generation L-Band MSS/ATC hybrid network is illustrated in Fig. 7. The space segment depicted consists of two L-Band satellites in geostationary orbit, each one with an aggregate EIRP of 79 dBW and G/T of 21 dB/K over a minimum coverage area of 95 % so that handheld cellular/PCS devices can close the forward and return links with some margin to spare. The second satellite acts as an in-orbit spare and also provides additional gain for faded

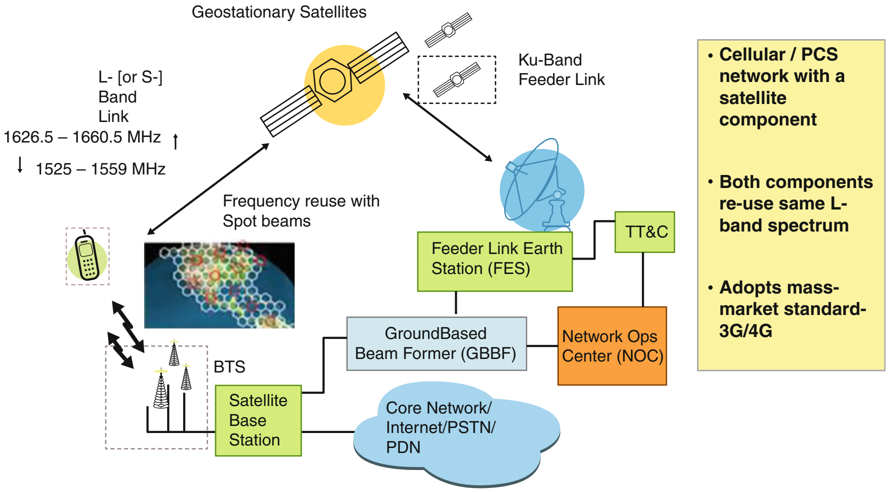


Fig. 7 Simplified block diagram of hybrid MSS with ATC (Graphic courtesy of R. Gupta)

mobile channels (because of multipath effects and shadows) through diversity combining. Four satellite gateways at Napa and Dallas (USA) and Ottawa and Saskatoon (Canada) are connected to satellites via Ku-Band feeder links and also provide connectivity to the terrestrial networks. These gateways serve as communications nodes and are connected to the same core public switched telephone network and public data network (PSTN/PDN), thus enabling seamless communications between the satellite and the terrestrial networks.

A new technology is the ground-based beam former (GBBF), which digitizes the beam signals and divides them into component beams so that appropriate beam weights can be applied digitally. The amplitude and phase beam weights are used together with the onboard satellite feed array, to form beams with maximum gain over the desired coverage and also to suppress side lobes for minimum co-channel interference as well as to suppress ATC signal-induced interference. The beam forming is also used to minimize interference to and from other systems using same or adjacent L-Band frequencies. The ability and flexibility offered by the GBBF for formation of multiple spot beams with interference suppression is a key design feature of the MSS/ATC networks that enables to increase the spectrum usage efficiency. The network operations center (NOC) monitors the satellite and terrestrial traffic and allocates bandwidth, frequency, and bandwidth resources to minimize interference.

The satellite network is designed to be largely independent of the air interface standards. Therefore, the air interface can be selected based on terrestrial offerings including 4G/5G LTE air interface. The satellite adaptation of the air interface requires the satellite gateway to compensate for satellite delay, and Doppler shifts (because of satellite motion), making the MSS/ATC network independent of air interfaces.

The MSS/ATC integrated network approach is quite different from existing Thuraya, Iridium, and Globalstar MSS networks in that these systems use different frequency bands to provide interworking between terrestrial and satellite networks. A handset for use in an MSS/ATC transparent network is designed to use the same frequency and air interface, so the handset is less complex and therefore likely to be less expensive.

FSS Convergence with MSS

VSAT systems operating in the FSS frequency bands were used for offshore maritime applications, especially in the oil exploration industry. With the development of low-cost antenna stabilization system, VSAT systems were introduced onto cruise ships operating within the coverage areas of domestic and regional FSS satellites. Subsequently, a VSAT became a serious alternative to an MSS portable terminal when the user needed a high data rate link, since bandwidths available in the FSS bands are much wider than in the L-Band MSS allocations.

Wide proliferation of relatively inexpensive and easy to install VSAT terminals with data rates exceeding 2 Mb/s resulted in integration of multimedia services, Internet and video distribution, IP video conferencing, video streaming, and IP multicasting services in the FSS satellite networks. These developments and trends have had a significant impact on the architecture of both FSS and MSS communications networks, capable of providing “global services,” wide area coverage, and emergency rapid response services at relatively low cost (Evans 1994). For example, communications satellites played a major role for disaster recovery operations during tsunami in Asia (2004), aftermath of hurricane Katrina in the USA (2005), and also during major earthquake and tsunami in Japan (2011). Regulatory agencies have recognized the critical role of satellites in the global telecommunications infrastructure and have approved new architectures and systems with convergence between the various satellite and terrestrial radio services.

A critical factor in the evolution of radio-communication services for mobile users will be the availability of bandwidth to support higher data rates. Even Inmarsat, with its history rooted in the L-Band allocations, is recently ordered a fifth generation of satellites that is to operate in the 20–30 GHz bands to enable much higher data rates to be supported not only for portable terminals but mobile terminals as well, such as those on aircraft. However, it should be noted that Inmarsat is not abandoning its L-Band user community and is continuing to also develop follow-on L-Band MSS satellite capacity.

Conclusion

MSS systems implemented over the past four decades have significantly improved the communications capabilities for the mobile user communities. Vital maritime, aeronautical, and remote area services will continue to be provided by reliable global

MSS system operators and several ongoing regional MSS operators. For land mobile services, 3G wireless services continue to be made available to areas outside terrestrial cells with roaming onto MSS satellite systems. With the deployment of ATC, users will have an option when to obtain seamless satellite and terrestrial access from a single device and service provider. For users with very high data rate requirements, more system choices are likely to be introduced with more portable and mobile terminals.

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ Satellite Communications Video Markets: Dynamics and Trends

References

Aerion News Release. http://www.Aireon.com. Last accessed 12 Dec 2015

Airbus Defence and Space signs contract with Inmarsat to build two next generation mobile communications satellites, https://airbusdefenceandspace.com/newsroom/news-and-features/airbus-defence-and-space-signs-contract-with-inmarsat-to-build-two-next-generation-mobile-communications-satellites/. Last accessed 24 Dec 2015

ECC Decision of 1 December 2006 on the designation of the bands 1980–2010 MHz and 2170–2200 MHz for use by systems in the Mobile-Satellite Service including those supplemented by a Complementary Ground Component (CGC), (ECC/DEC/(06)09) (2007/98/EC) amended 5 Sep 2007

J. Evans, Satellite and personal communications, in *15th AIAA International Communications Satellite Systems Conference*, San Diego, Feb/Mar 1994, pp. 1013–1024

B. Gallagher (ed.), *Never Beyond Reach – The World of Mobile Satellite Communications* (International Maritime Satellite Organization, London, 1989)

R.K. Gupta, Low cost satellite user terminals: lessons from the wireless industry, in *22nd AIAA Conference on Satellite Communications Systems*, Monterey, June 2004

Iridium Next, https://www.iridium.com/network/iridiumnext. Last accessed 31 Dec 2015

ITU Releases Facts and Figures (2015), https://www.itu.int/net/pressoffice/press_releases/2015/17.aspx. Last accessed 14 Jan 2016

P. Karabinis, S. Dutta, Recent advances that may revitalize mobile satellite systems. Light squared, www.lightsquared.com. July 2003

S. Lee et al., The wireless broadband system for broadband wireless Internet services. IEEE Commun. Mag. 106–113 (2006)

D.W. Lipke, D.W. Swearingen, Communications system planning for MARISAT. International conference on communications, Minneapolis, June 1974, IEEE Cat. No. CHO859-9-CSCB“Mobile Subcriptions Near the 7 billion-Mark. Does Everyone Have a Phone?” ITU News, No. 6, Nov/Dec 2015, https://itunews.itu.int/en/3741-Mobile-subscriptions-near-the-78209billion-markbrDoes-almost-everyone-have-a-phone.note.aspx

E.J. Martin, L.M. Keane, MARISAT-Gapfiller for navy satellite communications. Signal Mag. (Nov/Dec. 1974), pp. 6–12

E. Reinhart, R. Taylor, Mobile communications and space communications. IEEE Spect. **29**(2), 30–33 (1992)

D.W. Swearingen, D.W. Lipke, MARISAT multiple access capabilities, in *International Conference on Communications*, Philadelphia, June 1976, IEEE Cat. No. 76CH1086-0CSCB

D. Whalen, G. Churan, The American mobile satellite corporation space segment, in *14th AIAA Conference on Satellite Communications*, Washington, DC, March 1992, pp. 394–404

Store-and-Forward and Data Relay Satellite Communications Services

Joseph N. Pelton

Contents

Abstract

Commercial telecommunications satellite systems that provide fixed (FSS), mobile (MSS), or broadcasting (BSS) satellite services provide essentially "real-time" communications to satisfy the market needs of their commercial customers or support military communications requirements. The service is not precisely simultaneous in that there is close to a quarter second delay in the case of a satellite relay that travels from Earth to a geosynchronous satellite and then back to Earth. For normal commercial satellite services to support voice, data links, radio or audio channels, videoconferencing, or television service, the satellite link is provided on as close to a real-time basis as is technically possible.

There are, however, a variety of communications satellite services that are variously known as store and forward, business-to-business (B2B) relay, machine to machine (M2M), or data relay satellites. These types of "data relay" satellite services are typically not as instantaneous as is the case with the big three services – namely, fixed, mobile, and broadcasting. This type of service is usually

---

J.N. Pelton (✉)
International Space University Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_94

machine-to-machine data relays, and thus some delay in the transmission is usually not important.

Thus, what makes these types of satellite offering different is that there can be an acceptable time delay in the satellite data relay service. This delay, depending on the nature of the satellite service, can range from less than a second, to minutes, and to even hours. These various types of data relay satellite services will be addressed in this chapter. There is actually a wide variety of these satellite services that can also operate in different frequency bands. These diverse services and satellite types are designed to meet rather different types of communications and networking services. Some data relay satellites are very simple, small, and low-cost satellites that support amateur radio or volunteer efforts. Others are much more complex and actually support commercial customers. Yet others are designed for satellite-to-satellite interconnection and can be large, complex, and rather costly satellites.

Among these various types of data relay or machine-to-machine (M2M) services are the following: (i) amateur radio relay that is provided via so-called OSCAR satellites in the amateur radio band to allow global AMSAT connectivity and (ii) data networking using small satellites in various types of LEO or MEO orbits, or constellations, to provide non-real-time data relay services, often of a public service nature. Yet, this can also support commercial B2B or M2M services such as the Orbcomm satellite system. This commercial satellite system was designed to provide a minimum gap in connectivity and carry out such functions as near real-time tracking and communications with vehicles and ships and (iii) data relay services from GEO orbit to allow broadband communications with satellites or spacecraft in low or medium Earth orbit – or even UAV surveillance systems. These GEO-based data relay satellites are able to track and connect with lower orbit satellites and thus relay data from such satellites with minimum delay to ground communications centers half way round the world in close to real time. Such types of data relay satellites can be used to connect to spacecraft with passengers on board. These data relay satellites, such as NASA's Tracking and Data Relay Satellites, were used to support flights during the US Apollo moon program and then with the Space Shuttle. These TDRS allowed NASA to maintain connection to ground control facilities on close to a real-time basis. The different technical aspects of the various types of store-and-forward or data relay satellite systems, the frequencies they utilize, and the various types of services they support are all addressed in this chapter.

The common denominator for these diverse types of satellite services is that they are not real time but rather involve some elements of time delay. In the case of the most sophisticated data relay satellites, the connection may involve a delay on the order of a second. In the case of the most basic and low-cost store-and-forward satellite systems, the delay may be a period of several hours from the initial uplink to the ultimate downlink of the data message. Today, new types of store-and-forward data relay satellites can be quite sophisticated and high-cost systems that can handle high data rates. These new type of data relay

satellites have progressed a long way forward in terms of data throughput capabilities and can be more than a thousand times more capable than the first types of simple data relay satellites of the 1960s and 1970s. These much more broadband and sophisticated data relay satellites also operate in many higher frequency bands.

Keywords

Alphasat • AMSAT • ARTES satellite program • Business-to-business (B2B) data relay • Copernicus satellite program • European Space Agency (ESA) • European Data Relay System (EDRS) system • Japanese data relay satellite • Japanese Aerospace eXploration Agency (JAXA) • Ka-band • Ku-band • Machine-to-machine (M2M) data relay • NASA • NFIRE • Orbcomm • OSCAR • S-band • Sentinel program of Europe • Store-and-forward satellite system • TerraSAR-X • Tracking and Data Relay Satellite System (TDRSS) of NASA • Unmanned aerial vehicles (UAVs) • UHF band • University of Surrey • UoSAT • Very high frequency (VHF) band

Introduction

As noted above, one of the first instances of the use of a store-and-forward satellite system involved the amateur radio community that envisioned that amateur radio signals could be sent up to an orbiting satellite, stored, and then relayed to other parts of the world at a later time. In this instance, the signal did not have to be bounced off of the ionosphere but rather relayed by a low Earth orbit satellite. This satellite used the amateur radio band.

The success of the so-called OSCAR amateur radio satellite led to the idea that small satellites in a low Earth orbit constellation of only a few satellites could create a global network for relaying messages around the world. This approach to global data relay could provide an option to volunteer, rescue, and economic development organizations with worldwide services working in remote areas. This involved both governmental aid and nongovernmental organizations (NGOs) such as those providing medical, health, education, or aid services in areas without established communications services. The University of Surrey Space Centre in the UK was established to design and build small-class satellites that could provide telecommunications or remote sensing satellites and was among the first to design and build such satellites.

The UoSAT series designed at the University of Surrey were thus among the earliest small store-and-forward message relay satellite to be deployed. The first simple UoSAT-1 was built over a 30-month period by students and engineers for only about £250 pounds (sterling). It was launched for free by NASA in September 1981 (Watkins 2014).

These UoSAT-1 and UoSAT-2 were launched into a polar orbit so that they could provide a global communications network capable of carrying electronic mail, digitized voice, computer data, or even images. These small polar-orbiting satellites were

Fig. 1 The UoSAT-1 was the world's first store-and-forward data relay satellite (Graphic Courtesy of the Surrey Space Centre)

launched with sufficient power and antenna capability so that they could connect to low-cost and small-aperture ground terminals, although the data throughput was limited. The UoSAT-2 satellite, designed and built at the Spacecraft Engineering Research Unit at the University of Surrey, was constructed and launched in 1984 in cooperation with the Volunteers in Technical Assistance (VITA) and the Amateur Satellite group known as AMSAT (Uosat-1 Image) (See Fig. 1).

The initial UoSAT-2 store-and-forward transponder was of very low power and of very limited bandwidth but has amazingly managed to operate now for over 30 years. The global store-and-forward capability of this early experimental satellite included only an eight-bit central processing unit (CPU), and its message storage capability was limited to only 96 kb. This was enough capability to test that this type of system could operate effectively worldwide. This led to a number of much more capable store-and-forward small satellites. The UoSAT-3, for instance, was launched in January 1990 with a 16-bit CPU and an 8-MHz transponder. A number of such satellites are followed with increasing processing and message relay capability (Small Satellites).

One such follow-on project carried out via the Surrey Space Centre was funded by NEC of Japan. This was a two-satellite polar-orbiting store-and-forward satellite

(Lifesat) that supported global medical services to remote areas. It allowed remotely located doctors to request information from medical books and journals and stay in touch, albeit with a delay of about 2 h due to the store-and-forward message relay of information. Clearly, this was not as desirable as a direct "real-time" telephone connection via satellite, but it was much superior to having no communications capability at all.

Today the Surrey Space Centre has been acquired by Astrium-AirBus, and this facility is able to design and build store-and-forward CubeSats, such as one to six unit CubeSats, up to highly capable small satellites. Recently, the Surrey Space Centre, using essentially off-the-shelf commercial components, designed, built, and arranged for the launch of a nanosatellite that included a "smart" cell phone and an alcohol-based propellant system to assist with deorbit (Surrey Space Centre).

Space researchers plus student volunteers at the Surrey Space Centre (SSC) and SSTL developed STRaND-1, a 3-unit CubeSat containing a smartphone payload plus microthrusters for deorbit that was successfully launched into orbit in 2013. The STRaND-1 carries an amateur radio unit that operates on a packet radio downlink of 437.568 MHz (See Fig. 2). Frequency and telemetry information for STRaND-1 is provided on the AMSAT-UK website (STRaND Nanosatellite Smartphone).

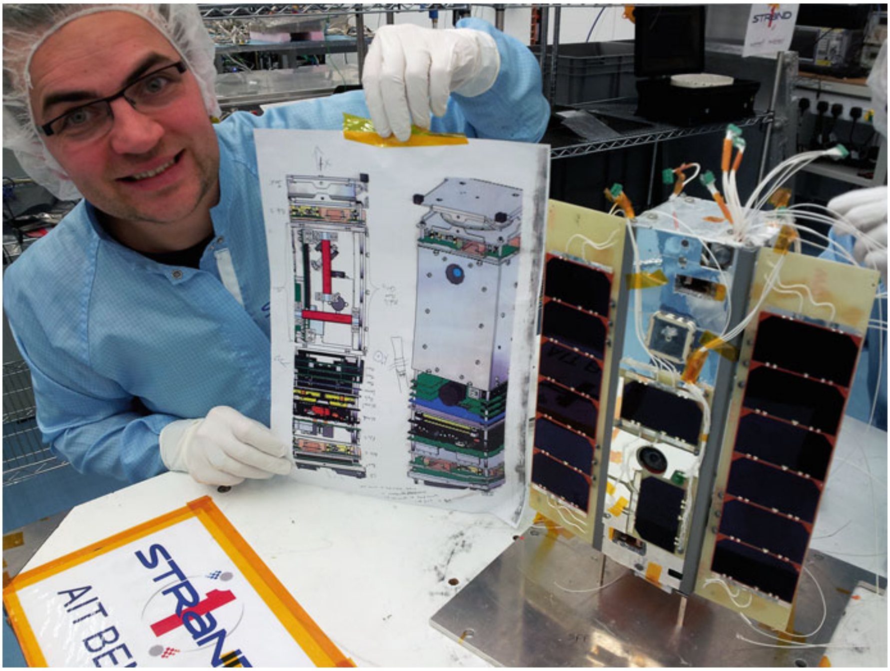


Fig. 2 The STRaND-1 nanosatellite built in 2015 at the Surrey Space Centre (Graphic Courtesy of the Surrey Space Centre)

Store-and-Forward Commercial Systems: The Orbcomm Inc. Constellation

The success of the AMSAT Oscar satellite and the small UoS satellites from the University of Surrey Space Centre led to more ambitious store-and-forward small satellite constellations. Perhaps the most ambitious of these is the so-called Orbcomm satellite network. This project was started by the Orbital Sciences Corporation (now Orbital ATK) and Teleglobe of Canada in the 1990s that deployed the original satellite constellation in the mid-1990s. This store-and-forward satellite system was spun off from Orbital when the network filed for chapter "▶ Ground Systems for Satellite Application Systems for Navigation, Remote Sensing, and Meteorology" bankruptcy in 2000. This company has reorganized and is now an entirely separate company known as Orbcomm, Inc., that issued public stock in 2006 (Orbcomm Inc.).

The first generation of satellites was designed by the Orbital Sciences Corporation. This network consisted of very compact Orbcomm global satellites, and the Orbital Sciences Corporation's Pegasus launch system was used to put them into orbit. The first-generation OG1 satellites were launched in the 1995–1998 time frame, of which 35 were successfully launched. Each of these satellites could be stacked together in a flat position, and each of them weighed only 42 kg (93 lbs). Two disk-shaped solar panels could pop-up from a stowed flat configuration after launch. These small round solar arrays were designed to track the sun and provide up to 160 W of power. Of the original 35 satellites in the low Earth orbit constellation, some 29 remain in service nearly 20 years later (Orbcomm System Overview).

Communication to and from subscribers, such as units installed in rental cars, trucks, and buses, was provided at the relatively slow rate of 2400 bits/s for the uplink and 4800 bits/s for the downlink. This allowed data messages to be sent from essentially anywhere on the planet and typically connect in about 6–10 min. In the case of the smallest, 1-byte (8-bit message) connection might typically be accomplished within 1-min time. This is the only commercial satellite that operates in the very small very high frequency (VHF) 137–150 MHz band that is allocated to little low Earth orbit (LEO) satellites. This VHF band allocation is only 13 MHz across and thus can only accommodate low-speed messaging. Every satellite in the Orbcomm network has an onboard GPS receiver for position determination, as do all of the ground antenna systems. Typical data relays for the first-generation Orbcomm system are 6–30 bytes in size. This brief satellite relay message system is still adequate for sending GPS positioning data or simple sensor readings.

Several such systems were planned in the early to mid-1990s. A Russian-based system named Courier was designed to operate in the 1.5 and 1.6 GHz bands that would start out as a store-and-forward data relay system and then, after building up the number of satellites in the constellation, convert to a real-time voice service that could compete with Iridium or Globalstar. Financing and problems with obtaining

US-based software eventually ended this project. Financing issues likewise ended the other erstwhile competitors. Ultimately, Orbcomm was the only such commercial data relay system to successfully launch. For the first generation of Orbcomm, independent tests showed well that over 90 % of the text messages were transferred within 6 min, but gaps between satellites could result in delays in message delivery of 15 min or more.

After emerging from bankruptcy and receiving new financing, the contract for the second-generation satellites for the system was awarded to Sierra Nevada by the newly reconstituted Orbcomm Inc. This second-generation satellite was more capable than the first, but maintained the very thin and modest weight characteristics that allowed six satellites to be launched at the same time (Orbcomm Second Generation).

With the full deployment of the 18 satellites in the second generation of the Sierra Nevada-built Orbcomm satellites, plus the 29 satellites in the first generation, there is robust coverage of the entire globe. This means that there is likely to be a satellite within range of almost any spot on Earth where an Orbcomm terminal might be operational. Therefore, data relays via the older and newer satellites in the overall constellation involve only the most minimal of time delays.

There are many examples of Orbcomm applications. Many of the services include the transportation industry and the relay of tracking or sensor data from rental cars, touring buses, delivery trucks, and cargo and luxury liner ships. There are a particularly large number of tracking services related to truck trailers and shipping containers and cargo security which can be carried out on a global scale. The other large application relates to the monitoring of remote equipment such as oil and gas rigs, equipment being remotely operated by supervisory control and data acquisition (SCADA) networks, as well as programs involved with data collection from scientific equipment located in remote or difficult-to-access locations. One of the most popular services now available involves the active tracking of vessels at sea using the Automatic Identification System (AIS).

On 3 September 2009, Orbcomm announced the selection of SpaceX (Space Exploration Technologies) to launch 18 second-generation satellites via Falcon launch vehicles. Initially, the plan was to launch via Falcon 1e rockets, but on 14 March 2011, it was announced that SpaceX would use Falcon 9 to carry the first two Orbcomm next-generation OG2 satellites to orbit in 2011. On 7 October 2012, the first SpaceX Falcon 9 launch of a prototype second-generation Orbcomm communications satellite failed to achieve proper orbit. Orbcomm thus filed a $10 million claim for this loss as the satellite burned up as it deorbited on 10 October 2012. But Orbcomm, despite this launch failure, continued to be confident in their launch arrangements with SpaceX. This has now paid off in the years that followed. On 14 July 2014, SpaceX used the reengineered Falcon 9 launcher to successfully launch six of the second-generation satellites (Peter de Selding 2015).

The launch of the remaining 11 satellites (the first of the series was lost in the March 2011 launch) for the completion of the next generation has now been achieved using the upgraded Falcon 9 launcher (OG2).

Tracking and Data Relay Satellites

Some of the most significant infrastructures routinely in use by space agencies are the data relay satellites that began with the deployment of the NASA Tracking and Data Relay Satellite System. In the earliest days of the space programs of the 1960s and 1970s, it was necessary to create tracking, telemetry, command, and monitoring Earth stations all around the world in order to stay in constant touch with rocket launchers and spacecraft during their launch and low Earth orbit operations. This meant operating expensive facilities with a large number of staff members around the world and even on shipboard-based antenna systems.

As early as 1966, engineers at NASA proposed the idea of being able to replace these many large tracking antenna ground facilities that were expensive to create, maintain, and operate, sometimes in locations that were subject to attack or looting raids, with a space-based system. NASA Administrator James Webb in 1967 agreed to proceed with the implementation of what became the NASA Tracking and Data Relay Satellite (TDRS) system despite being somewhat skeptical that the large, expensive, and sophisticated three geosynchronous network could achieve the demanding technical goals of tracking and relaying signals from spacecraft or launch systems in low Earth orbit or achieving the projected cost savings. This program was formally approved and put into action in 1973 and it was a decade before the first satellite in the TDRS network was launched and put into operation to support the Space Shuttle program.

The TDRS that was manufactured by the TRW Corporation turned out to be very difficult to design, build, and deploy. Initially, six of these satellites were ordered and a seventh added when there was a loss of the second of these satellites due to launch failure (See Figure 4).

It is not surprising that this technologically challenging system cost significantly more than first projected. Nevertheless, it did work very well and proved definitively that space-based tracking and data relay from LEO satellites to GEO and back to Earth for reception by processing and telecommunications centers could work effectively. The validation of the NASA TDRS system set the stage for subsequent generations of NASA TDRS. Since that time, a second generation of NASA TRDS was deployed, and now the third generation has become fully operational.

The system was designed to replace an existing network of ground stations that had supported all of NASA's manned flight missions. The prime design goal was to increase the time spacecraft that were in communication with the ground and improve the amount of data that could be transferred and also maintain continuous links with manned spacecraft such as Space Shuttle flights when they were operational. The earlier TDRS spacecraft were launched by Atlas IIA launchers, while the latest generation of spacecraft was launched by Atlas V rockets (DTRS Project Overview).

The third generation of NASA's Tracking and Data Relay Satellites, known as TDRS-K, TDRS-L, and TDRS-M, has been manufactured by Boeing Inc. The first satellite in this third-generation TDRS-K was launched in 2013; the second, TDRS-

L, in 2014; and the third, TDRS-M, in 2015 (National Space Science Data Center). Each of these satellites has a 15-year projected operational life. In addition, to telemetry, command, and mission data communication services, TDRS-K, TDRS-L, and TDRS-M will continue to provide tracking data used to determine the orbit and specific location of user satellites.

These satellites are now being used to relay the data from the Hubble Space Telescope and NASA Earth observation spacecraft to processing centers on the ground in as close to an instantaneous basis as possible. It can also communicate with the International Space Station as well. This is a much more efficient download process rather than waiting until the lower-orbiting spacecraft are over the ground communications and processing center and trying to download all the huge amount of data in a short span of only a few minutes. The various technical characteristics for the third generation of the NASA TDRS network, in terms of systems capabilities operating in different frequency bands, are provided in Table 1 below (DTRS Spacecraft Payload Capabilities).

The success of the NASA TDRS network also led to a parallel Japanese capability. The Japanese space agency (NASDA), and now in its current form JAXA, has created and operated its own TDRS system as well. Finally, it has also led to the European Space Agency (ESA) data relay program. ESA has designed its new state-of-the-art system based on laser communications relay. These are discussed below.

Table 1 The various capabilities of the latest generation of the NASA TDRS network

Subsystems of the NASA tracking and data relay satellite network	
Frequency band	Subsystem
S-band	**S-band multiple access:** the phased array antennas on the TDRS-K, TDRS-L, and TDRS-M are designed to receive signals from up to five spacecraft simultaneously and transmit to one at a time. Improvements in the multiple access performance and onboard processing have contributed to an increased return data. The third-generation forward (ground-to-space) service transmitting power for TDRS-K, TDRS-L, and TDRS-M is also increased
S-band	**S-band single access:** two 5-m diameter mechanically steerable antennas providing high-gain support to satellites with low-gain antennas or multiple access user satellites temporarily requiring an increased *data rate*. The antennas support manned missions such as the International Space Station; science data missions, including the Hubble Space Telescope; and satellite data downloads from Earth observation satellites
Ku-band	**Ku-band single access:** the two large antennas also operate at a higher frequency band supporting two-way high-resolution video and customer science data
Ka-band	**Ka-band single access:** two single access antennas in the Ka-band provide even broader band services for large volumes of science data. This frequency allows users to transmit data at up to 800 Mbps. Originally established on the TDRS-H, TDRS-I, and TDRS-J spacecraft, the Ka-band frequencies allow for continued international compatibility with Japanese and European space relay programs, enabling mutual support in case of emergencies

The Japanese Tracking and Data Relay Satellite System: Kodama

The Kodama TDRS network was launched on 10 September 2002 (See Fig. 3). This satellite was launched via a Japanese HIIA rocket. It has now operated successfully for over 13 years in orbit (Kodama on-orbit operations).

The Kodama successfully performed a data relay experiment involving remote sensing satellites and the relay of observed data to processing centers as quickly as

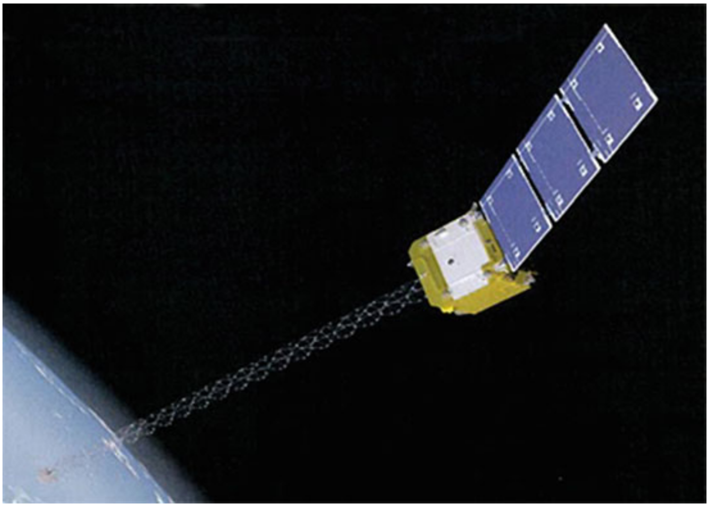


Fig. 3 The second-generation Orbcomm satellites (Graphics courtesy of Sierra Nevada)

Fig. 4 The first-generation tracking and data relay satellite (*TDRS*) that was a technical marvel (Graphic Courtesy of NASA)

possible after the data was acquired. This data relay experiment operated at the extremely rapid data transfer rate of 278 Mbps. This test was carried out using the data being transferred from the Advanced Land Observing Satellite known as "Daichi" and also other remote sensing satellites at lower speeds. In the highest-speed data transfers, data related to Daichi's global land observations and particularly with regard to disaster monitoring were related via Kodama to processing centers in Japan. Some 5 % of 6.5 million images taken by the Daichi remote sensing satellite – the equivalent to about one petabyte of data – were successfully relayed by the Kodama satellite at a very broadband speed of 278 Mbps in the 2002 time frame.

In many cases, the observed data from remote sensing satellites can simply be stored on the satellite and downloaded when traveling in the part of its orbit where it can provide a direct feed to processing centers. There can be situations such as crucial observations over disaster sites where vital information from the recovery areas is needed with great urgency. It is in these circumstances that the rapid data relay via Kodama can speed up the process by perhaps 1 or 2h (See Fig 5 and 6).

When a direct communication link is used to download data from a remote sensing Earth observation satellite, the total contact time between a low Earth orbit spacecraft and a ground station is limited to approximately 10 min per visible pass over Japan. Kodama relays data with much less restriction. It is able to provide a link between a LEO spacecraft and a receiving ground station for as much as 60 % of the flight path of the spacecraft. This is equivalent to close to 1 h out of approximately 90-min orbit. This enables Japan, despite its rather small geographic size, to extend contact time between its spacecraft and a small number of ground stations by up to six times and allow the crucial transfer of data when urgently needed during catastrophes. Although the relay involves some delay, it still constitutes a near-instantaneous connection (Kodama on-orbit operations).

Fig. 5 The Japanese tracking and data relay satellite known as Kodama (Graphic Courtesy of JAXA)

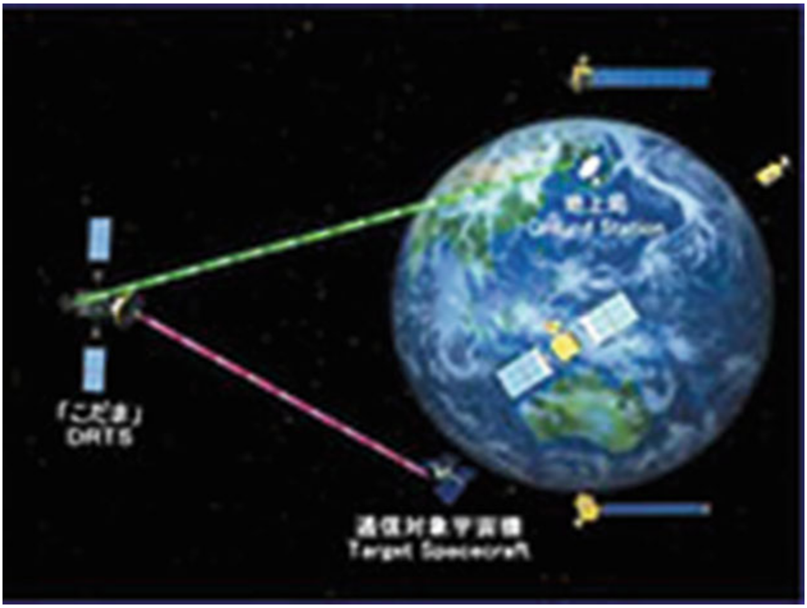


Fig. 6 Illustration of Kodama in GEO orbit connecting Daichi LEO satellite with Japanese processing center (Graphic Courtesy of JAXA)

The European Data Relay System

The European Data Relay System (EDRS) system is a planned constellation of GEO satellites that will relay information and data between satellites, spacecraft, UAVs, and ground stations. The designers intend the system to provide almost real-time and very broadband communication relays, even if the relay includes satellites in low Earth orbit. This system will allow ESA, as was the case with NASA's TDRSS network, to reduce substantially its need to maintain continuous visibility from its ground tracking, telemetry, and command Earth stations.

This program is being developed as one of the parts of the so-called ARTES 7 program. As now defined, EDRS is intended to be an independent European satellite system that reduces time delays in the transmission of large quantities of data from low and medium Earth orbit remote sensing satellites and other satellites to ground processing and communications centers.

The design is similar in concept to the US NASA Tracking and Data Relay Satellite System that was set up to support the Space Shuttle. There is one notable and significant technical difference in that EDRS will use a next-generation Laser Communication Terminal (LCT) technology to achieve extremely broadband throughput capability. The laser terminal for the EDRS is capable of transmitting at exceptionally broadband speeds up to 1.8 Gbps. It is also significant in that it is able to operate across the exceptionally long relay link of some 45,000 km that is associated with the distance involving a LEO to GEO and then back to Earth for processing or feeding into terrestrial communications links. Such a terminal was successfully tested during in-orbit verification between the German radar satellite TerraSAR-X and the American NFIRE satellite. In addition, the commercial telecommunication satellite Alphasat has been used to perform further system and operational service tests and demonstrations as part of the European Copernicus program.

EDRS is being implemented as a public-private partnership (PPP) between the European Space Agency (ESA) and Astrium, which is a part of AirBus.

ESA funds the infrastructure development and is the anchor customer that operates through the Sentinel satellite missions. Astrium will carry the overall responsibility for the implementation of the space segment including launch, as well as the ground segment. Astrium will then take over ownership of EDRS and will provide the data transmission services to ESA and customers worldwide on a contract basis.

Conclusion

The concept of a data relay satellite has evolved at a very rapid and significant way since the 1970s. Initially, the concept was to develop a very low-cost satellite that could relay very low-bandwidth text messages to support amateur radio operators. These initial low-cost and low Earth orbit satellites were only capable of relaying a few bytes of information. The only commercial satellite system that utilizes this concept is the so-called “Orbcomm” constellation that provides very small messages such as tracking and sensor information, largely for mobile vehicles and ships. The second generation of this type data relay system is now being deployed.

The next stage of the evolution of data relay satellites involved a significant change in the technology and the utilization of higher frequency bands that allowed the transfer of much larger amounts of data traffic at much higher rates to support broadband services including high-definition hyper-spectral imaging.

These two types of data relay satellites are thus quite different. The small, low-cost narrow band machine-to-machine-type satellites in LEO are often projects of developing countries and student experimental activities. The GEO-based broadband satellites that are designed to track satellites in LEO or MEO and relayed at high data rates back to GEO and then immediately down to Earth-based communications and processing centers are typically projects of space agencies and are quite expensive and technically sophisticated satellites. Currently, NASA in the USA, JAXA of Japan, and ESA have such tracking and data relay systems.

The first type might involve significant delays of up to an hour or more which is not a problem for machine-to-machine relays. The GEO-based tracking and data relay satellites can and typically do provide near to instantaneous relay and very broadband speeds. The latest generation of NASA TDRS can operate at speeds up to 800 Mbps. The European Data Relay System is projected to operate at speeds up to 1.8 Gbps.

Table 2 Summary chart of data relay satellites

Various types of data relay satellites		
Type	Frequency band	Orbit
Amateur satellite-OSCAR	(VHF) narrow band	LEO
UoSAT, AMSAT, and other small data relay satellites	VHF and UHF narrow band messaging	LEO
STRaND-1 cell phone sat	437 MHz	LEO
Orbcomm Inc.	VHF narrow band	LEO constellation (Gen. 1 and 2)
NASA TDRS (first generation)	S-band	GEO
NASA TDRS (third generation)	S-band plus Ku- and Ka-band with throughput up to 800 Mbps	GEO
Japan TRDS – Kodama	S-band plus Ku and Ka Band throughput up to 270 Mbps	GEO
ESA-Astrium DRS	S-band and laser throughput up to 1.8 Gbps	GEO

In addition to these systems providing data relay for the International Space Station (ISS), the Earth observation, and the Hubble Space Telescope, these sophisticated networks can help support the near real-time relay of ground images of areas that have experienced disaster events to assist rescue workers cope with things like volcanic explosions, hurricanes, typhoons, earthquakes, and similar crisis-based situations.

A summary chart, related to the various types of data relay satellites, is provided below (Table 2).

Cross References

▶ Space Telecommunications Services and Applications

References

DTRS Project Overview, http://tdrs.gsfc.nasa.gov/. Last accessed 5 Dec 2015

DTRS Spacecraft Payload Capabilities, http://tdrs.gsfc.nasa.gov/assets/files/. Last accessed 5 Dec 2015

Kodama on-orbit operations, http://global.jaxa.jp/projects/sat/drts/. Last accessed 5 Dec 2015

National Space Science Data Center -TRDS, http://nssdc.gsfc.nasa.gov/nmc/spacecraftDisplay.do?id=2002-011A. Last accessed 5 Dec 2015

OG2 – the next generation is here, ORBCOMM release, http://www.orbcomm.com/en/networks/satellite. Last accessed 5 Dec 2015

Orbcomm Inc.'s, 2007 10K. Security and exchange commission filings

Orbcomm Second Generation. http://www.spaceflight101.com/orbcomm-g2-satellite-overview.html. Last accessed 5 Dec 2015

Orbcomm System Overview, http://www.m2mconnectivity.com.au/sites/default/files/more-information/System_Overview_Rev_G.pdf. Last accessed 5 Dec 2015

Peter de Selding, Orbcomm to space X: Launch our satellites before October. Space News, 8 May 2015, http://spacenews.com/orbcomm-to-spacex-launch-our-satellites-before-october/. Last accessed 5 Dec 2015

Small Satellites, www.sstl.co.uk. Last Accessed 5 Dec 2015

STRaND Nanosatellite Smartphone, http://www.sstl.co.uk/Missions/STRaND-1–Launched-2013/STRaND-1/STRaND-1–Smartphone-nanosatellite. Last Accessed 5 Dec 2015

Surrey Space Centre, AmSat-UK, http://amsat-uk.org/tag/surrey-space-centre. Last Accessed 5 Dec 2015

UoSat-1 Image, University of Surrey Space Centre, http://www.google.de/imgres?imgurl=http://space.skyrocket.de/img_sat/uosat-2__1.jpg%26imgrefurl=http://space.skyrocket.de/doc_sdat/uosat-1.htm%26h=300%26w=194%26tbnid=cVl-OnNNY7oelM:%26zoom=1%26tbnh=130%26tbnw=84%26usg=__hCD-xwiW2Rnnc4GMv10Z7Hg3ckQ=%26docid=IqjpJoPaA65K-M. Last accessed 5 Dec 2015

J. Watkins, 33 years in space – celebrating guildford’s satellite experts. World Space Week, 10 Oct 2014, http://www.getsurrey.co.uk/news/surrey-news/33-years-space—celebrating-7911654. Last accessed 5 Dec 2015

Broadband High-Throughput Satellites

Erwin Hudson

Contents

Abstract

Rapid growth in demand for broadband Internet services has brought new challenges to the satellite industry. Satellite networks must have an incredible amount of bandwidth to deliver high-speed broadband service to large population of subscribers. Ku-band transponder satellites, which comprise a large fraction of the current worldwide fleet, are typically limited to 1-2 Gbps of total capacity.

E. Hudson (✉)
ViaSat, Inc. Carlsbad, CA, USA
e-mail: erwin.hudson@viasat.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_95

Ku-band satellites do not have the scale and bandwidth economics required to provide a compelling broadband service. To satisfy the demand for Internet bandwidth, large broadband satellites need 100 s of Gbps of capacity. The satellite industry has responded to this challenge with new payload designs and new satellite system architectures, advancing into higher-frequency bands and incorporating aggressive frequency reuse, advanced waveforms, adaptive coding and modulation, and other techniques. Broadband satellites approaching 150 Gbps of capacity are now in orbit. Satellites with up to 350 Gbps of capacity are being manufactured and will be launched in the 2016–2017 timeframe. The 1000 Gbps barrier will be exceeded in 2020 with the launch of recently announced third-generation broadband satellites from ViaSat.

Keywords
Access network • Broadband • Broadcast • Capacity • Error correction • Frequency reuse • Gateway • Internet • Modem • Ka-band • Modulation • Payload • Protocols • Satellite • Terminal

Introduction

Demand for broadband Internet access outside urban areas is both a great opportunity and a great challenge for the commercial satellite industry. Terrestrial broadband technologies such as digital subscriber link and cable modem are cost effective in populated areas. But in rural and remote areas, where customers may be located a mile or even several miles apart, these ground-based solutions become expensive and ultimately unaffordable.

Satellite systems do not suffer the dearth of distance limitation. From a geostationary satellite stationed high above the earth, all customers are about the same distance away – 22,500 miles give or take. The cost of serving a customer via satellite is the same regardless of where the customer is located.

The challenge for broadband satellite operators is to achieve the scale and economics required for a profitable and high-growth Internet-via-satellite business. Rapidly growing demand for Internet data has driven broadband satellite service providers to launch higher and higher-capacity satellites. Satellites with 100 s of Gbps of capacity are required to deliver a compelling service to enough customers to close a service provider's business plan. Such incredibly high-capacity satellites were entirely unimaginable just a few years ago.

ViaSat has become the industry leader in designing, building, and operating the highest capacity broadband satellite networks. ViaSat is both an infrastructure provider, building satellite networks for other network operators, and a broadband service provider using the same ViaSat satellite designs and ground equipment provided to others. ViaSat, as an infrastructure provider, delivered the ground system for the satellite part of Australia's National Broadband Network. ViaSat also provides ground equipment for Xplornet's satellite broadband service in Canada and

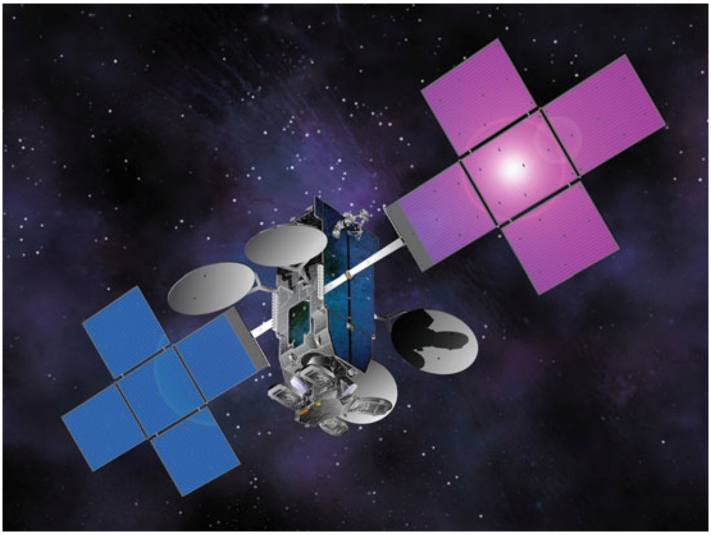


Fig. 1 ViaSat-1 delivers 140 Gbps and is the world's highest capacity satellite (Photo courtesy of Space Systems Loral. All rights reserved)

Eutelsat's Tooway service in Europe. As a service provider, ViaSat operates a highly successful satellite broadband service in the United States, selling residential, enterprise, and mobile services under the "Exede®" brand name (Fig. 1).

ViaSat/Exede delivers 12 Mbps download and 3 Mbps upload services to over 675,000 residential customers while serving thousands more customers on commercial aircraft with its Exede in the Air service. In addition to the two WildBlue legacy satellites, the Exede service is delivered over ViaSat-1, which has up to 140 Gbps of throughput. ViaSat-1 is recognized in the Guinness Book of World Records as the world's highest capacity satellite (ViaSat 2013a).

A new ViaSat-2 satellite with over 2 1/2 times the capacity of ViaSat-1 is completing manufacture and will launch in early 2017. Additional high-capacity satellites, with even more capacity than ViaSat-1, are being developed and will be launched over the next few years.

ViaSat's primary competitor in the United States is Hughes Network Systems (HNS), an EchoStar company. HNS operates EchoStar-17, a large broadband satellite serving the United States and Canada, and the company plans to launch a second broadband satellite EchoStar-19 in late 2016 (de Selding 2015a). Inmarsat has launched a constellation of three broadband satellites called Global Xpress, primarily focused on mobile broadband services for aviation, maritime, and enterprise markets (Hadinger 2015). Eutelsat Communications, a Paris-based satellite company, operates the KA-SAT satellite with approximately 90 Gbps of broadband capacity over Europe (Eutelsat Communications n.d.). Other broadband satellite operators around the world include Thaicom with their IPSTAR satellite that covers much of Asia, Avanti's Hylas-1 and Hylas-2 satellites with coverage over Europe and the

Middle East, and Yahsat with two satellites that cover the Middle East and parts of Africa.

The secret behind these ultrahigh-capacity satellites is a tightly integrated end-to-end system engineering approach. Everything matters. Each element of the network is purpose built and optimized for highest capacity at lowest cost. Data centers, earth stations, satellites, and customer terminals work together to seamlessly extend the Internet cloud up into space and bring it back down to many thousands of customers on the Earth's surface.

This chapter begins with a discussion of the scale and economic challenges created by the rapid growth of the Internet. A brief history of early satellite broadband using leased transponders highlights the importance of very high capacity and good bandwidth economics. The next sections cover high-leverage technology improvements that have allowed ViaSat to design satellites with 100 s of Gbps of capacity. An overview of ViaSat's Exede network architecture shows how these technology improvements are integrated together into a highly optimized satellite broadband system. The closing section looks beyond ViaSat-1 and ViaSat-2 and toward the long-term future of satellite broadband.

The Broadband Satellite Challenge

Rapid growth of the Internet over the past 20 years has created an insatiable demand for data. The telecommunications industry has responded with exponential increases in speed and capacity, and at the same time, with dramatic reductions in the cost of data delivery. Internet access pricing has dropped from between \$6 and \$8 per hour for 2400 bps dial-up service in the early 1990s to less than \$100 per month for a 100 Mbps always-on connection today. It is not unusual for a current home Internet user, particularly a moderate to heavy consumer of streaming video, to download over 100 GB per month. At early dial-up pricing, 100 GB of data downloads would have cost more than \$500,000. Between the early 1990s and today, the cost of Internet access has decreased by more than 5000 times — an amazing reduction in cost.

The new bandwidth economics established by the Internet have entirely transformed telecommunications. Telephone companies, for example, were among the world's most valuable businesses prior to the Internet. Today voice calls and even video calls to any point on earth are either free or almost free. Every sector of the telecommunications industry, particularly satellite, now delivers orders of magnitude more bandwidth at significantly lower prices.

Prior to the emergence of digital networking, bandwidth could be characterized as either point-to-point or broadcast. The public switched telephone network (PSTN) is an example of a large point-to-point system. The PSTN uses arrays of switches to allow any phone to connect to any other phone on the network. Radio and television broadcasters do one-way communications, blanketing their entire service area to allow thousands of customers to receive the same signal simultaneously. Although satellites are used for both point-to-point and broadcast, satellites, especially geostationary satellites, excel at broadcast. More than 30 million Americans receive television broadcast services over geostationary satellites from one of the two leading satellite TV providers, DIRECTV (2015) or DISH Network (2015).

Delivery of broadband Internet service requires a new type of bandwidth, referred to as network access. Rather than connecting customers to each other (point-to-point) or transmitting the same content to many customers in a service area (broadcast), a network access service connects a large number of customers to a network such as the Internet. While connectivity to the Internet is the enabler, it is the Internet itself that gives customers the desired functionality and enables the services they seek. Broadband Internet connectivity allows customers to do voice calls, browse the web, watch streaming video, and download files in a manner that is largely transparent to the Internet service provider.

Starting in the late 1990s, service providers began offering Internet access using leased transponders on Ku-band broadcast satellites, in effect doing network access using infrastructure designed for broadcast. While use of broadcast infrastructure worked technically, it failed economically. A 36 MHz Ku-band transponder, for example, leases for between $100,000 and $250,000 per month, depending on supply and demand in the region and other factors. In order to offer an affordable $100 per month service, Ku-band service providers had to limit their bandwidth cost to around $20 per customer per month, requiring them to put between 5000 and 10,000 customers on each transponder. With thousands of customers sharing a single transponder, the quality of service was poor. As a result, first-generation satellite broadband developed a reputation for being slow and expensive.

The challenge for both satellite manufacturers and satellite operators was to develop new broadband satellite systems optimized for network access that:

- Delivered broadband speeds and monthly downloads comparable to terrestrial technologies
- Were priced competitive with terrestrial technologies
- Offered excellent service everywhere, not just in population centers
- Provided mobile broadband to vehicles, including trucks, trains, ships, and aircraft, better than terrestrial alternatives

Could satellite technology meet these challenges? Was it possible to build and launch broadband satellites that exceed 100 Gbps of capacity? Could satellites ever achieve 1 Tbps of capacity? Would customers ever choose satellite broadband over other network access technologies? Could a satellite broadband operator deploy a compelling service, attract and maintain customers, and operate economically on an ongoing basis?

ViaSat's answer to all of these questions is yes. ViaSat is working on multiple generations of new satellite designs optimized to most efficiently deliver broadband services. Each of ViaSat's future system designs incorporates more advanced technologies and results in higher capacity and better bandwidth economics.

Optimizing Satellite Networks for the Internet

The Internet has become the global network of networks capable of providing live access to every person and thing on the planet. No one could have imagined, as few as 20 years ago, that the Internet would become such an integral part of government and industry. Internet applications such as texting, e-mail, banking, shopping, entertainment, and social networks have changed the lives of billions of people. The primary purpose of a broadband system, broadband satellite in particular, is to give customers access to these Internet applications that have become essential to modern life.

Computer-to-computer communication was the motivation behind the development of what became the Internet. In the 1960s, the Advanced Research Projects Agency (ARPA), an arm of the US Department of Defense, began experimenting with ways to connect large time-share computers at universities and military facilities.

Building on networking concepts developed by J. C. R. Licklider at Massachusetts Institute of Technology (MIT) and on packet switching theory developed by Leonard Kleinrock, also at MIT, ARPA designed a computer-to-computer network called ARPANET. In 1968, DARPA released a request for proposals (RFP) for an Interface Message Processor (IMP) to facilitate the exchange of data between computers. Bolt Beranek and Newman (BBN) won the contract and, within a year, installed the first IMP at the University of California Los Angeles (UCLA). A second IMP was installed at Stanford University Research Institute, and on October 29, 1969, the first computer-to-computer connection was established. BBN installed two more IMPs at the University of California, Santa Barbara and at the University of Utah, and in early 1970, the fledgling Internet consisted of a grand total of four computers (Internet Society n.d.).

Vint Cerf at Stanford University working with Bob Kahn at ARPA developed a new suite of protocols to improve the performance and scalability of the ARPANET. In 1973 Cerf wrote a specification known as the Transmission Control Protocol/Internet Protocol (TCP/IP). The Internet Protocol defines the addressing scheme that allows packets to be routed to their destinations. TCP ensures the assembly of IP packets into error-free messages, allowing for out of order packets and retransmission of corrupted and missing packets. The incredible scalability of TCP/IP allowed networks to expand to an almost infinite number of connections. The ARPANET implemented TCP/IP in 1983 and, at that point, what we know today as the Internet was born (Internet Society n.d.).

While the early Internet was a powerful resource for government and academia, it was Sir Tim Berners-Lee's development of the World Wide Web in 1990 that brought the Internet to millions and eventually billions of ordinary people. Lee, a researcher at Conseil Européen pour la Recherche Nucléaire (CERN) in Geneva, proposed an innovative concept of clicking on hyperlinks to allow one to access cross-references between scientific documents. Hyperlinking turned out to be incredibly powerful, making nearly every piece of information on the Internet instantly available to everyone. Marc Andreessen and Eric Bina at University of Illinois

Urbana-Champaign developed the first popular web browser called Mosaic, giving the Internet a simple point-and-click interface. Mosaic and its commercial successor Netscape brought the Internet to the masses. Netscape enabled true "e-commerce," and in July 1995 a little start-up called Amazon.com went live on the Internet (Brandt 2011). Early e-commerce companies grew at astounding rates, and today, many of them – Amazon, eBay, Yahoo, and Google to name a few – are corporate giants and household names.

As of year-end 2014, the Internet had grown to connect almost three billion people (Internet Live Stats n.d.). The popular social media site Facebook reports more than 1.55 billion monthly active users (Statistica 2015). To put the popularity of Facebook into perspective, up to one out of every seven human beings on planet earth checks a Facebook account each day (Zuckerberg 2015) (Fig. 2).

Internet data consumption is growing at annual rate of 20–30 %, doubling every 4–5 years. The world's rapid-growing appetite for data is a big challenge for Internet service providers (ISPs). If available capacity falls behind continuously increasing demand, network congestion and slow speeds occur, resulting in dissatisfied customers. In order to stay competitive, ISPs must continue to invest in higher-capacity technologies.

Figure 3 shows a high-level diagram of a broadband satellite network. Connectivity to the Internet may be through a large gateway Earth station with antennas as large as 13 m. Customer terminals are much smaller, typically less than 1 m in diameter. Following the arrows in Fig. 3, the Internet-to-customer or return link satellite payload collects signals from many customer terminals and retransmits those signals down to a gateway. The customer-to-Internet or forward link satellite payload, on the other hand, receives the signal from a gateway and retransmits those signals to many customer terminals.

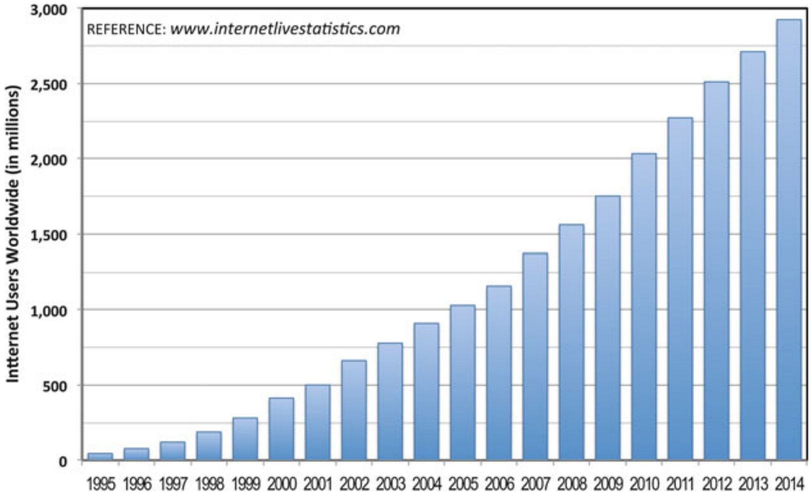


Fig. 2 Nearly three billion people now use the Internet – approaching half the world's population

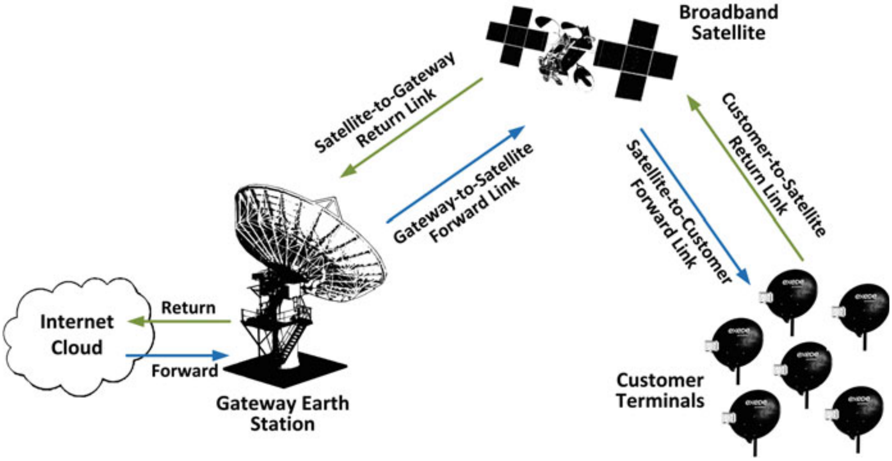


Fig. 3 Broadband satellite system network diagram

A well-designed broadband satellite payload takes into account the unique requirements of each of these four communications links. The links between the satellite and the gateway benefit from the high transmit and receive gain of the large gateway antenna. For the satellite payload engineer, the most challenging requirements are the "small antenna to small antenna" links between the satellite and the customer terminals. A high-capacity broadband satellite may need 100 or more high-power amplifiers to support the required forward link satellite-to-customer capacity. As a result, it is almost always the satellite-to-customer or forward link that stresses the payload design.

Broadband Satellite Service in the United States

By the mid-1990s demand for Internet access had grown to the point where homes and small businesses in even the most remote areas of the United States began looking for alternatives to dial-up service. Direct-to-home satellite television, particularly the high-power services launched by DIRECTV in 1994 (DIRECTV n.d.) and DISH Network (DISH Network n.d.) in 1996, proved how effectively satellites could connect rural homes regardless of location. Based on the success of satellite television, satellite technology seemed like the ideal way to deliver broadband Internet to these same rural markets.

High-speed satellite broadband, however, turned out to be more difficult than early satellite service providers anticipated. In 1996, DIRECTV and Hughes Network Systems, both Hughes Aircraft Companies at the time, began offering a hybrid satellite-dial-up service. Branded DirecPC, the Hughes service used leased Ku-band satellite transponders for downloads and dial-up modems for uploads, providing peak download speeds of 400 kbps (Conti 2000). Uploads were limited to dial-up modem speeds in the range 9600–28,000 kbps. In November 2000, StarBand, a partnership led by Gilat Satellite Networks Ltd., offered the first consumer two-way satellite broadband service with download speeds of 500 kbps and upload speeds of 150 kbps (Gilat Satellite Networks 2000). Hughes upgraded DirecPC to a two-way satellite service in 2001.

These early attempts to provide broadband over satellite were less successful than expected. The high cost of leased Ku-band transponders was the limiting factor. Squeezing thousands of customers onto each Ku-band transponder resulted in heavy congestion and slow speeds. Satellite customers were not happy paying a premium for a service that often worked poorly. Customer satisfaction was low, churn was high, and subscriber growth was limited.

By 2000, it was apparent that satellite capacity had to increase by at least an order of magnitude to cost-effectively deliver broadband services. A large Ku-band satellite with 32 standard 36 MHz transponders was state of the art at the time. With subscriber antennas in the 1.0–1.2 m range, an entire 32-transponder satellite could provide just 1.0–1.5 Gbps of total capacity – far below the minimum 100 Gbps needed to effectively deliver high-speed broadband to a large population of subscribers.

Colorado-based WildBlue Communications, in 2005, was first to offer Internet access service using purpose-built satellites. WildBlue entered service with a Ka-band-hosted payload on Anik F2, a large multi-payload satellite launched by Telesat Canada. WildBlue offered three tiers of service ranging from 0.5 Mbps to 1.5 Mbps and priced from $50 to $80 per month. Telesat's partner company Barrett/Xplornet offered a similar service in Canada. The service was fast and affordable relative to alternatives and demand was high. Less than 2 years later in December 2006, WildBlue launched the first dedicated Ka-band commercial satellite, WildBlue-1, which more than doubled the capacity of the network to approximately 15 Gbps. In the first few months after the launch of WildBlue-1, WildBlue was installing more than 5000 new customers a week (de Selding 2004). The combined bandwidth of the two satellites, however, was not sufficient to meet demand, and customer growth slowed as the satellites began to reach capacity. With both satellites entirely filled in areas of strong demand, WildBlue reached a peak of 420,000 customers in 2010.

The WildBlue network was based on ViaSat's first-generation SurfBeam ground technology. SurfBeam used the highly successful cable modem standard, Data over Cable System Interface Specification (DOCSIS). Designing satellite communications equipment to a commercial standard was highly innovative at the time, allowing a faster development cycle, lower cost, and better performance. With broadband speeds up to 3 Mbps download and 1 Mbps upload, ViaSat's SurfBeam was the first truly competitive satellite broadband system.

In January 2008, ViaSat announced a contract with Space Systems Loral to build ViaSat-1, the first broadband satellite to exceed 100 Gbps capacity. ViaSat acquired WildBlue in December 2009 and WildBlue became ViaSat Communications. The branding of the broadband service was changed from WildBlue to Exede to highlight the faster speeds and better performance available with ViaSat's second-generation

SurfBeam technology. The ViaSat-1 satellite was launched in October 2011 and entered service in early 2012, increasing the capacity of the WildBlue/Exede network by more than ten times, from 15 Gbps to over 150 Gbps. With this huge increase in capacity, ViaSat began offering Exede services with 12 Mbps download speeds and 3 Mbps upload speeds.

Exede was the first satellite service to compete head to head with terrestrial broadband. Based on customer surveys, up to 40 % of new Exede customers previously subscribed to a terrestrial service, often fixed wireless or DSL. The performance of the ViaSat-1 network was a big accomplishment by ViaSat, allowing the Exede network to offer Internet speeds to rural and suburban customers that were faster than the best speeds available to many urban customers.

ViaSat's Exede service has now grown to more than 675,000 subscribers. Residential offerings have been enhanced with telephony services using Voice over Internet Protocol (VOIP), unmetered downloads during off-peak hours, virtually unlimited service plans in selected areas, and other features. Exede service plans are available for small businesses and disaster recovery efforts, newsgathering, sporting events, and other applications where satellite has unique advantages over alternatives (Fig. 4).

In December 2013, ViaSat and JetBlue Airlines launched the first high-speed Ka-band broadband service available to commercial airline passengers (ViaSat 2013c). For the first time, passengers in flight could surf the web, send and receive e-mails, download files, and enjoy streaming video at true broadband speeds – capabilities that are all but impossible using other technologies. ViaSat began installing the service on United Airlines aircraft in February 2014 (Honig 2014) and announced a contract with Virgin America in July 2015 (Virgin America 2015). In September 2015 ViaSat and Boeing announced the availability of ViaSat's Exede in the Air as a line-fit option on selected Boeing wide-body aircraft, allowing airlines to order new aircraft with ViaSat's high-speed broadband service factory installed and ready for immediate use (Henry 2015).

Fig. 4 ViaSat's Exede in the air service is now available on US domestic flights on JetBlue, United/Continental and Virgin America Airlines (Photo provided by ViaSat. All rights reserved.)

Building on the success of ViaSat-1 and the Exede network, ViaSat has invested in a ViaSat-2 satellite for launch in early 2017 (ViaSat 2013b). ViaSat-2 will have up to 350 Gbps of capacity. The new satellite will cover the United States and Canada, Mexico, the north Atlantic, the Caribbean, and more. ViaSat-2 will support Exede and Exede in the Air services across North America and Latin America and will allow continuous coverage to aircraft flying from the west coast of the United States, across the continent, over the Atlantic, and into Europe.

Broadband Satellite Scale and Capacity

Large broadband satellite networks serving hundreds of thousands of customers need ultrahigh-throughput satellites. Two factors – the number of customers served and the average usage per customer – determine the required capacity. While somewhat counterintuitive, the speed at which the service is delivered to the customer has little impact on the satellite capacity required. It is peak demand for data and not speed of delivery that drives the design of broadband satellite networks.

Determining the capacity required to deliver high-quality broadband connectivity to thousands of customers is a complex statistical problem. Not all customers access the service at the same time and rarely do customers demand the maximum available speed. Even during periods of peak demand, a large fraction of customers will demand almost no data at all, and many customers who actively use the service may consume only a fraction of the throughput available. An active customer on a 25 Mbps service, for example, may be watching a high-definition movie that downloads at 6–8 Mbps, consuming less than a third of the maximum speed available.

So how does one determine how much satellite capacity is required? A simple way is to use average monthly consumption per customer to estimate peak busy hour load. While this method is strictly a rough approximation, it is often used as the initial capacity estimate for a new broadband system.

Consider, for example, a satellite ISP who wants to offer a broadband service with an average monthly consumption of 25 GB per customer in the forward direction $\left(MC_{fwd} = 25\ \text{GB}\right)$ and 5 GB per customer in the return direction $(MC_{rtn} = 5\ \text{GB})$. To avoid congestion at peak busy hour, the satellite must be sized to accommodate peak load.

On broadband networks dominated by residential customers, peak busy hour occurs somewhere between 7 PM and 10 PM and approximately 7 % of monthly broadband traffic is consumed during busy hour. The 7 % factor is surprisingly consistent across beams and even across networks. Figure 5 shows a typical beam loading profile over a 24-h period. The satellite must have enough capacity to handle 7 % of monthly uploads and downloads during the daily peak hour – in other words, a total of 7 % of monthly throughput over 30 one-hour periods.

Using this model, the busy hour demand per customer (BHDC) may be calculated as a function of monthly consumption:

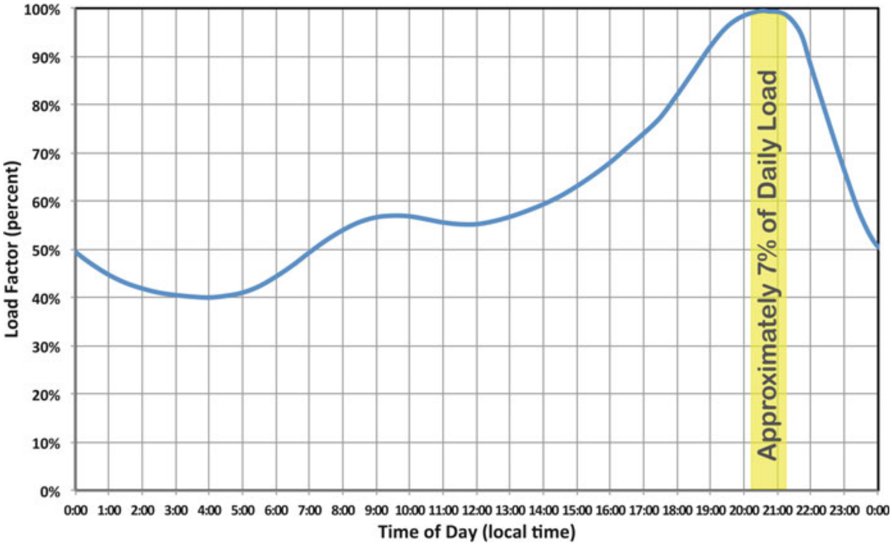


Fig. 5 Customer demand profile over a 24-h period (Residential behavior)

$$BHDC = (MC \times 7\%) / 30 \text{ days} \tag{1}$$

The required satellite throughput requirement per customer at busy hour is determined by converting *BHDC* to kbps. This parameter is referred to as customer provisioning rate (*P*) and is expressed in kbps/customer:

$$P = (BHDC \times 8 \text{ bits/byte}) / 60 \text{ min} / 60 \text{ s} \tag{2}$$

For a customer downloading an average of 25 GB per month, demand in the forward direction is

$$BHDC_{fwd} = (25{,}000 \times 7\%) / 30 \text{ day } = 58.3 \text{ MB} \tag{3}$$

and the forward link provisioning rate is

$$P_{fwd} = (58{,}300 \times 8000/1000) / 60 \text{ min} / 60 \text{ s} = 130 \text{ kbps} \tag{4}$$

Doing the same math for the return direction, the average demand per customer during busy hour is

$$BHDC_{rtn} = (5000 \times 7\%) / 30 \text{ days} = 11.7 \text{ MB} \tag{5}$$

and the return link provisioning rate is

$$P_{rtn} = (11{,}700 \times 8 \text{ bits/byte}) / 60 \text{ min} / 60 \text{ s} = 26 \text{ kbps} \tag{6}$$

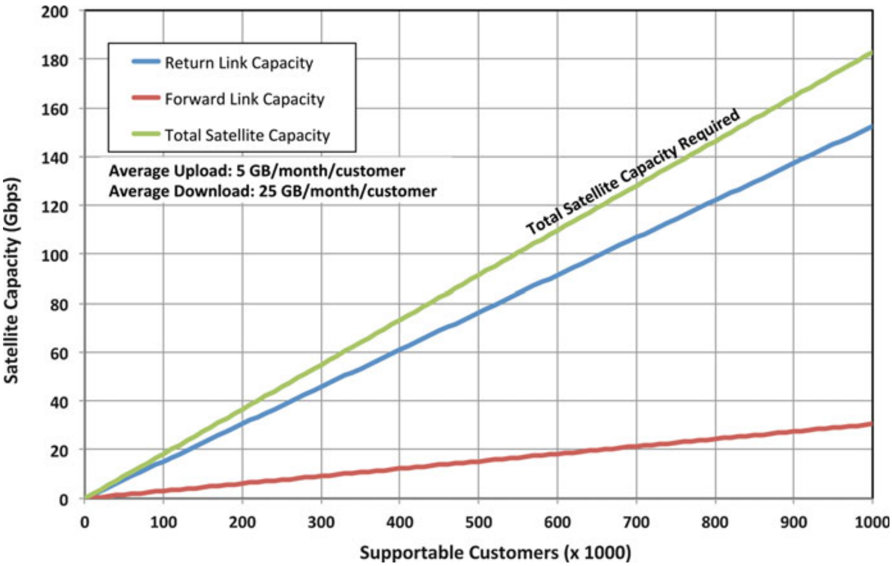


Fig. 6 Satellite capacity required to deliver an average 25 GB by 5 GB per month per customer

Much like a freeway needs spacing between cars to keep traffic flowing smoothly, broadband networks need additional capacity beyond total provisioned bandwidth to avoid congestion and to maintain good speeds. A commonly used rule of thumb is to limit channel loading to no more than 85 % of maximum throughput. For 25 GB by 5 GB average usage, and with the peak load limited to 85 %, the satellite must provide 153 kbps of forward link capacity and 31 kbps of return link capacity for each customer (Fig. 6).

Total satellite capacity determines the number of customers a broadband satellite network can support. Figure 6, computed for average downloads of 25 GB/month and average uploads of 5 GB/month per customer, shows the satellite capacity required to support up to 1,000,000 customers.

Figure 6 highlights the need for ultrahigh-capacity satellites. The figure shows that a 200 Gbps class satellite is required to provide an average 25 GB by 5 GB per month service to a million customers. This data also reinforces the prior discussion that broadband networks require purpose-built satellites optimized for broadband. Broadcast satellites and generic Ku-band transponder satellites do not have even a fraction of the bandwidth required to provide a competitive broadband service.

Designing High-Capacity Broadband Satellite Payloads

Designing satellites with 100 s of Gbps of capacity is a whole new problem requiring an entirely new solution. ViaSat payload engineers and ground system engineers, starting with a clean sheet of paper, have worked together to design the highest capacity broadband satellites. Achieving the bandwidth economics required to support an economically compelling Internet-via-satellite business has driven every aspect of the space segment and ground system design. In less than 10 years, ViaSat has reduced the cost of satellite bandwidth by over 100 times.

The largest increases in satellite capacity have come in a number of areas, most significantly:

- Higher frequencies and narrower beamwidth beams
- Aggressive frequency reuse
- Highly efficient physical layer
- Self-optimizing waveforms

These capacity improvements are discussed in the following sections. The approach is to evaluate each incremental capacity increase and then show how these improvements are combined together to allow for satellites with 100–350 Gbps throughput.

Higher-Frequencies and Narrower Beamwidth Beams

Large broadband satellites provide communications links between the Internet and many thousands of customers spread across the coverage area. Unlike television broadcast, there is no reason for two or more broadband customers to receive the same data transmission at the same time. To the contrary, given the importance of privacy and security on the Internet, it is important that data intended for a specific customer are available only to that customer. Broadband satellites rely on multiple beams, multiple carriers, TCP/IP connections, and data encryption to maintain the customer-by-customer connectivity required for Internet access.

In the satellite downlink, for example, any microwave energy transmitted by the satellite that is not collected by the one intended customer's ground terminal antenna is effectively wasted. The same argument can be made for the uplink direction as well. This thought process leads to the conclusion that the most efficient broadband satellite should have a separate dedicated beam for each customer, creating a tiny coverage spot around each customer's ground terminal.

Building a satellite with hundreds of thousands of microscopic beams, one for every customer, is not realistic, certainly not with today's technology. The principle, however, is directionally correct – high-capacity broadband satellites need a large number of small beams to minimize wasted microwave energy. In addition to improved microwave energy efficiency, multibeam satellites also have the option of reusing available spectrum over and over again to maximize total capacity.

The area covered by a satellite antenna beam on the Earth is established by the effective diameter of the antenna aperture on the spacecraft and by the frequency band of operations. Higher frequencies result in smaller beams, and smaller beams lead to higher-capacity satellites.

The primary frequency bands available for commercial geostationary communications satellites, in order of increasing frequency, are designated C-band, Ku-band, and Ka-band. Table 1 shows the range of earth-to-space and space-to-earth frequency allocations for each of these bands.

Figure 7 shows the relationship between antenna aperture diameter and downlink beamwidth for the three frequency bands. Based on Fig. 7, a C-band satellite with a 2.5-m antenna has a beamwidth greater than 2 °, while a Ka-band satellite with the same diameter antenna has a beamwidth of less than 0.4 °. Since a larger number of smaller beams results in more capacity, a satellite operating in the Ka-band will deliver significantly greater capacity than a Ku-band or C-band satellite.

Not only does Ka-band enable a larger number of beams, there is also more commercial satellite spectrum available at Ka-band. While satellite spectrum is regulated differently around the world, there is approximately 500 MHz available at C-band and Ku-band compared to as much as 2500 MHz at Ka-band. The

Table 1 Commercial Satellite Frequency Bands

Commercial satellite band	Earth-to-space (GHz)	Space-to-Earth (GHz)	Bandwidth available (MHz)
C-band	5.9–6.4	3.7–4.2	500
Ku-band	14.0–14.5	11.7–12.2	500
Ka-band	27.5–30.0	17.7–20.2	up to 2500

Note: Spectrum varies by region; extensions to standard bands may be available in certain areas. Allocated spectrum may be combination of primary and secondary

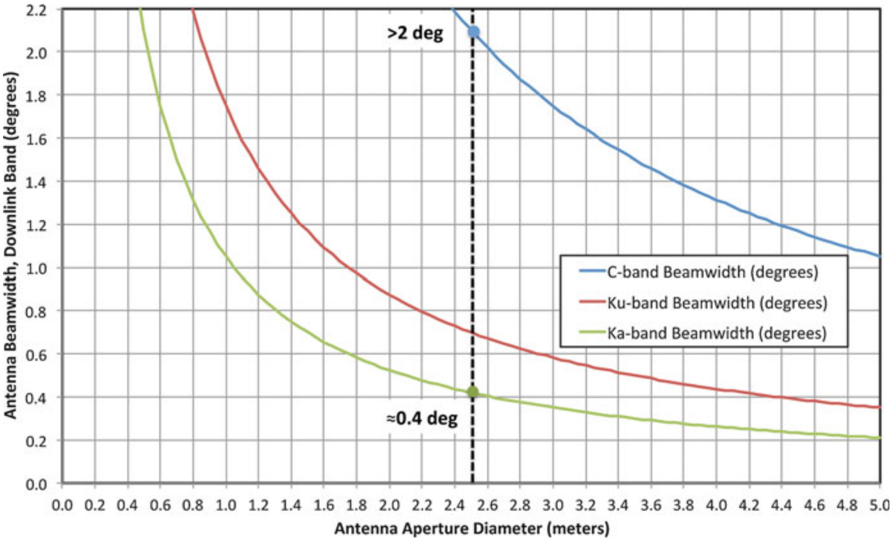


Fig. 7 Relationship between antenna aperture diameter and beamwidth

combination of more beams and up to five times more spectrum makes Ka-band the better choice for broadband satellites.

Additional spectrum available at Ka-band increases satellite capacity by three to five times.

Aggressive Frequency Reuse

Frequency reuse is a powerful way to increase capacity. Multibeam satellites reuse available frequencies over and over again across their coverage area. Maximizing network capacity is a careful balance between total bandwidth (which increases capacity) and interference between beams (which decreases capacity). The trade-off between bandwidth and interference requires broadband satellite engineers to pay careful attention to antenna beam roll-off and sidelobes to limit beam-to-beam interference to acceptable levels.

It is helpful to think of frequency reuse in terms of colors, where a "color" refers to a specific block of frequencies. Colors may be contiguous or non-contiguous and on one polarization or both.

Figure 8 shows an example of four-color frequency reuse where the available spectrum is continuous and both polarizations are available. Each beam is assigned one-fourth of the available spectrum. Interference occurs when microwave energy spills over between beams of the same color. Beams using the blue spectrum, for example, are placed as far apart as possible to minimize self-interference. The same reasoning applies to the other colors as well.

The right side of Fig. 8 shows a section of a uniformly spaced multibeam antenna beam pattern using four-color frequency reuse. The strategy is to assign the four colors in a repeating pattern that maximizes the distance between beams of the same color. A 100-beam satellite with a four-color coverage pattern has a frequency reuse

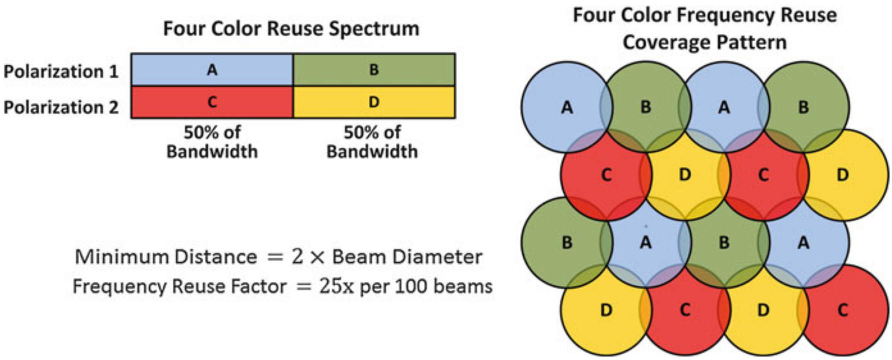


Fig. 8 Four-Color frequency reuse coverage

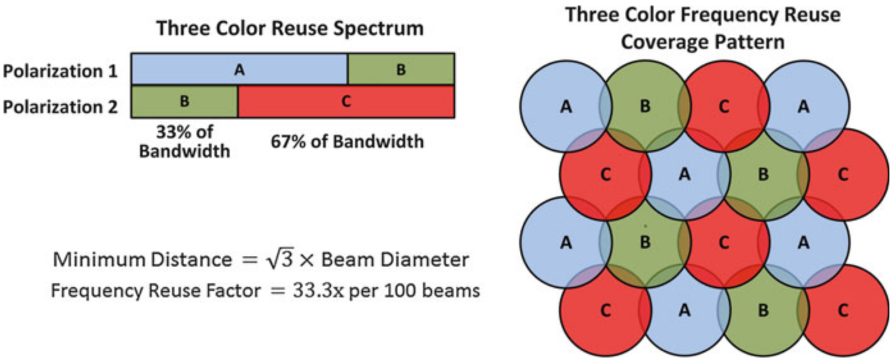


Fig. 9 Three-Color frequency reuse coverage

factor of 25 times. Same color beams, as measured from beam center to beam center, are exactly two beam spacing distances apart. It is same color beam spacing that determines the beam-to-beam interference that the uplink and downlink waveforms will be required to tolerate.

Figure 9 extends the concept to three-color frequency reuse. The green color in this example is non-contiguous, consisting of two frequency bands spread across both polarizations. While use of non-contiguous colors increases the complexity of the satellite payload implementation, the frequency reuse approach remains the same.

A three-color frequency reuse beam pattern is shown on the right side of Fig. 9. Compared to four colors, three-color frequency reuse increases total bandwidth per 100 beams by 32 %, from 25 times to 33 times, but reduces the minimum distance between beams of the same color by 13.4 %, from 2 to $\sqrt{3}$ beam diameters. Three-color frequency reuse results in more total bandwidth but increases beam-to-beam interference. Depending on the satellite antenna design, particularly close in sidelobes, three-color frequency reuse may or may not result in more capacity than four-color frequency reuse.

Frequency reuse results in a huge increase in capacity. Consider, for example, a 100-beam Ka-band broadband satellite with 1500 MHz of bandwidth available on two polarizations. Four-color frequency reuse allows 750 MHz to be assigned to each color. The total bandwidth of the satellite, with 750 MHz in each of 100 beams, is a staggering 150,000 MHz counting both forward and return directions. This example shows how Ka-band spectrum, multibeam antennas, and frequency reuse enable extremely high-capacity broadband satellites.

Frequency reuse may increase broadband satellite capacity by 25-times or more.

Highly Efficient Physical Layer

While frequency reuse allows a satellite to have tremendous bandwidth, it is the physical layer, the uplink and downlink waveforms, that converts bandwidth into customer capacity. Broadband satellite systems have adopted more and more complex waveforms with higher spectral efficiencies in order to maximize throughput.

Radio communications may be modeled as an exchange of abstract symbols between transmitter and receiver. The maximum symbol rate (R_s) is limited by the channel bandwidth (B). The relationship between maximum symbol rate and bandwidth is given by

$$R_s = {}^{B}/_{\alpha}\ \text{Hz} \tag{7}$$

where the parameter α is called the roll-off factor. The roll-off factor is determined by the amount of symbol shaping or smoothing done by the modulator prior to transmission. For broadband satellite waveforms, α is typically in the 1.10–1.25 range. Using Equation 7, a 100 MHz bandwidth channel using a roll-off factor of $\alpha = 1.2$, for example, supports a maximum symbol rate of 83.3 Msps.

A larger dictionary of allowable symbols, such that each unique symbol represents a larger number of bits, may increase capacity. The principle is the same as communicating between ships by waving flags – the more unique flags the sender has to select from, the more complex message each wave of a flag may represent. The number of bits per symbol is called the modulation order (N) and corresponds to the number of unique flags in the ship-to-ship signaling analogy. Higher modulation order waveforms deliver more bits per symbol and have the potential to increase throughput. The effective bit rate (R_b) for a waveform with modulator order N is just the product of modulation order and symbol rate:

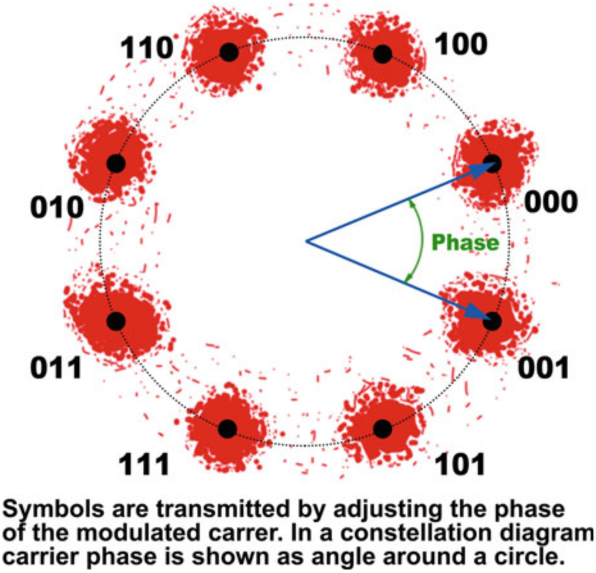
$$R_b = N \times R_s\ \text{bits/s} \tag{8}$$

Binary phase shift keying (BPSK, 2 symbols, N = 1 bits/symbol) and quadrature phase shift keying (QPSK, 4 symbols, N = 2 bits/symbol) have historically been the preferred modulation modes for satellite communications. BPSK and QPSK are simple waveforms, phase noise tolerant, and work nicely with nonlinear transmitters such as the traveling-wave tube amplifiers (TWTAs) used on spacecraft.

Advances in satellite modems, however, allow more complex higher-order modulation. Broadband satellite networks now incorporate 8-ary phase shift keying (8PSK, N = 3 bits/symbol), 16-ary amplitude/phase shift keying (16APSK, N = 4 bits/symbol), and 32-ary amplitude/phase shift keying (32APSK, N = 5 bits/symbol).

Using these higher-order waveforms, the satellite bandwidth supports higher bit rate, but at a cost – with the larger number of symbols to choose from, the risk that a mistake will be made determining which symbol was sent increases. Back to the analogy where flags are being used to signal between ships, as the number of

Fig. 10 8PSK constellation diagram



slightly different flags increases, it becomes increasingly difficult to determine exactly which flag was waved. Microwave communications have exactly the same problem.

Figure 10 shows a constellation diagram of an 8PSK (N = 3 bits/symbol), modulated waveform where each of the eight unique symbols, represented by eight equally spaced phase angles around a circle, is mapped to 3 bits. The 8PSK symbols are identified by black dots in the figure, and the red dots represent the blurring effect of random noise present at the receiver. While the satellite transmits the exact phase angles represented by the black dots, the ground terminal receives the red dots and has to make a best guess as to which black dot was actually sent. As random noise increases, the red dots get farther and farther away from the correct symbol, and it becomes harder and harder to determine exactly which symbol was transmitted.

One way to reduce the number of errors is to increase the symbol energy to noise density ratio (*Es*/*No*), often by increasing the power of the transmitter. Requiring larger space-based transmitters, however, where onboard electric power is costly and constrained, may not be the best option. The cost of increasing the transmitter power on hundreds of thousands of customer terminals on the ground may also be prohibitive. Use of error correction coding to minimize the probability of bit errors on satellite links is often a more cost-effective alternative to higher transmitter power. Broadband satellites rely on a combination of higher-order modulation and error correction coding to maximize capacity.

Correcting bit errors may appear to be almost magical, but the process is highly analytic, relying on well-established mathematics and the statistics of random noise. A simple but terribly inefficient way to do error correction would be

Fig. 11 Block code with $(n - k)$ Check bits appended to each (k) data bits

transmit every bit three times and have the receiver take a two out of three vote. While a two out of three vote would certainly allow the receiver to correct errors (as long as no more than one error occurred per three bits), it would reduce throughput by a factor of three. Fortunately there are more efficient error correction schemes that have much less impact on throughput. Efficient error correction is done by introducing mathematical structure, in effect rules that each data sequence must follow, into the transmitted data. The receiver checks the received data stream and if the data does not follow the rules, the bits that appear to be in error are corrected.

Figure 11 shows what is called a block code, where each block of n bits consists of a string of k data bits plus $n - k$ check bits. Check bits are appended to the data to give each code block-specific mathematical relationships that can be used by the receiver to correct errors. The receiver decodes each code block, performs the required mathematical calculations, and corrects any errors identified in the data bits. The check bits are then discarded and the stream of corrected data bits is passed on to the end user.

The most desirable error correction codes are those that allow the best error correction with the least number of check bits. Code rate (R) is defined as the ratio of data bits to block length, given by

$$R = k/n \tag{9}$$

Low-rate codes (smaller R) correct the most errors but have the greatest impact on capacity. High-rate codes (larger R), with more data bits and less check bits, have the least impact on capacity but cannot correct as many errors.

The combination of a particular modulation type and a specific error correction code is called the modulation-code point. 8PSK modulation with an $R = 3/4$ error correction code, for example, is referred to as the 8PSK 3/4 modulation-code point. The spectral efficiency (ξ) of a modulation-code point is simply the product of the modulation order and the code rate:

$$\xi = N \times R \ \text{bits/Hz} \tag{10}$$

Recalling that 8PSK has modulation order $N = 3$, the modulation-code point 8PSK 3/4 has a spectral efficiency of

$$\xi = 3 \times \frac{3}{4} = 2.25 \ \text{bits/Hz} \tag{11}$$

Selecting the modulation-code point with the highest spectral efficiency that the channel can support results in maximum capacity.

The spectral efficiency of a modulation-code point may be evaluated as a fraction of the Shannon limit, which establishes an upper bound on the capacity of a communication's channel (Shannon and Weaver 1964). Claude Shannon's famous noisy channel coding theorem, published in 1948, proves that for a channel of bandwidth *B* operating at a particular signal-to-noise ratio (SNR), maximum capacity (C_{max}) is

$$C_{max} = B \times \log_2(1 + \text{SNR}) \text{ bits/s} \tag{12}$$

Given that *Es*/*No* and SNR are related by the roll-off factor α as

$$\text{SNR} = (Es/No)/\alpha \tag{13}$$

Shannon's limit may be written as

$$C_{max} = B \times \log_2(1 + (Es/No)/\alpha) \text{ bits/s} \tag{14}$$

Shannon also proved that performing more and more effective error correction coding allows a channel to operate closer and closer to maximum capacity.

In spite of decades of research, the performance of error correction codes remained well below the Shannon limit for over 40 years. The first practical codes to approach Shannon capacity were turbo codes, patented in 1991. Turbo codes were a huge breakthrough but required licensing from the patent holders, which limited adoption. Then in 1996, Sir David MacKay at the University of Cambridge (MacKay and Neal 1996) rediscovered a PhD dissertation written by MIT student Robert Gallager way back in 1963 (Gallager 1963). Gallager had invented an incredibly powerful family of error correction codes called low-density parity-check (LDPC) codes but decided they were too computationally intensive to implement with 1960s technology, and his discovery was filed away in the MIT archives. With recent improvements, Gallager's LDPC codes are now better than turbo codes and, having been in the public domain for decades, are unencumbered by licensing requirements. LDPC codes have been widely adopted for 802.11 WiFi, 10G Ethernet, and other applications. The satellite digital video broadcast-second-generation (DVB-S2) standard incorporates LDPC codes (ETSI EN 302 307-1 2014).

The DVB-S2 standard defines a family of modulation-code points that covers an *Es*/*No* range from less than −2 dB to more than +16 dB.

Table 2 includes a subset of the modulation-code points in the DVB-S2 standard and is a good reference for evaluating required *Es*/*No* and spectral efficiency.

Higher-order modulation with error correction coding increases capacity by 2.0 to 2.5 times.

Table 2 Performance of DVB-S2 standard modulation-code points

Modulation mode	Ideal Es/No (dB)	Spectral efficiency (ξ)	Modulation mode	Ideal Es/No (dB)	Spectral efficiency (ξ)
QPSK 1/4	−2.35	0.490243	8PSK 2/3	6.62	1.980636
QPSK 1/3	−1.24	0.656448	8PSK 3/4	7.91	2.228124
QPSK 2/5	−0.30	0.789412	16APSK 2/3	8.97	2.637201
QPSK 1/2	1.00	0.988858	16APSK 3/4	10.21	2.966728
QPSK 3/5	2.23	1.188304	16APSK 4/5	11.03	3.165623
QPSK 2/3	3.10	1.322253	16APSK 5/6	11.61	3.300184
QPSK 3/4	4.03	1.487473	32APSK 3/4	12.73	3.703295
QPSK 4/5	4.68	1.587196	32APSK 4/5	13.64	3.951571
QPSK 5/6	5.18	1.654663	32APSK 5/6	14.28	4.119540
8PSK 3/5	5.50	1.779991	32APSK 8/9	15.69	4.397854
QPSK 9/10	6.42	1.788612	32APSK 9/10	16.05	4.453027

Refer to ETSI Standard EN 302 307 V1.4.1 (2014–11). Es/No based on Quasi Error Free Packet Error Rate of 10^{-7} (AWGN channel). Spectral efficiencies assume FECFRAME length = 64,000

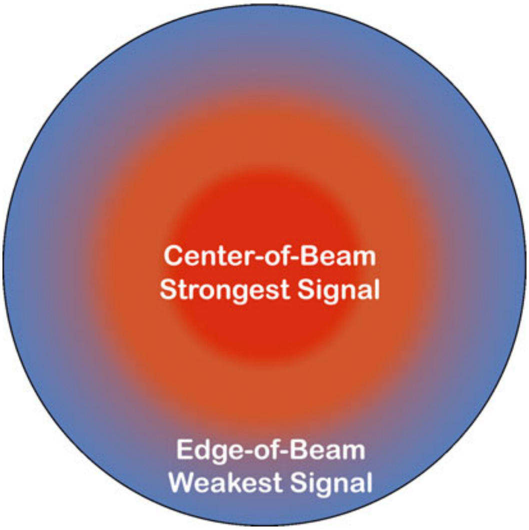


Fig. 12 Satellite beam "heat map"

Self-Optimizing Waveforms

Broadcast satellite networks must be designed such that customers at edge of beam, where the signal is weakest, can receive a high-quality signal. As shown in Fig. 12, broadcast customers at center of beam, where the signal is stronger and could deliver higher capacity, are limited to the lower data rates required to serve edge-of-beam customers. The potential capacity of the stronger signal at the center of beam is wasted in a broadcast satellite network.

Broadband satellites, however, enjoy a greater degree of flexibility than broadcast satellites. In a broadband satellite network, transmissions are intended for only one customer at a time. There is no requirement that every customer modem in a beam be able to demodulate the satellite signal at every moment in time. As a result, the modulation-code point may be optimized on a customer-by-customer basis. Data transmitted to a customer at center of beam may be done at a more bandwidth-efficient modulation-code point, while the same data transmitted to a customer at edge of beam may require heavier error correction coding and a less efficient modulation-code point in order to be successfully delivered with high confidence.

The technique of adjusting the modulation-code point on a customer-by-customer basis is called adaptive coding and modulation (ACM). Figure 13 shows how ACM increases the capacity of a broadband satellite system. The blue line in the figure represents the *Es*/*No* across the beam; in this case *Es*/*No* is approximately 12 dB at center of beam and decreases to 7.5 dB at edge of beam.

An engineer doing a traditional link budget would refer to the data in Table 2 and select the most efficient waveform capable of supporting the edge of beam *Es*/*No* – in this case choosing the 8PSK 2/3 modulation-code point. Every customer in the beam would receive data encoded at the same 8PSK 2/3 modulation-code point, even customers at the center of beam where the *Es*/*No* is more than 4 dB greater. An ACM link budget, on the other hand, allows the modulation-code point to be adjusted automatically based on the customer's *Es*/*No* as shown on the right side of Fig. 13. Customers at peak of beam would receive 8PSK 5/6 and customers between center and edge of beam would get other more robust modulation-code points. ACM allows engineers to harvest the additional signal power across the beam, delivering higher average throughput to the customer population.

Continuing with the example in Fig. 13, the effective improvement in capacity is calculated in Table 3. The spectral efficiency of the static non-ACM approach using 8PSK 2/3 is approximately $\xi = 2.0\ \text{bits/Hz}$, while the average spectral efficiency using ACM improves to $\xi = 2.7\ \text{bits/Hz}$ resulting in a 35 % gain in overall capacity.

Relative to static link budgets, incorporating ACM is better than buying three satellites and getting a fourth satellite for free. Satellites cost hundreds of millions of dollars! Almost all modern broadband satellite networks incorporate ACM in both uplink and downlink designs.

Adaptive coding and modulation increases satellite capacity by 35 % or more.

Combining Multiple Capacity Improvements to Design Very High-Capacity Broadband Satellites

Increasing the capacity of satellites by more than 100 times in a little over a decade is an amazing feat. Each of the features discussed in the previous sections contributes to the total capacity of a large broadband satellite. Table 4 shows how broadband satellites have achieved such huge increases in capacity.

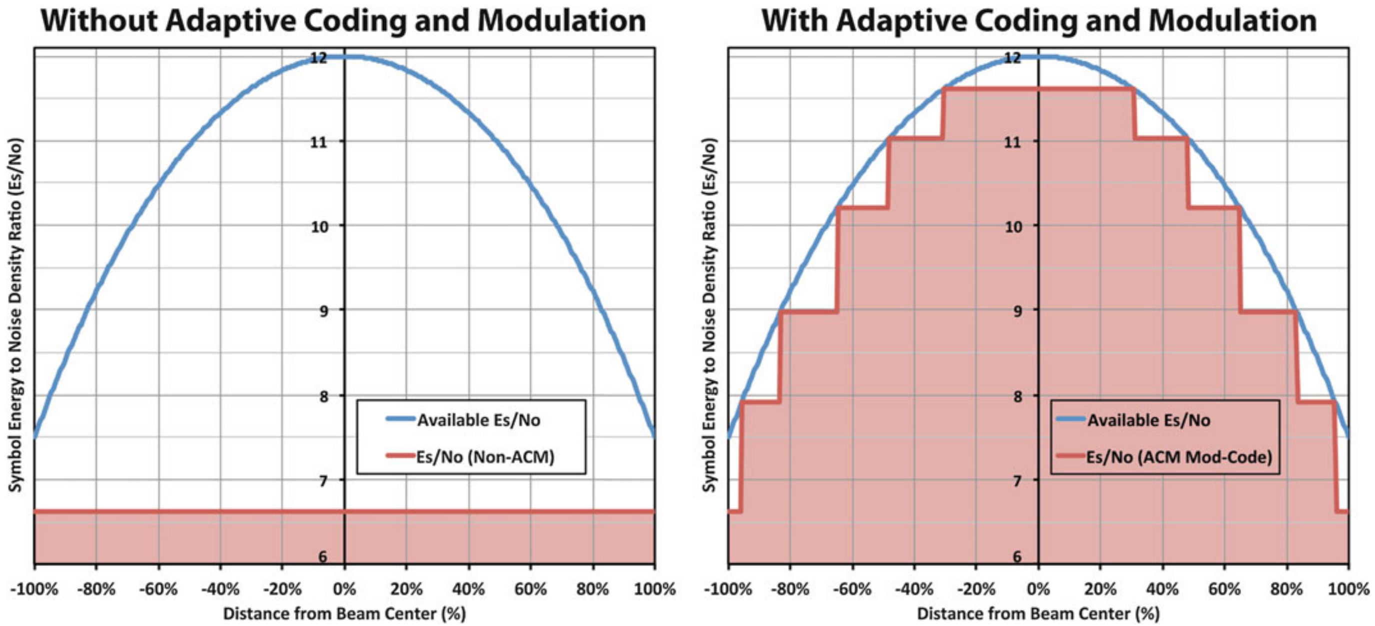


Fig. 13 Adaptive coding and modulation results in more effective use of the signal power available across each beam

Table 3 Adaptive coding and modulation increases capacity by 35 %

Beam location	Fraction of beam area	Modulation mode	Modulation order (N)	Code rate (R)	Spectral efficiency bits/Hz (ξ)
Center	9.6 %	16APSK 5/6	4	5/6	3.33
↕	13.9 %	16APSK 4/5	4	4/5	3.20
	18.7 %	16APSK 3/4	4	3/4	3.00
	27.5 %	16APSK 2/3	4	2/3	2.67
Edge	30.3 %	8PSK 2/3	3	2/3	2.00
Weighted average spectral efficiency					2.67

Table 4 Net improvement in capacity of over 200 times

Reference: late 1990s Ku-band satellite capacity	1.0 to 1.5 Gbps
Additional spectrum at Ka-band	×3 to ×5
Better frequency reuse at Ka-band	×25
Highly efficient physical layer	×2 to ×2.5
Adaptive coding and modulation	×1.35
Net improvement in capacity	**Over 200 times**

While the highest leverage features are the benefits of Ka-band spectrum and frequency reuse, all of the advanced features integrated together are required to build satellites with 100 s of Gbps of capacity.

ViaSat/Exede Broadband Satellite Network

ViaSat-1 was the first satellite to exceed 100 Gbps and established a new benchmark for space-based broadband capacity when it was launched in 2011. The ViaSat-1 satellite, integrated into the ViaSat/Exede ground system, provides an excellent example of a high-capacity broadband satellite network. Figure 14 has a diagram of the overall ViaSat/Exede network topology.

Data Processing Centers/Core Nodes

The ViaSat/Exede network has five data processing centers (DPCs), also referred to as core nodes. Starting from Internet and following the system connectivity out to the customer premises, the DPCs connect the satellite network to the Internet at a high-speed peering points. Each DPC performs data routing, switching, Internet acceleration, traffic management, authentication, and other functions required to support up to five Gateway Earth Stations.

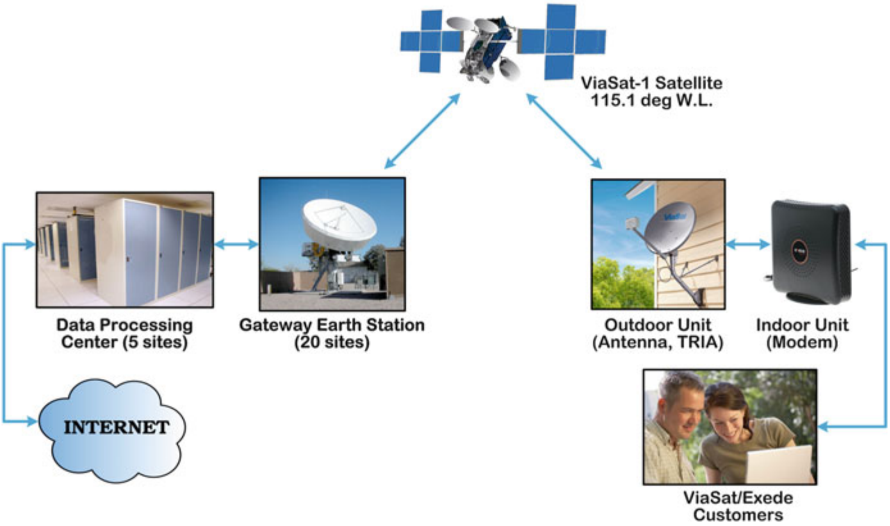


Fig. 14 ViaSat/Exede network end-to-end system diagram (Photos provided by ViaSat. All rights reserved.)

Exede Gateway Earth Stations

The ViaSat/Exede network has 20 Gateway Earth Stations that provide the bulk connections between the satellite and the Internet. The gateways are as identical as possible for ease of operations and maintenance. They reuse the same uplink and downlink spectrum and are located at least 300 km apart to avoid interference. Each gateway has 1500 MHz of allocated spectrum in the 30 GHz band and, by taking advantage of both LHCP and RHCP polarizations, supports 3000 MHz of forward link (Internet-to-customer) capacity. The gateway return links have 1500 MHz of allocated spectrum on both polarizations in the 20 GHz band, providing a symmetric 3000 MHz of bandwidth in the return link (customer-to-Internet) direction.

Each gateway consists of an RF subsystem and a baseband subsystem. The RF subsystem features a ViaSat 7.3 m Ka-band antenna with hub-mounted transmitters, receivers, frequency converters, and antenna-tracking equipment. The key functions of the baseband subsystem are performed by the ViaSat satellite modem termination system (SMTS). The forward and return link waveforms from every customer modem terminate at an SMTS. The SMTS does adaptive coding and modulation in the forward direction, demodulation of ACM waveforms in the return direction, data encryption and decryption, adaptive power control, time and frequency management, load balancing, switching and routing, and fault management of the end-to-end network. Each SMTS processes more than 500 MHz of bandwidth in each direction and can serve tens of thousands of customers. The baseband system has 6–8 SMTSs allowing each gateway to serve from 50,000 to 75,000 customers.

The 20 gateways and 5 DPCs are interconnected with fiber optic links. Each gateway connects to at least two DPCs over geographically diverse paths to ensure high reliability. In the unlikely event of a fiber outage between a gateway and its primary DPC, the gateway will switch all traffic to its alternate DPC to restore service.

ViaSat-1 Satellite

ViaSat-1 was manufactured by Space Systems Loral (SSL) and boosted into geostationary transfer orbit on October 19, 2011, by an ILS/PROTON launch vehicle from Baikonur Cosmodrome in Kazakhstan (International Launch Services 2011). SSL flew the satellite from transfer orbit to its assigned geostationary orbital slot at 115.1 ° west longitude (W. L.) (ViaSat 2011) (Fig. 15).

ViaSat-1 has 72 customer beams over the United States and Canada, including coverage of Hawaii and the more populated areas of Alaska. The satellite also has 20 gateway beams that complete the two-way connection between each customer's terminal and the Internet (Table 5).

A map of the ViaSat-1 beam coverage is shown in Fig. 16. Customer beams are shown in blue and gateway beams are shown in green. The ViaSat-1 coverage area is unique and was designed to achieve two primary objectives:

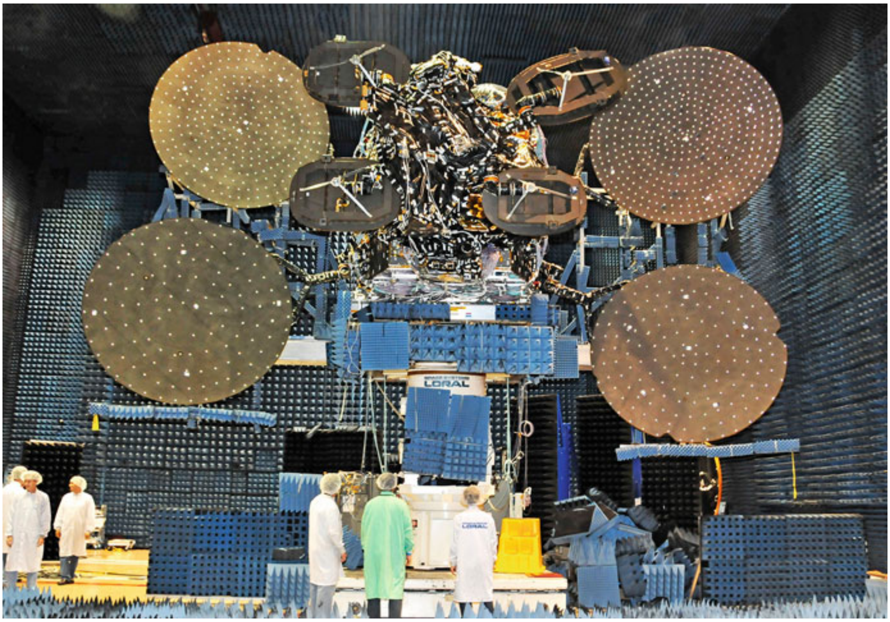


Fig. 15 ViaSat-1 satellite in compact antenna test range at Space Systems Loral (Photo courtesy of Space Systems Loral. All rights reserved)

Table 5 ViaSat-1 satellite key characteristics

NORAD ID/international code	37843/2011-059A
Manufacturer	Space Systems Loral
Separated mass	6740 kg
Launch vehicle/date	Proton M/Briz M 19-October-2011
Geostationary orbital position	115.1 deg W.L.
Coverage area	Contiguous United States, Alaska, Hawaii, Canada
Antenna beams	Customer beams: 72 Gateway beams: 20
Total capacity	>140 Gbps
Operational life	>15 years

1. Place customer beams over the areas of highest demand, generally where the population density is above a minimum threshold.
2. Place gateway beams away from customer beams to allow all allocated spectrum to be reused for both gateway beams and customer beams.

Four exceptions were made to objective #2 – gateway beams and customer beams do overlap at four sites: Denver, Colorado; Tucson Arizona; Honolulu, Hawaii; and Anchorage, Alaska. These exceptions were made, and a modest compromise to maximum capacity was accepted, to allow the satellite to serve these strategically important markets.

Since the ViaSat/Exede network consists of three satellites, the lower demand areas in the western United States not covered by ViaSat-1 are serviced by the combined capacity of WildBlue-1 and Anik F2. The Exede service is available across all 48 contiguous US states, although customer speeds in areas served only by the lower capacity WildBlue-1 and Anik F2 satellites are not as high as the peak speeds offered in areas served by ViaSat-1.

The space-to-ground (forward) links are in the 20 GHz segment of the Ka-band and the earth-to-space (return) links are in the 30 GHz segment of the Ka-band. The customer beam pattern uses four-color reuse, with two 500 MHz colors and two 1000 MHz colors. Half the customer beams has 500 MHz in each direction; the other half has 1000 MHz in each direction.

Adding the bandwidth across all 72 customer beams, the total network access bandwidth is 108 GHz. Based on an average spectral efficiency of $\xi = 1.5\ \mathrm{bits/Hz}$ in the forward direction and $\xi = 1.2\ \mathrm{bits/Hz}$ in the return, the total capacity of the satellite exceeds:

$$\mathrm{C} = 54{,}000 \times 1.5 + 54{,}000 \times 1.2 = 145{,}800\ \mathrm{Mbps} \tag{14}$$

When ViaSat-1 was first announced, even satellite experts questioned the claims that ViaSat-1 would have more than 140 Gbps of capacity. The capacity is quite real however. At the time of launch, ViaSat-1 had more capacity than that of all other satellites over North America combined.

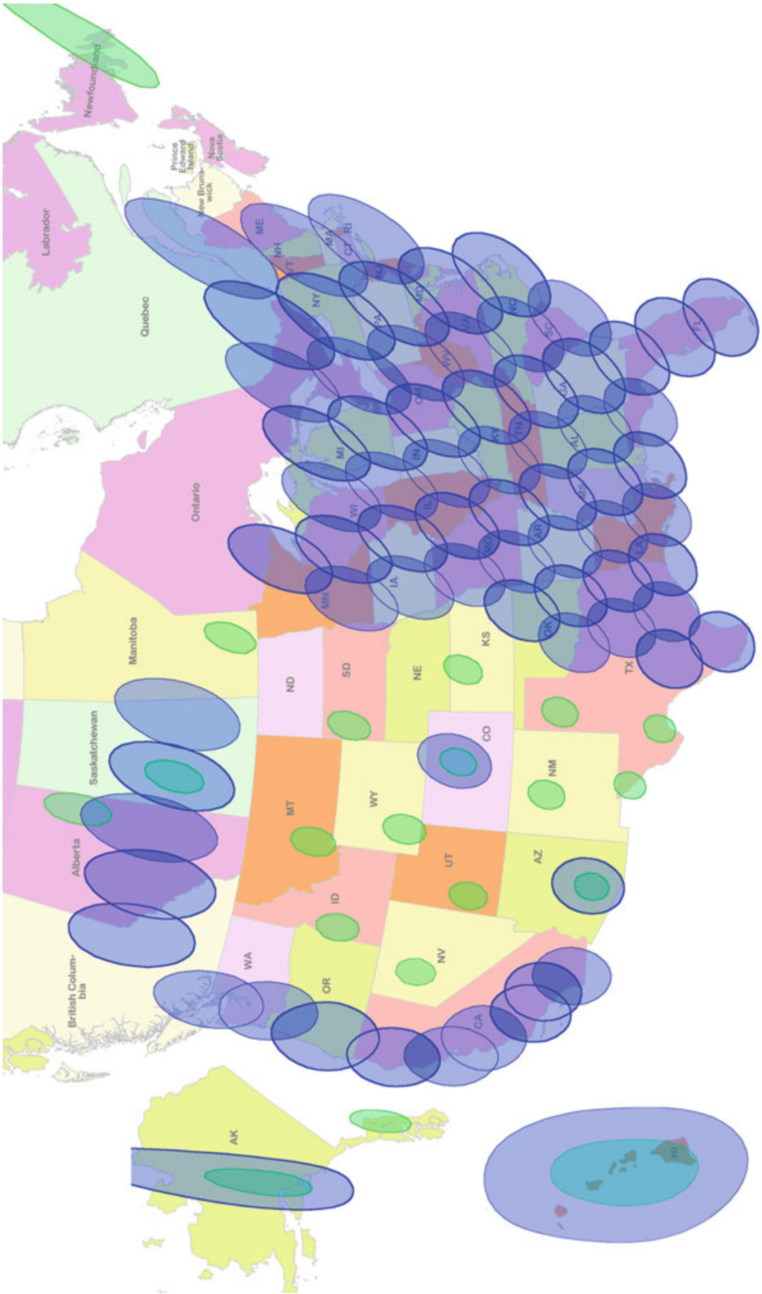


Fig. 16 ViaSat-1 beam coverage over the United States and Canada (Photo provided by ViaSat. All rights reserved)

Exede Customer Terminals

The availability of low-cost customer terminals has been a key enabler for widespread adoption of satellite broadband. The total cost of a million customer terminals, including installation at the customer premises, may exceed the capital cost of the entire system – satellite, launch, launch insurance, gateways, and DPCs all added together. ViaSat has manufactured over 2,000,000 broadband satellite customer terminals, focusing on continuous improvements in cost, performance, reliability, and ease of install (Fig. 17).

A ViaSat customer terminal consists of an outdoor unit and an indoor unit. The standard outdoor unit has a 75-cm antenna and a microwave electronics unit known as a transmit/receive integrated assembly (TRIA). The modem connects to the customer's local area network through a wired or wireless Ethernet connection and acts as the customer's bridge or gateway into the Exede network.

The ViaSat customer antenna uses a lightweight stamped steel reflector with very low surface tolerance to ensure high efficiency and low sidelobes. The mount has an oversized mast and braces for extra stiffness under the highest expected wind conditions. A fine adjust mechanism is built into the mount to make it easy for installers to point and peak the antenna with minimal pointing error.

The TRIA is an integrated transmitter-receiver fully sealed in an aluminum-zinc housing and designed for high reliability and long life in even the most severe outdoor environments. The state-of-the-art receiver uses very low noise transistors and has an excess noise temperature of approximately 75 K. The transmitter features

Fig. 17 ViaSat/Exede service outdoor unit (Photo provided by ViaSat. All rights reserved)

Fig. 18 ViaSat/Exede service indoor unit or modem (Photo provided by ViaSat. All rights reserved)

a nominal 4-W gallium nitride (GaN) single-chip high-power amplifier, delivering more than 38 dBm at the waveguide output. ViaSat has developed a novel way of integrating the orthomode transducer and circular polarizer into the TRIA housing for compactness and low cost. The unit has an electronic switch that changes the transmit/receive polarization between RHCP/LHCP and LHCP/RHCP in response to commands from the indoor unit.

To address both ease and accuracy of install, the TRIA has a built-in beeper to aid the installer in pointing and peaking the antenna. Using the antenna fine adjust feature and the TRIA beeper function, an installer can easily achieve better than 1 dB pointing accuracy without a meter and with just a single tool to tighten the antenna bolts and coaxial cable connectors (Fig. 18).

The indoor unit allows customers to connect end user devices or local area networks to the Exede network. The Indoor Unit performs adaptive coding and modulation, power control, TCP/IP data processing, Internet acceleration, encryption/decryption, performance reporting and fault detection, and other functions. The customer interface is a standard Ethernet port and offers the same Internet connectivity to PCs, home routers, and other end user devices as fiber, wireless, and cable modems. The fact that the Internet service was delivered over satellite is entirely transparent to customer.

ViaSat indoor units are optimized for high-volume manufacture and produced on fully automated production lines. The units are highly reliable and easy to install. The modem software is upgradable over the air at initial installation and throughout the life of the device. A low-power sleep mode minimizes electric power consumption when little or no traffic is transiting the network.

Next-Generation High-Throughput Satellites

ViaSat has multiple generations of new broadband satellites in development, with each generation designed to deliver higher capacity and better bandwidth economics. While the best bandwidth economics are obtained with larger 6 metric ton class satellites, ViaSat is also developing architectures that scale down to smaller more affordable satellites was well. Mid-sized 2–4 metric ton satellites may be a better match for national or regional systems where the coverage area is half or a quarter the size of the contiguous United States. The total capital investment required for a well-designed 2 to 4 metric ton satellite may be less than half that of a large satellite yet still provide excellent bandwidth economics over a smaller coverage area.

In May 2013, ViaSat announced the construction of an all-new ViaSat-2 satellite with more than 350 Gbps of capacity, greater than 2 1/2 times the capacity of ViaSat-1 (ViaSat 2013). In addition, ViaSat-2 will have more than seven times the coverage of ViaSat-1, providing service across North and Central America and the Caribbean, with extended coverage of commercial air and maritime routes over the North Atlantic between North America and Europe. ViaSat-2 is being built by Boeing Space & Intelligence Systems in El Segundo, California.

In July 2014 ViaSat and Paris-based Eutelsat Communications announced a roaming agreement allowing mobile satellite customers, on ships and aircraft, for example, to roam between ViaSat-2 coverage and the area covered by Eutelsat's KA-SAT satellite (ViaSat 2014). The combined coverage of ViaSat-1 and KA-SAT is show in Fig. 19. When ViaSat-2 becomes operational, commercial aircraft will be able to take off from Los Angeles, California, and fly to major cities in Europe (London, Paris, Barcelona, Rome, etc.) with no loss of connectivity.

Fig. 19 Map showing the combined coverage of ViaSat-1 over the North Atlantic connecting to Eutelsat coverage to allow mobile roaming into Europe (Photo provided by ViaSat. All rights reserved)

In March 2015, ViaSat announced the availability of a Flexible Broadband System incorporating a new satellite design referred to as “ViaSat-2 Lite” (ViaSat 2015). The system is designed to provide the best satellite bandwidth economics possible in a smaller more affordable package, tailored to regional operators. ViaSat-2 Lite is a low-cost Boeing 702SP-based satellite using next-generation ViaSat payload technology. ViaSat’s Flexible Broadband System provides high capacity with a whole new level of flexibility, scalability, and affordability.

In November 2015, ViaSat Chairman and Chief Executive Officer Mark Dankberg announced the upcoming procurement of a ViaSat-3 satellite constellation (de Selding 2015b). The constellation will consist of three ViaSat-3 satellites providing full global coverage. Each satellite will have an incredible 1 Tbps of broadband capacity. Construction of the first satellite is expected to take approximately 4 years. ViaSat plans to launch the first ViaSat-3 over the Americas and then launch additional satellites approximately 1 year apart to cover the rest of the Earth.

Conclusion

ViaSat continues to develop advanced space and ground system technologies to ensure that each new generation of satellite networks will deliver the broadband capacity and performance that the market demands.

Geostationary broadband satellites, such as ViaSat-1 and ViaSat-2, can deliver 100 s of Gbps of capacity over a continent or into a regional area, providing Internet access to hundreds of thousands of people regardless of location or population density. No other network access technology offers anywhere near the degree of geographic coverage, high flexibility, and time to market as a satellite in orbit. Ground-based technologies become unaffordable in rural and remote areas whereas satellite service costs the same everywhere.

Non-geostationary satellites butter-spread their capacity around the globe and add the additional complexities of large constellations, satellite-to-satellite handoffs, and expensive ground terminals with tracking antennas to follow satellites across the sky. For customers outside urban areas, geostationary satellites deliver better and more affordable broadband services than any other technology.

In just a few years, with the launch of ViaSat-2 and ViaSat-3 satellites, ViaSat will have fast and reliable satellite broadband available almost everywhere on earth. We believe geostationary satellite will remain the preferred broadband platform for aircraft, ships at sea, and land mobile applications that require high-speed access to the Internet. We also believe that geostationary satellite will remain the most effective broadband delivery technology for residential and enterprise customers outside more urban areas.

ViaSat is incredibly proud of its employees whose hard work, dedication, and commitment have allowed the company to reach its goal of becoming the world leader in high-capacity broadband satellites. ViaSat is confident that the innovation, creativity, technology advancements, and engineering breakthroughs will continue as the company takes on the challenges of building higher and higher-capacity satellite networks that deliver more and more cost-effective bandwidth.

References

R. Brandt, Birth of a Salesman. Wall Street J. (2011, October 15)

C. Conti, New two-way satellite internet services battle for rural market. Streaming Media Mag. (2000, December 06), http://www.streamingmedia.com/Articles/News/Online-Video-News/New-Two-Way-Satellite-Internet-Services-Battle-for-Rural-Market-63453.aspx. Viewed 15 Jan 2016

P. de Selding, Unexpected demand takes wildblue, telesat by surprise. Space News (2004, June 29)

P. de Selding, EchoStar explains why Jupiter-2 is launching on Atlas instead of Ariane. Space News (2015a, August 06)

P. de Selding, ViaSat willing to bet big on super-high-throughput satellites. Space News (2015b, November 10)

DIRECTV, *Form 10Q Quarterly Report*, El Segundo, CA, 08 May (2015), p. 44

DIRECTV, *Company Profile* (n.d.), http://www.directv.com/DTVAPP/content/about_us/company_profile. Viewed 15 Jan 2016

DISH Network, *Quarterly Investor Summary*, Englewood, Colorado, 09 Nov (2015), p. 11

DISH Network, *Company Info* (n.d.), http://about.dish.com/company-info. Viewed 16 Jan 2016

Eutelsat Communications, *Eutelsat Communications via KA-SAT, Global Communication Solutions for any Situation*, brochure, Paris (n.d.), p. 4–6

ETSI EN 302 307–1, *Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications; Part 1: DVB-S2 v1.4.1* (2014), https://www.dvb.org/standards. Viewed 23 Jan 2016

R. Gallager, *Low Density Parity-Check Codes* (Massachusetts Institute of Technology, 1963), http://www.rle.mit.edu/rgallager/documents/ldpc.pdf. Viewed 23 Jan 2016

Gilat Satellite Networks, *Gilat-To-Home Becomes StarBand Communications – Creating a New Category of High-Speed Internet,* Press Release, 11 Sept (2000)

P. Hadinger, Inmarsat global Xpress: the design, implementation, and activation of a global Ka-band network, in *33rd AIAA International Communications Satellite Systems Conference*, 7–10 Sept (2015)

C. Henry, ViaSat, boeing sign technical agreement to offer line-Fit Wi-Fi.\ ViaSatellite Mag. (2015, September 17)

Z. Honig, United activates ViaSat's blazing-fast satellite WiFi on select 737s. EnGadget (2014, February 26), http://www.engadget.com/2014/02/26/united-satellite-wifi/. Viewed 15 Jan 2016

International Launch Services, *ILS Proton Successfully Launches ViaSat-1 for ViaSat, Heaviest Satellite Launched on ILS Proton,* Press Release, 20 Oct (2011)

Internet Live Stats, *Internet Users in the World* (n.d.), http://www.internetlivestats.com/internet-users/#trend. Viewed 15 Jan 2016

Internet Society, *Brief History of the Internet* (n.d.), http://www.internetsociety.org/internet/what-internet/history-internet/brief-history-internet. Viewed 15 Jan 2016

J. MacKay, R. Neal, Near Shannon limit performance of low density parity check codes. Elect. Lett. **32**(18) (1996)

C. Shannon, W. Warren Weaver, *A Mathematical Theory of Communication* (The University of Illinois Press, Champaign, 1964)

Statistica, *Number of monthly active Facebook users worldwide as of 3rd quarter 2015* (n.d.), http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/. Viewed 15 Jan 2016

ViaSat, *ViaSat-1 Satellite Reaches Geosynchronous Orbit*, Press Release, 03 Nov (2011)

ViaSat, *ViaSat Sets Guinness World Records Title for the Highest Capacity Satellite*, Press Release, 06 Mar (2013a)

ViaSat, *ViaSat Announces Next Generation Broadband Satellite*, Press Release, 16 May (2013b)

ViaSat, *High-Speed Internet Now Flying as JetBlue Launches Service Using ViaSat High-Capacity Ka-band Broadband,* Press Release, 12 Dec (2013c)

ViaSat, *ViaSat and Eutelsat Enter First-of-a-Kind Agreement to Link High-Capacity, Ka-band Satellite Networks*, Press Release, 01 July (2014)

ViaSat, *New Space/Ground System From ViaSat and Boeing Designed to Accelerate Spread of High-Speed Broadband Worldwide*, Press Release, 16 Mar (2015)

Virgin America Airlines, *Virgin America Partners with ViaSat to Offer Faster, Higher Quality WiFi in the Sky: Travelers at 35,000 Feet Can Now Stream Video Content from the Internet*, Press Release, 07 July (2015)

M. Zuckerberg, Web log post, 27 August. Facebook (2015), https://www.facebook.com/zuck/posts/10102329188394581, Viewed 15 Jan 2016

Distributed Internet-Optimized Services via Satellite Constellations

Joseph N. Pelton and Bernard Jacqué

Contents

Abstract

One of the most significant recent developments in satellite communications has been the sudden resurgence of large-scale constellation satellite programs to provide broadband services. This has occurred some 20 years after the several unsuccessful attempts to deploy such huge constellations like Teledesic in the USA and Skybridge in Europe. These were never deployed for several reasons

---

J.N. Pelton (✉)
International Space University Arlington, VA, USA
e-mail: joepelton@verizon.net

B. Jacqué
Thales Alenia Space, Cannes, France
e-mail: bernard.jacque@thalesaleniaspace.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_96

that included financing and the bursting of the Internet bubble at the end of last millennium.

Of course, other telecommunication constellation programs have been deployed since that time (Globalstar 1st and 2nd generation and the Iridium and soon Iridium Next for instance). These systems, however, were designed to address narrower band services in low band frequencies (L and S-band) for mobile telephony and low-medium rate data.

These new types of constellations that are currently either recently operational or under design and development are intended to mainly provide broadband Internet-optimized services with the ability to offer low latency performances compared to geostationary satellite alternatives. These new systems, and in particular the O3b and OneWeb networks, both headquartered in the tax haven Jersey Island, UK, as well as the Leosat initiative, from a Delaware registered company, have been described as "disruptive," "game-changing," and "innovative" in their architecture (P. de Selding, "Never Mind the Unconnected Masses: Leosats Broadband Constellation is Strictly Business", Space News, Nov. 20, http://spacenews.com/nevermind-the-unconnected-masses-leosats-broadband-constellation-is-strictly-business/, 2015).

One remarkable aspect is that each one represents very different and specific approaches to addressing broadband applications by satellite. Some are in Low Earth Orbits (LEO) at about 1000–1500 km altitude while O3b is flying much higher in Medium Earth Orbits (MEO) at about 8000 km altitude. O3b is an equatorial MEO system of 12 full-sized satellites. Others are intended to represent a network of some 100 satellites of 700–1300 kg mass while yet others are requiring several hundred spacecraft of 175–200 kg mass. Some are envisioned to provide "local" services connecting a gateway with users in visibility of a sole spacecraft scale (Proposed Leo Sat Constellation, Space News, March, 2015 http://spacenews.com/proposed-leosat-constellation-aimed-at-top-3000/Last. Accessed 9 Dec 2015).

Other systems are designed with a more interconnected architecture for connecting users to a gateway or another user that can be located far away in an another continent thanks to inter-satellite links.

However, each concept in its own way is raising a number of regulatory and technical challenges.

This trend to deploy new broadband constellations for fixed and mobile satellite services started with the deployment of the medium earth orbit O3b constellation in 2013 and 2014, and now OneWeb has selected in mid-2015 Airbus Defence and Space as a joint venture partner to invest and manufacture some 900 small satellites (i.e., operational plus spares) to be deployed starting in 2018. There may be other companies that follow suit to deploy similar so-called mega-constellation systems, but currently OneWeb is the only such LEO constellation system under a development contract to manufacture and launch such a large-scale network. Another possible system has started design and engineering phases such as Arlington, Virginia-based LeoSat (although officially headquartered in Delaware). This system is exploring an 80 satellite that might be expanded to about a 110 satellite constellation. This project involves Thales Alenia Space. Then there is the announced effort whereby Singapore Space Intelligent IoTS Pte. Ltd. (SSII) is partnering with German satellite maker OHB System to develop the world's first Asia-based low Earth orbit Internet constellation (Singapore Space Intelligent IoTS Pte. Ltd. http://www.ssii.sg/Last. Accessed 8 Mar 2016).

Finally, there have been reports of a Space-X backed system that might deploy as many as 4000 satellites in a massive mega-LEO system.

These systems are, however, not yet contracted to manufacturing and thus are not addressed here in details. The Space X system would presumably like OneWeb involve quite a huge number of small satellites and be aimed at underserved developing countries' markets among other more mature markets. The LeoSat system in contrast would involve much larger and capable satellites with more than 2 kW of power and would be aimed at meeting the special needs of the largest corporations in the world (Propose LeoSat).

The implications of such large-scale constellations of small satellite are manifold. These new type satellite networks would seem to revolutionizing the cost of manufacturing and launching spacecraft, concerns about radio frequency allotments and protection from interference, orbital debris build-up and removal, collision avoidance, management of liability concerns, and more. What is clear is that the deployment of those satellite constellations in low earth orbits will provide a satellite network that is quite different in many ways when compared to GEO satellites. The LEO satellites would typically be some 30 times closer to the Earth's surface than GEO satellites with about 60 times less transmission delay for a round trip. Clearly such a network can accommodate latency-sensitive applications in Internet data transmissions (i.e., TCP/IP protocols) with greater efficiency and support voice conversation services with greater facility. On the other hand, their closer vicinity with the earth's surface restricts their coverage reach, and this requires many more satellites for a continuous earth coverage.

This new trend to deploy satellite constellations for broadband satellite services is occurring in close parallel with the development and deployment of very high throughput satellites (VHTS) in geostationary orbit that provide much greater capacities at lower costs. Clearly these parallel and potential "disruptive" trends to deploy even more capacitive HTS and low earth orbit constellations could serve to drive down costs and make available new digital services to consumers around the world at much lesser costs. We can even expect in a midterm the integration of both complementary solutions, the very high capacitive geostationary HTS systems providing a much higher data rate per user together with mega-constellation services offering low latency data flow and a world coverage including the poles. The involvement of well-known international satellite operators of geostationary fleet such as Intelsat that is involved in the OneWeb project or SES in the O3B is probably a revealing clue.

This chapter describes the various systems that have been implemented or now in production to be deployed in the coming years – especially O3b and OneWeb. This chapter provides some of the basic technical and operational characteristics

of these new systems. It also addresses the various types of new services that are being offered or planned by these types of networks.

It was thought in the 1990s, a mega-LEO satellite system for broadband fixed satellite services similar to OneWeb might deployed. This network, which was named Teledesic and financially backed by Bill Gates, Greg McCaw, and venture capitalist Ed Tuck, was proposed along with about 15 other Ka-band satellite networks. The Teledesic system and the other proposed Ka-band systems were never deployed – except for the Wild Blue Geo satellite network (renamed the Ka-band satellite system) and which was delayed over a decade in its actual launch and deployment. Today, some 20 years later, the viability of such large-scale lower earth orbit satellite systems now seem to be economically feasible again.

Thus, the first generation of O3b has been designed, manufactured, and successfully deployed, and rapid progress is being made to design, manufacture, and launch OneWeb in a not so distant future. The advent of 3D printing, advanced manufacturing techniques taking benefits of more automated processes for large-scale production and testing, more extensive use of commercial off-the-shelf (COTS) components, and new commercial systems to launch small satellites at low cost have combined in a positive fashion to greatly reduce the cost of building and launching such satellites.

New satellite networks, born out of Silicon Valley, such as the Skybox constellation for remote sensing, now acquired by Google, have served to unveil a whole new pattern of commercial satellite business. “Disruptive” technologies and new satellite system architectures are thus the hot trend of the day driven by so-called New Space commercial ventures.

Keywords

Airbus Defence and Space • Disruptive technologies • Google • High throughput satellites (HTS) • Hughes Network Systems • Ka-band satellites • Ku-band satellites • Leosat • Liberty Media • Mega-LEO Constellations • O3B constellation • OneWeb • Qualcomm • SES Global • Silicon Valley • SpaceX • Thales Alenia Space • Via Satellite Virgin Galactic’s Launcher One • Greg Wyler

Introduction

The field of commercial satellite communications has evolved in a continuous but sometimes “jerky” manner since its start in 1965 for a half century. The past 5 years has been one of those “jerky” periods, where there have been a number of disruptive technologies introduced. This is, in part, in response to expanded global networking needs and the greatly expanded capabilities of broadband fiber optic cable systems that can operate at terabit/s speeds and very little transmission delay. These changes are currently having a significant and explosive impact on the communications satellite industry and network design. These new technical capabilities and new market needs have helped create new patterns of competitive technologies and given rise to new competitive systems and new players on the global scene.

First of all perhaps the biggest change is that there are now high throughput satellites in GEO orbit that have 10–25 times the capacity of previous satellites, thanks to multibeams frequency reuse technologies and solutions, and are sharply competitive with more "conventional" satellite networks. These new networks are having a huge impact on the cost of service, a shift from transponder pricing to channel pricing, and competitive markets on a global scale. The race for more capacity seems to be accelerating since the beginning of 2016 with a new generation of Terabit satellites announced for a start of operation in 2020 and new concepts such as satellite optical rings being seriously considered. In close conjunction with the new more cost-competitive networks, there is new and increasing demand for broadband data services worldwide (even at high attitudes and over the pole) for mobile services to support commercial aircraft and new northern maritime routes, some of them being particularly sensitive to transmission delays associated with satellites in GEO orbit. In addition, there are also new techniques associated with the design, manufacture, acceptance testing, and low-cost launching of small satellites that have and will create new economic efficiencies with regard to the deployment of larger scale constellations of satellites in medium or low earth orbit.

Finally, there are new designs under development for ground systems. These new "smart" ground antennas will be able to track using electronically-formed beams rather than mechanically formed beams. This is expected to lead to relatively low cost and efficient tracking of LEO/MEO satellites. Those efficient and low cost terminals are key enablers for the good overall economic figures of these evolving mega-LEO constellations businesses. All of these factors have come together to create new opportunities – as well as challenges – for medium and low earth orbit constellations that are designed to meet previously unserved networking needs. Some of these systems are particularly optimized to meet market demands related to the developing countries of the world and their unmet needs. These new constellations as now designed and engineered can meet the needs of many developed economies as well. This chapter addresses the evolution of these new constellations for telecommunications and networking purposes and provides an overview of the space segment and ground segment system design, the markets to be served, and the issues that these new systems engender.

The first system of this kind is the O3b medium earth orbit that was deployed in 2013 and 2014. The OneWeb constellation, now under contract for manufacture and launch, that is planned for low earth orbit is the second. This constellation is planned to contain nearly 800–900 small satellites plus many spares and is scheduled for deployment in 2017 and 2018. This LEO-based system represents an even more ambitious initiative in terms of a large-scale constellation. This project and its technical design in many ways recalls the Teledesic or Skybridge satellite networks that was envisioned in the 1990s but ultimately both went bankrupt. Other aerospace entities have indicated less specific plans to build and deploy such large-scale constellations, perhaps on an even larger scale.

These new type constellations could give rise to new opportunities and lower costs for broadband data services around the world, but they give rise to concerns with regard to frequency interference to GEO satellites, a possible significant increase in orbital debris in the event of a collision in the orbital regions where these satellites are to be deployed or after satellite end of service period in their postmission disposal orbits, and the extent to which deployment of such large scale systems could preclude other countries or entrepreneur from being able to launch similar systems in the future.

This chapter provides information about the precursor satellite systems that have set the stage for the 03b and OneWeb networks and others that may follow. It then describes the two systems that have been or will be deployed and the markets they intend to serve. It then concludes with a consideration of the various technical, regulatory, market, frequency coordination, and orbital debris related issues that these new satellite constellations could give rise to and the various processes now underway to address such concerns.

History and Background to Medium and Low Earth Constellations for Communications

In the 1990s, a team of engineers were exploring new satellite architectures that might allow the successful deployment of satellite systems using the previously unutilized Ka-band frequencies. These designers were seeking a satellite system design that was optimized to provide Internet-based broadband services and also overcome the problem that rain attenuation posed, in particular, at that time for these very high frequencies (since, dedicated mechanisms has been elaborated within the waveform to manage the coding and the data rate during attenuation events).

This new type satellite network that was first called the Calling Satellite System and then renamed Teledesic represented a radical departure from the geosynchronous – or Clarke orbit – satellite networks that had dominated satellite communications architecture up to that time. The new system would deploy a massive amount of satellites in low earth orbit. The total network plus spares would have involved the launching of some 920 satellites (i.e., 840 plus 80 spares). The key elements of the design involved several new ideas: (i) the satellites would be designed and qualified so that they could be manufactured and quality tested like VCRs on a largely automated production line so that their unit cost would be much less than a typical satellite; (ii) the Teledesic satellites, by being deployed some 30 times closer to the ground than a GEO satellite (that orbited at 35,870 Km from the Earth's surface), would thus have 60 times less latency roundtrip (and thus better conversation continuity and data quality); and (iii) further, as a special bonus for the Ka-band satellites, there would also result in 900 times (i.e., proportional to $1/30^2$) less transmission power loss or what engineers called free-space path loss (Pelton 2013, Satellite Communications).

(The reason why that advantage would be "30 squared" was because the spreading out of transmitted power from the circular-shaped parabolic antenna required a calculation of the area of the expanding circle of the transmission. This is based on the formula $1/\pi r^2$.)

The Ka-band satellites because of their high frequencies are more sensitive to signal attenuations when there is heavy rain. Possible use of even higher frequencies in the Q/V spectrum bands in the future will represent an even greater challenge. The frequencies that are involved are of a wavelength similar to a drop of rain, which means the rain can, in effect, "bend" the path of the transmission. Such tiny radio wavelength is much more sensitive to rain attenuation and other forms of precipitation than the C-band or Ku-band. The heavy precipitation serves to distort the signal and make it harder to receive and thus power loss (or path loss) compounds the problem. This major power advantage obtained by having the satellites in low earth orbit, plus relative high masking angles (i.e., high elevation look angles to the satellites) were seen as key to this type of Ka-band satellite constellation being able to function during a rain or snow storm. This historical background is provided, because the envisioned OneWeb satellite network, although planned for the Ku-band, intends to utilize a technical design similar to that of the Teledesic system.

There was of course a down side to their innovative design – either for OneWeb or Teledesic. First the various beams on the satellite would have to be acquired, tracked, and handed off among many different beams as the satellite traveled overhead. This would be very hard to do because with so many satellites in the constellation, each with several beams, there would be very frequent handovers to successfully achieve continuity of service. These handovers would, for instance, be measured in seconds rather than minutes. Second, the design of the satellite and it's especially designed antenna system would "paint" a coverage on the ground as the satellite moved over head. This meant that the satellite antenna design and computer processing capabilities would need to be more complex so that the ground antennas could be simpler. This, of course, adds complexity and possibly costs to the satellites. Third, there would be yet another driver of complexity in the satellite design which is needed to avoid interference with GEO satellites when the lower orbiting satellites were transmitting in the orbital arc. In the case of OneWeb, the satellite earth pointing axis moves as needed to steer away transmissions that would illuminate and interfere with GEO satellites.

The added cost of the rapidly tracking and steerable antennas on the satellite, the design elements to avoid interference with GEO satellites, the added cost of building and launching quite so many satellites, and the difficulty of switching quickly between and among many beams certainly were a tall order for the technology of the 1990s. The bottom line was that these economic and technical issues proved fatal to the Teledesic system and the system was never placed under contract and built.

Although this Teledesic system was formally filed and licensed by the FCC in the USA and then filings made with the International Telecommunication Union (ITU), the cost of the system when put out to bid was judged to be too high (estimated about 9 billion USD at that time). The project was ultimately canceled despite quite a few millions of dollars having been spent on this very ambitious satellite program. Despite the fact that more than a dozen other Ka-band satellite systems were filed with the FCC in the USA in the 1990s, none of these systems – whether LEO constellations or GEO orbit satellites – were built and deployed until years later. Only about 15 years later were Ka-band GEO satellites such as “Wild Blue,” renamed the “Ka-band Satellite,” actually deployed and operated by Viasat together with the Viasat-1 satellite to offer a continuous CONUS broadband coverage.

Today, there are a number of GEO-HTS based Ka-band satellites operating – most notably ViaSat 1 and Echostar/Jupiter-1 over North America, Ka-Sat from Eutelsat and Hylas-2 from Avanti over Europe, Yahsat 1B from Yahsat and Badr-7 from Arabsat over Middle East, the Thaicom IPstar over Asia, and the three Inmarsat 5 satellites for a worldwide coverage. Others systems are to be deployed or entering operation soon with Viasat-2 (2017 launch), the Intelsat EPIC (29e and 33e later in 2016), and Echostar Jupiter-2 (later in 2016) over North America. The need for Ka-band satellites with its broader spectrum range and the saturation of the lower C-bands and growing congestion in the Ku-bands has made the transition to Ka-band more and more desirable as far as retro-compatibility with legacy system is not a must for telecommunications companies, TelCo. But all of the Ka-band systems to date have been GEO satellite networks. None had been deployed in the low earth orbits as constellations to be used for the end-user broadband connection to satellite.

Today, only the O3b satellite constellation using Ka-band frequencies for fixed satellite services (FSS) has been deployed in a lower orbit (and this deployment is in medium earth orbit within the equatorial plan.) The mega-LEO OneWeb constellation that is planned for deployment in 2017 and 2018 has opted to use Ku-band frequencies for its satellite links but Ka-band spectrum for gateway links. This choice was driven by regulatory reasons (frequency filing in Ku-band for users). This has become a key enabler to implement such a system. The technical and economic challenges posed by the Teledesic design still have to be faced in the design and deployment of the O3b. This is even more so for the case of the OneWeb systems. Both of these systems were envisioned and championed by a man named Greg Wyler, who would set to bring new broadband capabilities to the developing world. His role in the creation of these systems will be discussed below.

And if O3b and OneWeb prove successful, then other lower earth orbit systems may well seek to follow even though frequency coordination issues and concerns about orbital debris may serve to lock additional systems. The high throughput LeoSat Constellation, with an initial network of nearly 80 satellites that can be increased based on demand, has indicated its intention to try to deploy its network as early as 2019 or 2020. It is seeking in its design to use powerful and capacitive intersatellite links to securely connect very distant users with very high data rates and will also see to utilize the benefits of new types of flat panel ground antennas currently under development that can generate beams electronically and thus address and track rapidly the various moving satellites in a more efficient and hopefully more cost-effective way.

The largest Leo constellation that has been actively considered by SpaceX has been conceived as having as many as 4000 satellites. The future of these types of constellations is both largely unknown and highly dependent on a wide range of quite different design considerations. These various known or pending constellation designs differ in terms of satellite size and power, constellation orbital configurations, satellite operating lifetime, nature of intersatellite links, launch arrangements, satellite antenna and ground antenna design, and technical arrangements to avoid interference with GEO satellites. One of the largest unknowns is the extent to which these communication satellite constellations can still avoid in-orbit collisions that would greatly increase the orbital debris problem and also successfully deorbit their satellites in a controlled manner at the end of life.

The First Communications Satellite Networks Were for Mobile Services

Design concepts for low and medium earth orbit constellations were conceived for the design of mobile satellite systems in the 1900s that were deployed or planned for deployment in the late 1990s and early 2000s. The Iridium Satellite System and the Globalstar Satellite System were designed as low earth orbit constellations. The ICO satellite system, which was a spin-off of INMARSAT, in contrast, was first envisioned a MEO constellation. Over time, when Iridium, Globalstar, and ICO declared bankruptcy, ICO was re-envisioned as a GEO satellite system design. The Iridium system of 66 satellites plus spares was deployed beginning in 1996–1997 and the Globalstar system was deployed very shortly after. The bottom line was that all three ventures – namely Iridium, Globalstar, and ICO all collapsed financially and went through bankruptcy proceedings. In the case of these three satellite networks, the main issue seemed to be a lack of market demand, but the technical performance and the high service price of the system perhaps partially contributed to the market failure. Certainly, the Iridium and Globalstar networks demonstrated that the overall control and network management of the satellites was difficult. During the first months of operation of the Iridium system so-called cockpit errors in terms of wrong commands to the constellation were a problem with just over 70 satellites in the network. This type of problem certainly raises questions about the difficulty of “cockpit management” of a network of over 800 operational satellites and spares of satellites in two grids.

The various mobile satellite communications systems, regardless of orbit, used much lower radio frequency bands (i.e., around 1.6/1.5 GHz rather than 30/20 GHz) and provided only narrow data/voice channels rather than broadband. In these mobile satellite systems design, it was recognized that the mobile user would have to have tracking or omnidirectional antennas in any event since the customer would be moving and not be at a fixed location. Iridium and Globalstar also envisioned using networks with far fewer satellites (i.e., 50–70 satellite plus spares rather than 840 satellites plus spares envisioned for the Teledesic system). The backers of the Iridium, Globalstar, and ICO systems also thought that many millions of people would pay a premium for this premium service. The customer base ultimately turned out to be quite small. This was due to the fact that during the time the Iridium and Globalstar satellite networks were being designed, manufactured, and launched, the terrestrial mobile cellular systems had been built out and upgraded in their performance throughout the 1990s and early 2000s. The ultimate result was that all three mobile satellite systems, namely Iridium, Globalstar, and ICO were forced to declare bankruptcy because the market that they had hoped to serve had been captured by terrestrial cellular systems that were now much more pervasive in urban areas and had been upgraded so they had a thousand to ten thousand times more power so that inside calling and calling within cars was now possible. The bankruptcies of Teledesic, Iridium, Globalstar, and ICO convinced the financial markets that LEO and MEO constellations were risky propositions. Meanwhile, the GEO satellite networks that were succeeding were seen as much better business investments.

But over time, new technology, manufacturing techniques, and innovations in launch services have helped to change perspectives. What failed 20 years ago is now being seen with new eyes today. The fact that the new systems have lower latency and are seen as better suited for Internet services and optimized for unserved or underserved developing economies has opened the door to new investment – especially by digitally oriented companies in Silicon Valley such as Google. As noted earlier, one of the key people that have restored interest in communications satellite constellations is a man named Greg Wyler. He is today considered the father of both the O3b (Other three billion people) satellite system and OneWeb. The history of how these two systems came to be is useful and instructive.

Wyler initially engaged in an effort to upgrade the rural communications of Rwanda to meet modern telecommunications needs. Every concept that he explored using terrestrial technology failed to come close to providing a viable business plan that could one day even break even. Slowly he came to realize that only a satellite network that provided integrated coverage to the entire equatorial region of the planet where three billion people lived that were ill-served communications networks that could provide Internet connections. Wyler grasped that it was only satellites that could provide the connectivity and the modern information and communication technology (ICT) for Africa, South America, the Caribbean, the Middle East, and Asia in any reasonable time period and that could possibly be economically viable.

It was from this realization that the idea for the O3b satellite network was born. Wyler was a dynamo that used his financial investment "smarts" to convince a range of technology and communications companies to invest in O3b. He was able to convince Google, Liberty Global, SES of Luxembourg, Satya Capital, North Bridge Venture Partners, Sofina, and Allen & Company, plus HBSC bank to invest as well as to retain HSBC also to arrange debt financing to fund his ambitious project. Altogether a total of $1.2 billion in financing was put together in a remarkably short period of time in order to build and launch a medium earth orbit constellation that would circle Earth's equatorial orbit. Wyler formed the O3b company in 2008 and the satellites in the initial constellation went up in 2013–2014.

Teledesic, Iridium, Globalstar, and ICO projects all essentially began as the result of "technology push" provided by service providers, equipment suppliers, and investment backers. All of these systems failed financially. The market that was envisioned unfortunately never materialized.

O3b, in contrast, started from a market of "wannabe Internet users" that were seeking to be served in Africa, the Middle East, Asia, the Caribbean, and South

America. The question is still pending as to the extent to which OneWeb, in contrast to O3b, is indeed a response to market demand, or technology push, or perhaps a useful combination of both. What is clear is that the investors in OneWeb are indeed dominated by suppliers of the satellite system, ground equipment, and launch services.

What is known is that Wyler, after getting O3b underway, embarked on an even more ambitious project and for this project he developed essentially a whole new group of investors. After spending only a short stay at Google that was a major funder of O3b, he left and spun off his WorldVu company that began the even more ambitious OneWeb constellation. Many perceived this as moving from being a Google-sponsored project with O3b to become an Elon Musk and SpaceX sponsored enterprise.

By his concerted efforts Wyler has been able to raise most of the capital for the 12 satellite O3b network that costed about $1.2 billion. He has now raised $500 million from among his suppliers for the building and launching the OneWeb network of 648 satellites.

Wyler has been most adept in finding investors that would also be his equipment suppliers, his launch operators, as well as to find backing from the world's largest satellite service provider in Intelsat. In these arrangements, AirBus became the manufacturer of the satellites, SpaceX, Virgin Galactic's Launcher One, and Arianespace/Soyuz became the provider of launch services, while Echostar/Hughes Network Systems became the supplier of the innovative new ground systems for the OneWeb network. The danger that could arise in such an arrangement is that as the project shifts from one designed to meet market demand to one in which the suppliers push the products forward, the problems that manifested itself with Iridium and its bankruptcy could happen again. Intelsat's investment of a modest $25 million seems clearly an attempt to learn what the new market demand really is and whether this system is truly viable (de Selding 2015, One Web's Partners).

This project has moved ahead with remarkable speed from idea to firm contracts. It may represent a remarkable case of where the Arthur C. Clarke laws of prediction and his three stages of evolution of a project have been compressed in a remarkably short span of just a few years. Clarke's three stages of a project are whimsically set forth as:

> "**Stage 1**: It's impossible; **Stage 2**: It's possible but it's not worth doing; **Stage 3**: I said it was a good idea all along." (Pelton 2015, The Oracle…)

To date it seems O3b to have established itself as a new satellite system that has evolved at a time when market need for low latency data-oriented satellite networks and new technological and manufacturing capabilities have coincided in a positive way. The past history represented by Teledesic, Skybridge, Iridium, Globalstar, ICO, and even Orbcomm may have helped to overcome technological, economic, and market pitfalls that earlier networks have encountered. The future of new systems such as OneWeb, and others that may follow, clearly face stiff technical, economic, market, and other challenges. These challenges are numerous and include: orbital debris avoidance and removal, avoidance of interference with GEO satellite networks that enjoy protected status, coping with potential liability claims, avoidance of interference from terrestrial and high altitude platform system (HAPS) networks. In addition there is the business challenge of matching capital costs and operating expenses to revenue flows. The market demand and the technical, regulatory, and economic challenges remain to be clearly understood. Even so the current backers of the OneWeb system have an impressive array of technical competence and financial resources to address these challenges.

The Development of the O3b Satellite Network

The basic idea that the O3b network represents was proposed by Brazil's space agency (INPE) almost two decades ago. At this time, they proposed what was known as "the string of pearls" concept for an equatorial constellation of six to eight satellites in the equatorial band that would serve all nations near the equator. O3b began as an eight satellite equatorial constellation but has now been upgraded to a more intensive 12 satellites constellation. The O3b network can be further upgraded to an 18 satellite network in the future as demand might warrant. This upgradability based on market demand is one of the O3b constellation positive features – both from a business and a market-responsiveness perspective. As noted above, the Ka-band based Teledesic satellite constellation plus the design and operation of the Iridium and Globalstar constellations have also provided useful information with regard to the design and operation of the O3b satellite constellation as well as the OneWeb network.

One of the key design features of the O3b satellite is the many steerable antennas on each spacecraft. This allows the steerable antennas on-board each satellite to be continuously steered so that parabolic dish antennas on the ground can be continuously illuminated. This design feature is key to keeping the ground systems simple and lower in cost. The satellite's steerable beams can also be used to minimize interference to GEO satellites. To date, interference issues involving O3b satellite and GEO communications satellites have been avoided. The prime manufacturer of the satellites for O3b is Thales Alenia Space. The future concern, however, could be a failure in the steering mechanism for the antennas that could in time create a problem. In the future, the antenna beams might be electronically generated and steered by computer software, but again even electronically formed beams could malfunction (Fig. 1).

With the 12 satellite configuration, the O3b system actually covers a good deal of human populated Earth. The entire area from 45° North to 45° South can be effectively covered by this unconventional network. This means that the many billions of people that reside in this area including a very high percentage of those countries with developing economies are reachable via the O3b constellation. Many developed economies such as the USA, Japan, South Korea, and Australia are also within the coverage area as clearly shown in Fig. 2.

Fig. 1 The O3b satellite in systems test in the Thales Alenia production plant (Graphic courtesy of Thales Alenia)

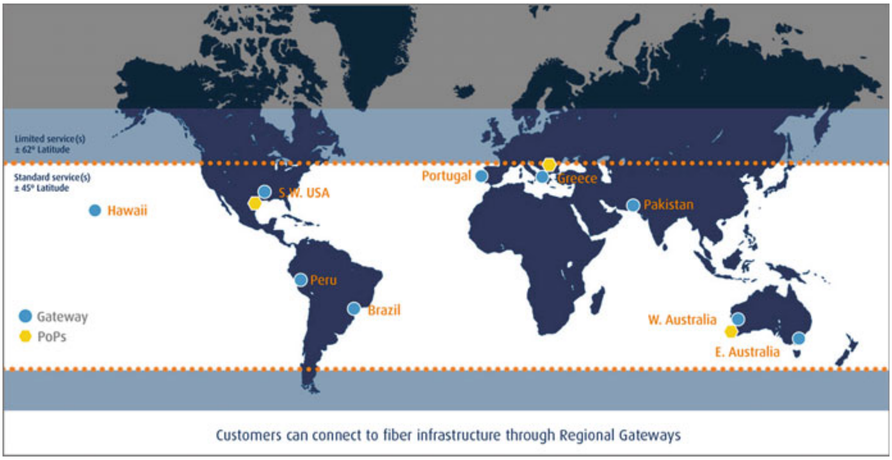


Fig. 2 The coverage of the O3b networks satellite constellation (Graphic courtesy of O3b)

The fact that all of the satellites are precisely maintained in an orbit that is 8,062 km (i.e., 5000 miles) in altitude means that the transmission path even to the extremes of the coverage area is on the order of only 12,000 km (7500 miles). This is still three times less than the minimum transmission path of a GEO satellite and perhaps five times less than the maximum transmission path for a GEO satellite connecting to high latitude regions (O3b Networks Frequently Asked Questions).

The key to the success of the O3b network is in many ways dependent on its ground segment. If the ground network is efficient in allowing a significant level of throughput, then the system can support a significant amount of traffic to be supported by the satellite network. If the ground system is composed of all small, lower traffic volume earth stations, then the total system throughput is significantly reduced. The key design element of O3b is that the space segment can be increased and network throughput enhanced as demand for service grows. This type of constellation design that allows network growth as traffic volume grows is one of its attractive features that distinguishes it from OneWeb that requires a very large network which 648 satellites deployed in 20 planes to be deployed to activate the system.

The basic technical characteristics of the O3b Ka-band network are provided as follow below: (The O3b Non Geostationary Orbit).

- Non-geosynchronous orbit (NGSO), fixed satellite service (FSS), Ka-band satellite network
- Initial constellation of eight medium-earth orbit (MEO) satellites increased to 12 MEO satellites
- Complete constellation will be at least 18–20 satellites
- Spacecraft provided by Thales Alenia Space
- Orbital height = 8,062 km; Equatorial inclination: 0°
- Ground period = 360 min/Number of contacts = 4 per day
- 30° spacing with 12 satellites
- Initial Ka-band frequencies (TT&C and Data Gateways)
- Downlink: 17.8 GHz – 18.6 GHz and 18.8 GHz – 19.3 GHz
- Uplink: 27.6 GHz – 28.4 GHz and 28.6 GHz – 29.1 GHz
- Global coverage
- Optimal coverage between 45° N/S latitudes
- Ten beams per region (seven regions) with 105 remote beams with 12
- Satellite constellation
- ~1 Gbps per beam (600 Mbps × 2); 126 Gbps available per 12 satellite constellation
- Beam coverage: Beam diameter to 600 Km
- Transponder bandwidth: 216 MHz; 2 × 216 MHz Fwd/Return Pair

The network is thus envisioned as providing high speed gateway access but also providing an air interface capability for wireless services. Today O3b is providing services to a wide mix of customers such as remote oil and mining operations, cellular operators in Samoa, etc. The potential of O3b to meet unmet needs in developing economies has been widely praised.

Dr. Hamadoun Touré, Secretary-General of the International Telecommunication Union (ITU), the UN agency for Information and Communications Technologies (ICTs) has said back in 2013:

> I am delighted to welcome an innovative newcomer to the ICT market, especially one whose strategy offers the potential to extend connectivity to broadband networks to millions more people worldwide. O3b's plan adds an exciting new piece to the puzzle through a low-cost solution that could help quickly bridge the emerging broadband divide separating rich and poor nations. The company's plan to have services available by 2013 means this solution could also play a significant role in harnessing ICTs to help meet the UN Millennium Development Goals by the target date of 2015. (O3b Network raises)

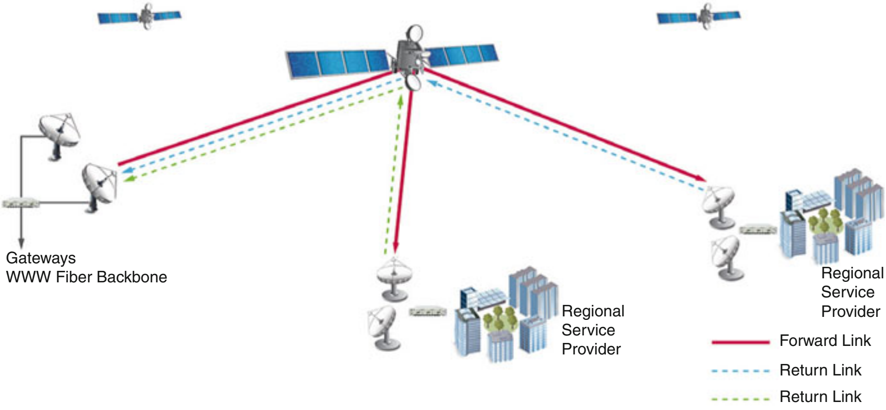


Fig. 3 Schematic showing MEOLINk ground antennas that can support data links up to 810 Mbps (Graphic courtesy of Qualcomm)

The Viasat-developed modems and encoding systems for the O3b system are highly efficient and the gateway stations that interconnect with fiber backbone are able to support data links at speeds up to 810 megabits/s. The network design is optimized for asymmetrical traffic so that thin-streams of traffic can be supported in the return link while very broadband services are supported for such throughput requirements as fiber interconnections (Fig. 3).

The OneWeb Network System

The OneWeb network has had evolved in its design in terms of its likely orbital configuration and number of satellites. At the current time, the network will involve some 648 satellites deployed in 20 different orbital planes some 18° apart and in a 1200 km (750 mile) orbit that will service the entire populated world. The design envisions a network that can provide very high speeds through gateways as well as air-interface standards that can support thin route services to villages and homes (Fig. 4) (OneWeb Taps Airbus To Build 900 Internet Smallsats 2015).

Qualcomm Research, the R&D division of Qualcomm Technologies, is designing many of the technology innovations required for the OneWeb network. The announced objective is to develop a new, high-performance wireless air interface for end-to-end satellite communications including system design of a new approach for wireless coding, modulation, and protocols. The specific objective is to allow OneWeb's architecture to provide layer 2 and layer 3 services that can be used by any ISP or telecommunication provider to extend any network using IP protocols. The plan is to develop and provide low-cost small cell terminals and a core network that is fully3GPP compatible with the 3rd Generation Partnership Project (3GPP) standards that seven standards organizations have joined together to develop. This in

Fig. 4 The mega-LEO constellation known as OneWeb

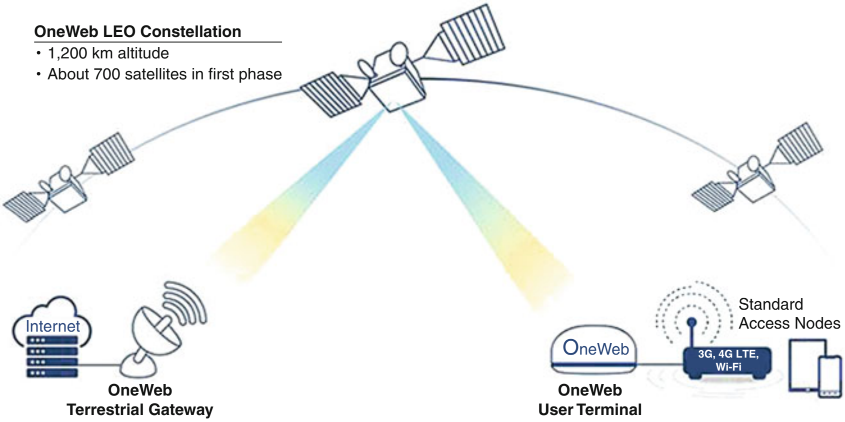


Fig. 5 Qualcomm engineering concepts for OneWeb broadband gateways and smaller user terminals (Graphics courtesy of Qualcomm)

theory will allow OneWeb to work together with providers in any regulatory environment, anywhere in the world.

To deliver reliable connectivity, the wireless air interface will enable intrasatellite, intersatellite, and intergateway handoffs. It will also be designed with advanced interference avoidance techniques to adhere to spectrum requirements. To help ensure the OneWeb system will be ready for commercialization, Qualcomm Research is also developing a modem hardware and software reference design for the OneWeb User Terminals – the terrestrial access nodes to enable connectivity to the satellite network (Connecting the Unconnected) (Fig. 5).

Table 1 Technical design aspects of the OneWeb satellites

Major design elements for the OneWeb satellites	
Orbital configuration	Initial configuration is 648 satellites deployed in 20 orbital planes at an altitude of 1200 km (750 miles). Note this is a change from earlier concepts of satellites at lower altitudes in two types of circular orbits at two different altitudes around 800 km
Satellite mass	Between 175 and 200 Kg
Satellite antenna characteristics	Phased array antenna measuring 36 by 16 cm (14.2 by 6.3 in)
Frequency band	Ku-band
Capacity of each small satellite	Theoretical throughput of 6 gigabits/s per satellite. Internet service speeds to ground antennas at 50 megabits/s
Approach to avoid interference to GEO sats	Patented "progressive pitch" system in which the satellites as they cross orbital arc are slightly turned to avoid interference with Ku-band satellites in geostationary orbit and then return to after crossing orbital arc
Construction schedule	The objective is to produce three satellites per day when into full production mode
Debris mitigation system	Details to be identified, but will meet minimum requirement of removal from Earth orbit within 25 years of end of life

The final design of the OneWeb Satellites are still in development and may during design reviews be modified to some extent. Nevertheless, the key technical specifications are largely now fairly clearly identified as shown in Table 1.

François Auque, Head of Space Systems at Airbus Defence and Space, at the first public announcement of the award stated that: "Teaming with OneWeb with a requirement to produce several small satellites each day has inspired us to develop innovative designs and processes that will dramatically lower the cost in large volumes for high performance space applications……Without doubt, this program is challenging but we're ready for it because we have leveraged resources and expertise across the entire Airbus Group." Airbus officials have claimed that they will adapt manufacturing techniques developed in the manufacture of the A380 aircraft to the rapid production of the nearly thousand satellites they will produce for OneWeb (France's President).

New LEO Constellations on the Horizon

To date the O3b satellite has been deployed in a 12 satellite MEO constellation and the OneWeb constellation is being manufactured at AirBus facilities and launch arrangements are in place. The Leosat constellation, in an initial constellation of 78 satellites, is likely to be the next LEO constellation to be deployed perhaps as early as 2019 or 2020. This network is different in that it would entail the launch of much larger and capable satellites with high capacity intersatellite links in order to support higher throughput requirements for large corporations and to support backhaul requirements for 4G LTE broadband cellular networks (LeoSat Constellation).

The Leosat constellation is also unique in its plan to use flat panel, metamaterial ground antennas that are able to track satellites as they transit in low earth orbit, rather than requiring satellite antennas that track ground stations mechanically. In the particular case of the Leosat constellation, there appears to be a close relationship between the Leosat constellation and the Kymeta Corporation that is one of the leaders in the new flat panel ground antennas that provide electronic (and computer controlled) beam formation and tracking. Kymeta is also working with Intelsat.

Kymeta's innovative antennas use metamaterials technology to electronically and dynamically adjust the antenna beam towards transiting satellites and does so with no moving parts. The technology enables flatter, smaller, and less expensive antennas compared to traditional parabolic dish satellite antenna technologies. Kymeta has already begun large-scale production of its antenna products for a variety of applications for terrestrial and space communications applications (CNBC Names Kymeta).

Finally, there is the least defined project of all. This is the SpaceX constellation that is being considered by Elon Musk that would possibly contain as many as 4000 quite small satellites that would, as in the case of OneWeb, be manufactured rapidly in a production line at a small cost well below $1 million per unit. While OneWeb and the SpaceX constellations would involve small satellites and optimized for Internet-related traffic for developing countries, the Leosat constellation is presumably envisioned for the top 3000 corporations in the world. The common element among the OneWeb, Leosat, and SpaceX constellations would be the difficulty of launching and managing such large satellite networks and avoid the problem of satellite collisions and to minimize the difficulty of problem of orbital debris build-up issues at the stage of launch, network management, or deorbit at end of life.

Issues Posed by New FSS Satellite Constellations

The advent of fixed satellite service (FSS) systems that are non-geosynchronous orbit (NGSO) systems creates a number of issues. By far, the greatest number of communication satellite networks are indeed concentrated in GSO, but one MEO network like O3b adds 12 satellites to Earth orbit and OneWeb mega-LEO will add, just in the initial configuration, something like 700 satellites. The disproportionate number of satellites added by mega-LEO triggers a number of key issues.

Frequency Allocation and Number of Large-Scale Constellations that Can Be Deployed

Satellites networks registered and coordinated through International Telecommunication Union (ITU) procedures have a protected status against other systems. There is also an agreed model of acceptable interference between GEO systems and NGSO that was agreed at the last ITU World Radio Conferences. Once a network has met these agreed interference criteria, then it too becomes protected. The problem is that each new system that is planned to be deployed that it has greater and greater difficulties of being successfully coordinated. Thus once the O3b systems and the OneWeb systems are coordinated, and if other systems such as the LeoSat NGSO and the Space X NGSO systems also go forward, it may be practically impossible for other systems such as those that might be envisioned by European, Chinese, or other countries to be deployed and achieved successful coordination as well.

Frequency Interference

There is concern that the ITU procedures agreed at the International Telecommunication Union (ITU) World Radio Conference (WRC) 23 January–17 February 2012 that established model levels of "acceptable" interference between LEO constellations and GEO satellite networks may in time be increased so that meeting ITU standards may become much more difficult. Even with the existing levels set for interference will make it increasingly difficult for additional systems that may be filed with the ITU and launched in the future. In short, there are two major problems that new LEO constellations represent with regard to GEO networks and even to additional MEO and/or LEO networks. These are the problems of frequency interference between various satellite systems, and, as more LEO constellations are added, the problem of actual physical collision.

Orbital Debris Concerns

At the current time, the largest practical concern is that of orbital debris increase. Since the 1960s there has been a steady build-up of orbital debris. On January 22, 2007, the Chinese missile destruction of the defunct Fen yun (YC-1C) weather satellite created an impulse jump of well over 2000 new trackable debris elements. The collision of the Iridium 33 and Cosmos 2251 satellites on February 22, 2009 created well over 2000 new debris elements. The NASA scientist Donald Kessler who warned of orbital debris and the possible growing cascade effect of debris that might ultimately grow out of control has warned that even with the amount of debris that is currently in orbit – without adding thousands of new satellites – will likely result in a significant new collision once every 10 years. The addition of the nearly 1000 satellites represented by OneWeb constellation, the 100 + satellites of the Leosat Constellation, the potential 4000 satellites of the SpaceX constellation present significant challenges to the future management of the ever growing space debris problem.

And this concern with regard to space debris does not include the Iridium current and generation NEXT constellation, the Globalstar network, plus the remote sensing networks of Skybox (Google), Northstar (Norstar Space Data) as well as US Defense mobile satellite network, LEO meteorological satellites, etc. There are currently some 22,000 + orbital objects more than 10 cm in diameter being actively tracked. When the new S-band radar "space fence" ultimately comes on line around 2018, it is estimated that over 250,000 objects will be trackable and only a small percentage will represent active satellites with active deorbit capabilities. If there is one problem to highlight with these new systems that is of greatest concern then orbital space debris is clearly the number one issue.

Liability Provisions

There is an associated concern that relates to all of the above issues and this is who pays for liability claims if there is a future situation where physical or financial damage is engendered as a result of a satellite or satellite constellation creates a low to others either in space or on the ground. Although increasingly it is private companies that own and operate satellite networks or entire constellations, it is not they that are liable. The liability is that of space insurance companies or ultimately of the "launching state" as explicitly identified in the Outer Space Treaty and the so-called Liability Convention. If SpaceX, a US company, for instance, launches a LEO constellation of 4000 satellites and this deployment somehow triggers a run-away cascade of space debris with many space objects crashing into one another and creating a deadly shield of space debris encircling Earth and traveling at over 10,000 km/h, the USA could presumably be liable for trillions of dollars in damages. Vital networks such as for weather forecasting, communications, remote sensing, and navigation, positioning, and time, etc., could ultimately be lost since replacement satellites could not be safely launched.

For many years, the liability claims related to space have been minimal and issues have largely involved concerns related to the use of isotope fuels. Today it appears that a whole new era of concerns have arrived.

Conclusion

The concept of deploying constellations of satellites in low earth orbit or medium earth orbit is not a new idea. Even Arthur C. Clarke anticipated the use of low earth orbit satellite systems. The advantages that such LEO or MEO constellations can bring include low transmission latency (which is particularly useful for Internet-related services) and much less path loss for the RF signals transmitted to and from the satellite. New spacecraft and ground antenna technology and improved processing and coding techniques today make the commercial and operational feasibility of such systems much higher. There are still a number of challenges for these new systems and concerns about such issues as orbital debris. The success of the O3b system is a hopeful sign that broadband services can indeed be efficiently provided from non-geostationary satellite systems (NSGOs). The experience achieved with the various new systems discussed in this chapter will provide a much clearer pathway to the future of communications satellite services and the types of spacecraft that will be deployed in the decades ahead. These new systems will also greatly affect the design of ground systems as well.

Cross-References

- ▶ Broadband High-Throughput Satellites
- ▶ Satellite Orbits for Communications Satellites
- ▶ Trends and Future of Satellite Communications

References

“CNBC Names Kymeta Corporation Among World’s Top 50 Disruptors” Business Wire, June 14, 2014 http://www.businesswire.com/news/home/20140617005402/en/CNBC-Names-Kymeta-Corporation-World%E2%80%99s-Top-50. Last accessed 11 Dec 2014

Connecting the Unconnected, Aug. 26, 2015 https://www.qualcomm.com/news/onq/2015/08/26/connecting-unconnected-qualcomm-and-oneweb-developing-global-network-extend. Last accessed 9 Dec 2015

P. de Selding, “One Web’s Powerful Partners in their Own Words” Space News, (June 26, 2015) http://spacenews.com/onewebs-partners-in-their-own-words/Last. Accessed 8 Mar 2016

France’s President + CEOs Celebrate As OneWeb Selects Airbus Defense + Space To Connect The World w/Microsatellites, Space News June 15, 2015 http://www.satnews.com/story.php?number=1735784868#. Last accessed 4 Dec 2015

Leosat constellation, http://www.leosat.com/Last. Accessed 11 Dec 2015

O3b networks Frequently Asked Questions, http://www.o3bnetworks.com/faqs/Last. Accessed 9 Dec 2015

O3b Networks raises total funding of US$1.2 billion http://www.o3bnetworks.com/o3b-networks-raises-total-funding-us1-2-billion/Last. Accessed 7 Dec 2015

OneWeb Taps Airbus To Build 900 Internet Smallsats”. SpaceNews. 2015-06-15. http://spacenews.com/airbus-wins-oneweb-contract/Last. Accessed 12 Dec 2015

J.N. Pelton, *Satellite Communications* (Springer, New York, 2013)

J.N. Pelton, *The Oracle of Colombo: How Arthur C. Clarke Revealed the Future* (Springer, New York, 2015)

The O3b Non-Geostationary Orbit Fixed Satellite Service System: http://www.sspi.com.br/portal/images/stories/pdfs/spectrumday2010/spectrum-day-2010-o3b.pdf

An Examination of the Governmental Use of Military and Commercial Satellite Communications

Andrew Stanniland and Denis Curtin

Contents

Abstract

This chapter provides information concerning the use by governments of military and commercial satellites systems for strategic and defense purposes. It discusses dedicated communications satellite systems designed for particular uses and the so-called dual use of commercial systems to support military and strategic

---

A. Stanniland (✉)
Inmarsat, London, UK
e-mail: Andrew.Stanniland@inmarsat.com

D. Curtin
Communications Satellite Consultants, Rockville, MD, USA
e-mail: dns.crtn@gmail.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_8

purposes. It explains various pathways that can be followed by governments to obtain communications satellite services to support military uses. These paths include: (1) dedicated satellites, (2) hybrid satellites (both military and commercial payloads on a single satellite), (3) shared satellite facilities via intergovernmental agreements, (4) guaranteed long-term leases, (5) ad hoc leases of capacity on demand, and (6) a long-term partnership between a government and a commercial partner as is the case with the Skynet 5 program in the United Kingdom.

In this chapter the authors will also examine how various countries obtain their national satcom, how and why commercial capacity has become, and will continue to be, a significant part of national satcom capabilities. It will examine the present and future contracting approaches and procedures used in various countries but primarily in the United States and other NATO countries.

Finally, there will be a discussion of the issues involved when nations decide between purchasing nationally owned satellites and leasing capacity commercially. In this regard, it is noteworthy that in many cases the major investments in new technology for satellite defense communications systems are now more often coming in the commercial communications world. Governments are more and more changing their procurement models to take advantage of commercial procurements or long-term leases. This allows military communications units to spend their financial resources more strategically on any small adjustments to make their satellite acquisitions more military specific. Technology is typically moving too fast for a “normal” 5-year military R&D program followed by procurement cycles to be at the cutting edge of the latest technologies in today’s world.

Keywords

Ad hoc leases • Advanced Extremely High Frequency (AEHF) satellite • Anik satellite of Canada • Athena-Fidus joint French and Italian satellite • CBERS satellites of China • COMSATBw of Germany • Dedicated satellites • Defense Information Systems Network Satellite Transmission Services-Global (DSTS-G) contract • Defense Satellite Communications System (DSCS) • Dual use of commercial satellites • End-to-end services • European Defence Agency (EDA) Satellite Communications Procurement Cell (ESCPC) Global Broadcast Service (GBS) • Haiyang satellite of China • Hispasat • Hosted Payloads • Hybrid satellites • Joint Tactical Radio System (JTRS) • LEASAT • Marisat • Mexsat-Milstar • Ministries of Defense (MODs) • Mobile User Objective System (MUOS) • NATO • NATO NSP2K • Paradigm • SICRAL satellite of Italy • Skynet – Tranformational Satellite (TSAT) System • Transformational Communications Architecture • TURKSAT • Turn-Key Services • UHF Follow on (UFO) • Unmanned aerial vehicles (UAVs) • Wideband Global System • X-band • XTAR • Yahsat • Yaogan of China • Zhongxing satellite of China

Introduction

Communications have been critical to military organizations for centuries. With the advent of the use of space and the increased sophistication of military forces, it was natural that the military would look to space to help supply their communications needs. After the launch of Sputnik in 1957, it was clear that satellites offered opportunities for all types of applications including communications. Almost immediately military organizations around the world began to look at ways to use satellites to provide communications to their forces. Over the last 50 years, a variety of satellites, deployed into different orbits, have been developed by many nations to help meet their government's communications needs.

Governments tended to focus on the geostationary orbit because of its specific beneficial characteristics for supporting communications. A single satellite deployed into geostationary orbit provides visibility of almost a third of the earth's surface. In addition, the most distinct characteristic of the geostationary orbit is that satellites placed there have the same orbital period as the earth and appear to be fixed above the same point on Earth, and therefore satellite terminals on the ground do not have to track the satellite's movement. This means that the terminal will be lower in cost than a more sophisticated terminal with the ability to track a moving satellite and can also be much easier to use. Ease of use and minimizing training requirements are key factors that are considered when implementing infrastructure for military forces.

The United States, Soviet Union, and some NATO countries initially set about procuring their own, unique, national satellite systems in the 1960s when the field of satellite communications was just being pioneered. These same nations also played a key role in the start-up of the commercial satcom industry. The United States was, and still is, the leading pioneer in the military satellite communications (satcom) arena. The United States, alone now has multiple constellations of satellites in geostationary orbit, with the total number estimated at 25 spacecraft.

A number of different contract approaches have been used to obtain defense-related communications satellite capacity, and a whole industry has been developed around providing communications requirements for military purposes. In some cases, industry has provided the military user with full end-to-end (sometimes known as turn-key) services including the capacity leased from a commercial provider, terminals, other hardware, and operation of the full service to supplement the military capacity.

This chapter will discuss a number of the satellite solutions that have been developed and deployed by different countries and how some of these countries have procured their military communications. It will also discuss the continuing emergence of new and diverse communications requirements, the issues raised by these new demands, and the approaches utilized to satisfy them. It will also look at what the future might bring for both the military and the commercial industry.

A good summary of the history of both commercial and defense-related satellites is discussed in an earlier chapter of this book entitled "▶ History of Satellite

Communications." The authors refer the reader to that chapter rather than repeating the history here.

From the mid 1960s until the mid 1990s, commercial satellites and military satellites were on different development paths. Military systems were initially designed to meet the demands of the Cold War. Starting in the 1980s, there were only a few dedicated satellite services designed to support tactical level forces via a small number of satellite ground terminals. Such systems began to be fielded in larger numbers, predominantly by the United States, to tactical level forces during the mid- to late 1980s.

The generation of military communications satellites available for use during the 1990s and 2000s were, for the most part, defined, designed, developed, and produced beginning in the late 1980s. This planning was largely carried out before the widespread advent of cell phones, the ubiquity of personal computers, the evolution of the Internet to a public global utility, and the subsequent innovation of what became known as the World Wide Web. Technology innovations developed specifically for commercial systems could be, and often were, adopted by military systems and vice versa.

This result was unsurprising since the firms that built the satellites built them for both military and commercial customers. Commercial systems being developed during the late 1980s were evolving faster than their military counterparts to meet the needs of international telephony and the broadcast industry, especially with the advent of using communications satellites to broadcast television channels direct to people's homes.

During the Cold War period, the vast majority of the United States and NATO defense-related communications traffic was carried by a country's own national satellite(s) with some minor military traffic being carried by commercial satellites. This began to change in the 1990s as more international conflicts started to occur simultaneously in different geographical locations, for example, the Bosnia, Kosovo, and Gulf War I conflicts between 1990 and 1999. This evolution in the operational context meant that the existing military satcom systems were becoming outdated and struggling to meet the new demands as flexible coverage and increasing throughput to support a wider range of applications became the order of the day.

By the beginning of the second Iraq war in 2003, the bandwidth requirements had increased to the point where the available capacity of the defense satellites was no longer able to meet the demand. This resulted in the US DoD and other NATO Ministries of Defense (MODs) turning to commercial satellite capacity to meet the shortfall. This was a fundamental and far reaching change, which has continued and indeed greatly increased since 2003.

Due in no small part to budget constraints and delays in satellite procurement programs, as well as the greater demand for capacity, the amount of on-orbit defense satellite capacity in the United States and elsewhere has not been able to meet the demand. In 2004, the US DoD began acknowledging that commercial satellite capacity was providing over 80 % of US satcom bandwidth that the US military used. This continued to evolve in the 2006–2007 time period, and during this period, over 95 % of the Satcom used in the US Central Command (CENTCOM) area of responsibility was provided by commercial systems (DISA conference proceedings 2009).

This trend has been advanced by the growing sophistication of the communications devices employed by the military, including communications on the move terminals, man-packs, and the increasing use of unmanned aerial vehicles (UAVs) with sophisticated sensors and cameras necessitating a large amount of capacity to transmit the data to both tactical and garrison facilities where it can be stored, processed, and analyzed. It is predicted that these trends will continue and that even with the launch of additional DoD and other countries' government satellites, the vast majority of the defense-related bandwidth requirements will be met by commercial satellites for the foreseeable future.[1]

The US DoD employs commercial capacity more frequently and in greater volume than any other country or alliance, including NATO. The implication of this fact for the US DoD is that commercially procured satellite communications can no longer be considered just an adjunct to military communications but must be regarded as an integral part of the warfighters' communications inventory and must therefore be treated as a critical part of a nation's infrastructure. The Transformational Communications Architecture (National communications system fiscal year 2007) and the Joint Space Communications Layer (JSCL) Initial Capabilities Document (Satellite 2001) both state that commercial satcom is now an integral part of DoD's overall satcom capability portfolio.

This fundamental shift means that many commercial satellites currently in orbit, as well as those being developed, will often be dual-use satellites with a significant portion of their capacity employed for noncommercial and often very sensitive communications. In some cases (for example, the XTAR X-band satellites), the satellites are developed specifically for the provision of government services but under commercial terms. In another example, a portion of the UK military Skynet 5 satellites, owned by Paradigm Secure Communications Limited, has been specifically designated under the UK MOD agreement with Paradigm to providing services to both the US DoD and other MODs under commercial terms.

The widespread use of commercial capacity by the DoD and others has led to some difficult questions for defense users and policy makers as well as for commercial owner-operators and commercial service providers. Likewise, it also poses significant issues for defense contractors that manufacture and develop unique technologies for military satcom satellites. Such defense contractors have made the development of dedicated military satellites one of the cornerstones of their order book for over 40 years. Questions that now arise include the following: Can defense budgeting continue to fund extensive capital procurement programs? Can capital expenditure budgets be refocused to fund a leased solution? What type of traffic can be sent over commercial capacity for the longer term? Are commercial systems reliable enough for sending sensitive traffic? Is commercial encryption sufficient?

---

[1] Defense systems article: commercial satellites plug bandwidth gap for military satcom, http://defensesystems.com/Articles/2011/02/28/Cover-Story-Commercial-Satellites-Evolve.aspx?Page=1

Can a commercial operator be trusted to guarantee communications at all times and in all places? Is the lack of nuclear, and other physical, hardening measures an issue given the high percentage of defense-related traffic on commercial systems? Where are the teleport facilities of the commercial operator? Will any military traffic be "landed" there? What control will the military operators have over the routing of their traffic, if any? Will supporting military operations have any knock-on effects for the day to day business of a commercial operator? Can a commercial operator separate out military and commercial traffic within its operations and its business processes? Can industry build military payloads that fit in with commercial operator business plans?

Solutions are already being found for many of the above questions. Clearly the nationally owned defense satellites will continue to be used for the most critical traffic by most nations, but, with the growing demand from users, the nationally owned resources can no longer be relied on to have sufficient capacity and, therefore, it is inevitable that some sensitive traffic will have to be transmitted commercially. It is expected that as new commercial satellites are developed, the United States and other governments may ask that those satellites which could be used partially for government purposes be fabricated with some additional attributes such as specific frequency bands, a limited jamming resistance, or steerable antenna.

Various approaches have been used over the years to acquire the necessary capacity. In some countries with relatively small requirements, such as Spain with the original HISPASAT satellites or Turkey with the TURKSAT satellites, the military has already added X-band or other military frequency transponders to commercial national satellites. Other countries like the United Kingdom and France have launched national multimission defense satellites having a combination of UHF, X-band, and EHF transponders and are following this up with sophisticated outsourcing programs. In the United States, the government programs have largely employed satellites dedicated to a single mission and frequency band although, beginning in the 1970s and 1980s, the US government also leased capacity on satellites such as the MARISAT and LEASAT satellites where commercial operators had added specific capability to commercial satellites for government use. This then paved the way for the future use of commercial satcom assets for military requirements.

In addition to leasing commercial capacity from a commercial operator, another approach to providing this needed capacity is through the use of Hosted Payloads where a payload for a specific government purpose is designed, built, and installed on a commercial satellite planned or under construction. The hosting of payloads is not a new concept: the US government has Hosted Payloads on other government satellites for decades, as stated above with the LEASAT program dating from the late 1970s, but increasingly hosting of military payloads has found its way back into the military satcom policy makers' thinking over the last decade. More recently, there have been a number of meetings between industry and the US DoD discussing the technical issues surrounding Hosted Payloads, as well as different ways to overcome any contractual, coordination, or other issues that might delay the implementation of Hosted Payloads.

The commercial satellite industry is constantly evaluating what capacity will be needed in the future, with what frequencies and with what spot or zonal beams that could be available at different orbital locations. The industry's planning would be more accurate if it knew when and where defense-related capacity would be required and for what period of time the requirements would endure. This is a difficult question for any government to answer as operational requirements are not only often unpredictable but are also likely to be classified. How does a defense ministry contract for capacity servicing a specific geographical location for several years if there is a likelihood that the military situation at that location might change and the capacity would no longer be needed, or might be needed elsewhere? How does a commercial satellite operator manage the transient nature of the military requirements with the fact that the commercial mission is more than likely to be focused on a specific population area, which prevents the satellite being moved to a different orbital location? In the United States, the commercial satellite industry and US government officials regularly meet to discuss the best way to frame and work through these types of issues.

Nationally Critical Satellite Communications

The justification behind procuring highly survivable, dedicated military systems, which often represent much higher cost facilities, versus the leasing of commercial satellite capacity has become a highly contentious issue for procurement agencies worldwide. More nations are looking to implement dedicated military satcom systems than ever before. While some are satisfied with an initial reduced capability system in order to obtain a limited military satcom capability before upgrading to more capable assets in subsequent phases of infrastructure rollout, many nations are choosing to implement state-of-the-art resources at the first attempt in the understanding that their requirements will expand to fill the available capability (e.g., UAE and Norway).

At the beginning of 2011, the main players in the field of designing, building, and procuring dedicated military satellites, or dedicated military payloads that are owned and operated by the military, are China, France, Germany, Italy, Russia, Spain, the United States, and the United Kingdom (although as will be explained later, the UK capability has now been outsourced as part of the Skynet 5 program).

Several other countries have developed a dedicated military satcom capability, typically by adding a military frequency payload to a nationally owned and operated commercial satellite. Among other countries, Australia, Brazil, Japan, South Korea, and Turkey have done this. Usually one or more small military satcom payloads on national satellites are sufficient to meet the national defense needs of the particular country rather than paying for dedicated, and sophisticated, military satcom satellites as are required by other countries with greater and more demanding national interests, associated military responsibilities, and resulting communications requirements. This approach has been used successfully for the last 20 years and it is expected that additional countries who believe that they have a need for a specific, but limited, military satcom capability will either add military frequency capacity to a national commercial satellite or possibly arrange for a Hosted Payload on an international commercial satellite under development (for example, Australia's successful negotiation to have a commercially built UHF Hosted Payload included as part of the Intelsat 22 satellite).

Arguably the most critical attribute of military satcom systems is the confidence that the military communicator has to have that their communications equipment will work – always and on immediate demand. These systems are used by nations to support both peacekeeping and hostile deployments and, therefore, there is always a threat of hostile (or accidental) interference to the radio signals and a physical threat to either the ground or the space-based equipment, coupled with the political threat that the nation providing any commercial communications may not agree with the military activity being pursued and seek to deny access to the communications.

Nations with military satcom capability have analyzed their individual security and survivability requirements and typically followed one of five paths to secure access to guaranteed communications for their militaries:

- Dedicated satellites: These are satellite that are specifically designed, procured, and launched by the government itself. The whole satellite is solely for military or governmental purposes and both the payload and the platform can be designed to satisfy the demands for security and survivability.
- Hybrid satellites: This type of satellite has a payload that is designed and procured by the government but launched as a co-payload on a commercial satellite. The payload can satisfy the security needs, but the platform is normally built to commercial standards to keep cost to a minimum. Marisat, LEASAT, and Telecom are early examples. These are the forerunners of today's Hosted Payloads.
- Intergovernmental agreements: Under this type of arrangement, countries who are natural allies enter into agreements to provide each other with dedicated (or backup) communications capability as an alternative to the procurement of stand-alone capacity or infrastructure.
- Guaranteed, long-term leases: This type of lease provides for assured access to communications that are fully or partially guaranteed by agreeing to a long-term reservation or usage contracts with commercial operators. Usually the protection of the communications capacity itself cannot be guaranteed and so this approach is often favored by those nations with either a low threat assessment or who can rely on allied military satcom for requirements with a higher threat assessment.
- Ad hoc leases of capacity: This type of arrangement provides excellent value for money since a nation only pays for what it uses, but there is no guarantee that the capacity will be there, or what it will cost, unless it uses a national commercial provider as its conduit. There will also be no guarantee of any information assurance features on the capacity procured.

Since the advent of the Skynet 5 program in the United Kingdom, a sixth option has emerged – a formal, contractual partnership between government and industry to provide a service-based approach for both commercial and military communications. In this case, the total need is satisfied by a commercial company, but this is accomplished via a mix of guaranteed access to protected military communications capacity that is owned and operated by the commercial service provider and long-term commercial leases managed and secured by the service provider.

The lead taken by the United Kingdom in this arena is indicative of the trend in Europe where decreasing defense budgets are forcing governments to explore financial and commercial innovation with just as much rigor as technical innovation. For example, the military satcom systems of France, Germany, Norway, Italy, Spain, and the United Kingdom have all been procured using different commercial and financial methodologies by the respective national defense departments. The next generation (beyond 2020) European systems are expected to further explore methods of increasing defense budget utility with increasing international consolidation, but today each nation is ensuring that its diverse geographical and interoperability requirements can be met by a nationally procured solution using a multifrequency payload. In Japan, the government is currently exploring how it can best provide future military communications capacity as well.

The Dedicated Satellites Approach

This section examines some key examples of dedicated defense-related satellite systems that have been implemented.8pt?>

NATO

NATO has been a user of military satcom since 1970 and has owned and operated the NATO 1, 2, 3, and 4 series of satellites. The NATO 4 (sometimes designated NATO IV) satellites, which were launched in 1991 and 1993, were built to the same design as the UK Skynet 4 series. NATO changed its approach to the procurement of military satcom in 2004, as will be discussed later.

United States

The United States currently possesses the largest number of military satcom satellites, and it has developed into a nation with multiple constellations of satellites. These constellations tend to be frequency specific. This often results in an additional, alternate payload of lower capability being used to provide cross compatibility with the frequencies employed on other constellations. The United States has divided its communications into four elements:

- Narrowband, unprotected communications using UHF
- Wideband communications with limited protection features on X-band and, more recently, Ka-band frequencies
- Protected communications with full hardening and survivability features using EHF frequencies

- Leased commercial satellite communications using L-band, C-band, Ku-band, and more recently X-band and UHF (though these commercial capabilities are not the subject of this part of the chapter)

Despite being an extremely limited service in terms of throughput, UHF has become an enduring technology for troops worldwide due to its utility for highly mobile, deployed forces. This utility is unlikely to change in the future despite the advent of handheld commercial satcom systems such as Iridium, Thuraya, and Globalstar.

The Mobile User Objective System (MUOS) satellites integrate commercial cellular technology, Wideband Code Division Multiple Access (WCDMA) waveform, and Universal Mobile Telecommunications System (UMTS) infrastructure and has now replaced the UHF Follow On (UFO) satellite network.[2]

The MUOS constellation thus now provides UHF secure voice, data, video, and network-centric communications in real time to US mobile warfighters through 2030 and will be fully interoperable with the Joint Tactical Radio System (JTRS) and current radio systems. The system was designed to maintain compatibility with UFO system and legacy terminals by the inclusion of a UFO legacy payload on the earliest MUOS satellites. However, the UFO is now retired from service and MUOS provides mobile military satellite services for the US Department of Defense. Individual terminals (users) will be able to access up to a 64 kbps link.[3]

The majority of the US military communications are supported by the Wideband Global System (WGS) that has now replaced the Defense Satellite Communications System (DSCS).[4,5] In contrast to the now retired UFO and MUOS, both of these constellations are operated by the US Air Force. The first DSCS III was launched in 1982 and the first DSCS IV launched in the 1990s, but all of these are now retired. The first WGS was sent up in 2007 and these continue to be launched to support all US military forces and some US allies.

The replacement to DSCS, WGS, is a satellite communications system which was originally conceived as an interim system to meet the military needs of the first decade of the twenty-first century and that provides flexible, high-capacity communications for US warfighters. In fact the original name was Wideband Gapfiller Satellite. The Wideband Global System (WGS) provides a quantum leap in communications bandwidth over DSCS. Although one key difference is that while the DSCS satellites included technical features to make the success of denial of service threats by enemies more difficult, the DoD decided to design the WGS satellites

---

[2] US Navy homepage for UFO system, http://www.public.navy.mil/spawar/PEOSpaceSystems/ProductsServices/Pages/UHFGraphics.aspx

[3] Description of MUOS system, http://www.globalsecurity.org/space/systems/muos.htm

[4] US Air Force fact sheet for DSCS system, http://www.losangeles.af.mil/library/factsheets/factsheet_print.asp?fsID=5322&page=1

[5] US Air Force fact sheet for WGS system, http://www.afspc.af.mil/library/factsheets/factsheet_print.asp?fsID=5582&page=1

without such features, relying instead on the Milstar and future AEHF systems to support the highly survivable communications requirements. The WGS satellites are therefore nearly identical to commercial satellites from the early 2000s in terms of their ability to respond to denial of service threats.

The WGS satellite system, originally designed as a constellation of three satellites, has now been expanded to at least six satellites and provides service in both the X-band and military Ka-band frequency spectrums. The decision to obtain more WGS satellites came when the ambitious TSAT (Transformational Satellite System) was canceled due to cost overruns.

As well as replacing the DSCS X-band communications and the Global Broadcast Service (GBS)[6] one-way Ka-band service, WGS provides a dedicated two-way Ka-band service for US DoD users for the first time. The first WGS satellite entered service in 2007 with WGS 2 and 3 following in 2009. The second batch of WGS satellites has a modified Ka-band payload configuration specifically intended to improve throughput for unmanned aerial vehicles.

WGS supports US DoD high data rate intelligence, surveillance, and reconnaissance (ISR) requirements as well as tactical warfighting units, many of the ISR requirements have hitherto been supported using commercially available capacity due to the shortfalls in high data rate capable capacity.[7,8]

Finally, the highly survivable and protected national communications requirements are supported on the Milstar[9] and Advanced Extremely High Frequency (AEHF)[10] systems.

Milstar represents a joint service satellite communications system comprising five satellites launched between 1994 and 2003 (an additional satellite was lost on launch). The system provides secure, jam resistant, worldwide communications to meet essential wartime requirements for high-priority military users. An important difference between Milstar and AEHF and other military communications satellites is that each satellite processes the communications signal within the payload, restoring the signal to its original form, and serving as a smart “switchboard in space” which can direct traffic from terminal to terminal anywhere on the Earth. Inter-satellite links further reduce the requirement for ground controlled switching. Milstar can support individual user link data rates from 75 bps through 1.5 Mbps. The AEHF, however, will replace Milstar service at the end of their service life.

AEHF provides global, secure, protected, and anti-jam communications for high-priority military ground, sea, and air platforms. The system will consist of at least

---

[6] LA AFB fact sheet for the GBS system, http://www.losangeles.af.mil/library/factsheets/factsheet.asp?id=7853

[7] Intelsat general UAV services, http://www.intelsatgeneral.com/services/applications/uav.aspx

[8] Satellite markets and research: government/military demand for commercial satcom remains steady, http://www.satellitemarkets.com/node/769

[9] U.S. Air Force fact sheet MILSTAR satellite communications system, http://www.af.mil/information/factsheets/factsheet_print.asp?fsID=118&page=1

[10] U.S. Air Force fact sheet advanced EHF system, http://www.afspc.af.mil/library/factsheets/factsheet_print.asp?fsID=7758&page=1

Fig. 1 Advanced EHF satellite (Image taken from US Air Force website; US Air Force fact sheet advanced EHF system, http://www.afspc.af.mil/library/factsheets/factsheet_print.asp?fsID=7758&page=1)

four satellites in geosynchronous earth orbit (GEO) and will provide 10–100 times the capacity of the Milstar satellites which they will eventually replace, with maximum data rates on individual user links up to 8 Mbps instead of the 1.5 Mbps possible from Milstar. Without question, the AEHF system is the most complex satcom satellite now in service for assuring communications to US military forces and is, in terms of survivability and security capabilities, the most advanced military communications satellite in the world (Fig. 1).

The first AEHF satellite was launched in August 2010. However, the satellite failed to initially achieve geosynchronous orbit due to a malfunction in the liquid apogee engine and was left in a low earth orbit. Utilizing the other thrusters on board the satellite, AEHF Flight 1 (AEHF F1) was subject to a long duration orbit raising exercise to raise this satellite to its correct geostationary orbital slot at 68° West Longitude. It was found that the liquid apogee engine had a malfunction rather than a design flaw; consequently, preparations and plans continue toward the launch of AEHF F2.

The AEHF system will be used not only by the US DoD for its highly critical communications links but also by a multinational consortium of allies who have all invested in the satellites and the associated ground systems. These nations (Canada, Netherlands, and United Kingdom) will be granted access to a specific amount of capacity across the AEHF constellation and will purchase appropriate terminal and teleport equipment to be able to access the system via government to government agreements.

The above summarizes the US DoD's operational military satcom systems; the future of US military satcom beyond MUOS, WGS, and AEHF was intended to be satisfied by a program called Transformational Satellite Communications System (TSAT). This program was a US DoD program to provide high data rate military satcom and Internet-like services. TSAT was planned as a five satellite constellation with a sixth satellite as an in-orbit spare, with the first launch in the 2019 time frame.

An extremely ambitious project, utilizing many state-of-the-art space-borne technologies, TSAT was intended to ultimately replace the DoD's current satellite system and supplement the constellation of AEHF satellites. It was designed to support net-centric warfare and would have enabled high data rate connections to space and airborne intelligence, surveillance, and reconnaissance (SISR, AISR) platforms. The total RF throughput projected for the TSAT program was more than ten times that of the AEHF system.

In April 2009, after almost $2 billion (US) of R&D expenditure and 6 years of development, Secretary of Defense Robert M. Gates asked that the project be canceled in its entirety.[11] High cost, technological risk, and development delays were given as primary reasons.

As an interim replacement strategy, Secretary Gates recommended the procurement of two additional AEHF satellites, bringing the total constellation to four satellites. Although some industry analysts would say that the cancellation of the TSAT program was inevitable in the current US defense budget climate, it is clear that the decision is already having an effect on the future of dedicated military satcom programs around the world. The appetite for governments to fund the design and development of quantum leaps in technology and capability is decreasing, which in turn is forcing industry and the military alike to examine the potential for incremental capability increases along with more innovative use of existing technologies. It is possible that a more advanced design for a dedicated US defense satcom program may be restarted in future years, although this does not seem to be a near-term prospect.

United Kingdom

The United Kingdom has been a military satcom user since the late 1960s, and the Skynet 2 satellites were actually the first communication satellites to be built outside either the United States or the USSR. Because of the breadth of geographical coverage needed by the UK armed forces, the United Kingdom has always opted for a multisatellite constellation – despite its relatively small size. As stated earlier, the Skynet 4 design was reused by Matra Marconi Space Ltd as the basis of the NATO IV series of satellites, thus guaranteeing interoperability between the United Kingdom and NATO satcom equipment. Skynet 4 was the last series of UK satellites to be wholly owned and operated by the UK MOD and is now retired. The advent of the Skynet 5 private finance initiative (PFI) program has transitioned this responsibility into industry and will be discussed in more detail later in this chapter.

---

[11] Defense budget recommendation statement made by secretary of defense Robert M. Gates, 06 April 2009, http://www.globalsecurity.org/military/library/news/2009/04/dod-speech-090406.htm

France

France has been a member of the military satcom community since 1980 with Syracuse 1 and Syracuse 2 (now retired) and currently operates the Syracuse 3 constellation.[12,13] France has also followed the hybrid satellite route, sharing satellites with the "Telecom" commercial payloads, which were owned and operated by France Telecom. Syracuse 3A was launched in October 2005 and Syracuse 3B followed in August 2006. The Syracuse 3 series is hardened and protected to NATO standards, similarly to the UK Skynet series, and unusually for European military satcom does not have a narrowband UHF payload on board, concentrating instead on military X-band and EHF frequencies to support the French military.

In 2007, a third spacecraft was expected to be ordered, but this was canceled in favor of including the Syracuse-3C payload on the Italian SICRAL 2 satellite, ushering in a new era of allied collaboration which will no doubt have far-reaching impacts on the whole of the military satcom arena.

Germany

Germany has only recently entered the military satcom arena with its own dedicated assets, relying for many years on NATO capacity, intergovernmental agreements, and commercial leases. In 2006, the German Bundeswehr awarded a contract for the construction of two satellites for narrowband and wideband communications, a comprehensive ground user terminal segment and the upgrade of the network management center to a special company set up to deliver the capability, MilSat Services GmbH, which was a joint venture between EADS SPACE Services and ND SatCom.[14]

COMSATBw 1 and 2 were launched in October 2009 and June 2010, respectively, and are now fully in operation and owned by the Bundeswehr. However, their operation is carried out by MilSat Services GmbH, who is also responsible for the ongoing maintenance of the ground network as well as the long-term leases of any required commercial satellite capacity.

Italy

Italy entered the military satcom arena in 2001 with the launch of Sicral 1A[15] into geostationary orbit, providing UHF and X-band capacity to Italian armed forces. Like France and the United Kingdom, Italy has been a mainstay of the delivery of X-band and UHF capability to NATO forces since the signing of the NSP2K memorandum of understanding in 2004 between the Ministries of Defense of Italy,

---

[12] Alcatel press announcement on Syracuse 3B, http://www.home.alcatel.com/vpr/vpr.nsf/DateKey/16012004uk

[13] Description of Syracuse 3 system, http://www.deagel.com/C3ISTAR-Satellites/Syracuse-III_a000283001.aspx

[14] Satcom BW overview, http://www.astrium.eads.net/en/programme/satcombw-comsatbw2.html

[15] Sicral program overview, http://www.telespazio.it/pdf/Tes53_impg_3_4_09_ing_lowresolution.pdf

France, the United Kingdom, and NATO. This agreement meant that the then planned Sicral 1B satellite (which was duly launched in April 2009) was now even more critical for Italy.

As mentioned earlier, Italy and France have now jointly embarked on the SICRAL 2[16] program, which is expected to enter service in 2013.

USSR (and Now Russia)

The USSR was the first country to orbit a satellite in 1957. The Soviet Union, and now Russia, has been very active in using their space-borne capability. It is reported that between 1960 and 1990 the vast majority of Soviet satellites that were launched carried military payloads, even though until the last decade of the twentieth century there was no official acknowledgment of a military space program. During the first decade of the twenty-first century, Russia has continued its launch program and now identifies specific military satellites but with no specific information as to individual missions.[17]

China

China launched its first satellite in 1970. Since then its satellite activity has increased, particularly in the last decade of the twentieth century and the first decade of the twenty-first century (Annual report to congress 2010). China has a large program of both reconnaissance and communications satellites utilized for military purposes:

- Reconnaissance Satellites: China continues to deploy imagery, reconnaissance, and earth resource systems that can also be used for military purposes. For instance, the Yaogan 1 through 6 satellites, the Haiyang 1B, the CBERS-2B satellite, and the eight planned Huanjing disaster/environmental satellites are capable of visible, infrared, multispectral, and synthetic radar imaging.
- Communications Satellites: China utilizes communications satellites for both regional and international telecommunications supporting both military and commercial users somewhat like a number of other countries. China also operates a single data-relay satellite, the TianLian-1, launched in 2008. Most recently, China launched the Zhongxing 20A dedicated military communications satellite into geosynchronous orbit from a Long March 3A launch vehicle in November 2010, making that launch the 14th successful Chinese space launch in that year (Long March launch of Chinese Military Satellite 2010).

Hosted Payloads or Hybrid Satellites

For nations who do not have the budget or the overall requirement for their own military communications satellite, then a more limited payload is often the best

---

[16] Sicral 2 press release, http://www.thales-transportservices.com/Press_Releases/Markets/Space/2010/Thales_Alenia_Space_and_Telespazio_sign_contract_for_Sicral_2/

[17] Overview of Russian space activities, www.russianspaceweb.com/spacecraft_military.htm

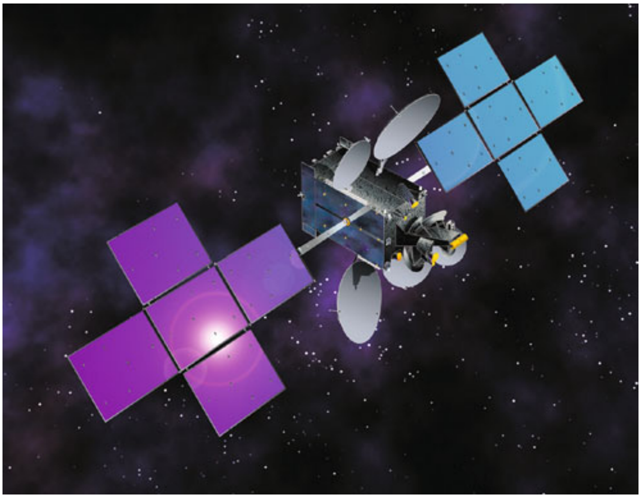


Fig. 2 Intelsat 14, host to the IP Router in Space (IRIS) Joint Capability Technology Demonstration

solution. However, since the satellite bus or platform and launch costs are the dominant factors in procurement of satellite capability, countries often search for less expensive alternatives. One such idea is the Hosted Payload concept. The term has been developed to refer to the utilization of available space on a commercial satellite's platform to accommodate additional transponders, instruments, or other items needing to be orbited. The original hybrid satellite concept is now more generally taken to mean a satellite developed with both the commercial and military payloads in mind from the start. Where the commercial operator and the government originate from the same country, these are often referred to as hybrid satellites (for example, France, Australia, and Japan have followed this approach) and where the military payload is opportunistically launched on another entity's commercial satellite they are referred to as Hosted Payloads (for instance, the Intelsat 22 UHF mission for Australia[18]). Either a Hosted Payload or a hybrid satellite can be interpreted to mean that a specific satellite fulfills multiple missions for different customers. For this chapter, we will treat the terms interchangeably for simplicity.

By offering "piggyback rides" or "hitchhiking" opportunities on commercial spacecraft already scheduled for launch, satellite firms allow organizations such as government agencies to have sensors and other equipment launched into space on a timely and cost-effective basis. The Hosted Payloads concept is similar to the ridesharing or multiple manifesting launch concept, but instead of sharing a space launch vehicle, the partners share a satellite bus (Fig. 2).

---

[18] Intelsat announcement of Intelsat 22 satellite procurement and procurement of UHF payload by Australian Defence Force, http://www.intelsat.com/press/news-releases/2009/20090427-2.asp

Hosted Payloads allow the government to plan, develop, and implement predefined space missions on much shorter cycles compared to the time it takes to design, procure, and launch an entire government satellite – typically 24 months versus many years. This is especially important for agencies facing impending gaps in operational capability. The partnership with the commercial satellite firm gives the government an opportunity to leverage an already planned or existing satellite bus, launch vehicle, and satellite operations.

Placing a Hosted Payload on a commercial satellite costs a fraction of the amount of effort required for planning, building, launching, and operating an entire satellite. The commercial partner only charges for the integration of the payload with the spacecraft and the incremental costs associated with the use of spacecraft power and fuel, launch services, and other resources. This means that the main contributor to the government costs is the dedicated payload and, therefore, the total price is far below that of deploying an independent, government-owned satellite.

Countries Employing Hybrid Satellites or Hosted Payloads

Australia

Australia, being fairly remote from the rest of the military satcom innovators, has long had a history of being innovative with its use of satcom for the Australian Defence Forces (ADF) and has implemented a multiphase program to investigate, develop, and deploy a range of military satcom capabilities. This philosophy is reflected in its Defense White Paper which is updated regularly by the ADF.[19]

In 2003, Australia launched the Optus C1/D satellite[20] into a Pacific Ocean coverage area. The satellite was owned and operated by Optus, an Australian telecommunications company, and in addition to the primary commercial satellite communications payload contained a military payload funded by, and solely for the use of, the Australian Defence Force (ADF). This payload consisted of X-band, Ka-band broadcast, and UHF payloads and served to initially augment and eventually replace the heavy reliance that the ADF had up to that point on commercial leases of satcom capacity.

France

As stated previously, France's military satcom history is dominated by hybrid satellites, both Syracuse 1 and 2 employed defense payloads on board the Telecom series of satellites, which were owned by France Telecom. The Telecom 2 series of

---

[19] Australian white paper on defence, http://www.defence.gov.au/whitepaper/

[20] Press announcement of Australian involvement in US WGS program, http://www.australiandefence.com.au/F4F2FBC0-F806-11DD-8DFE0050568C22C9

four satellites were launched between 1991 and 1996 and allowed the Direction Générale de l'Armement (DGA) access to military capability on board a national French satellite.

Japan

Japan has a long history in using satcom for its forces and has focused primarily on utilizing the frequency bands of the Superbird series of satellites. The Ku- and Ka-band payloads are used by commercial customers as well as by the Ministry of Defense and the Self-Defense Forces, but the X-band payload on each Superbird satellite is reserved for the exclusive use of the Ministry of Defense and provides military satcom for all three of the Self-Defense Forces.[21] Currently, Japan has been directed by the Japanese Diet to study the deployment of new satellite capabilities for surveillance and defense communications but no specific new programs have yet been launched.

Spain

The Spanish MOD[22] became a military satcom user with the launch of the Hispasat 1A satellite in September 1992. The satellite was placed at 30°W to provide transatlantic Ku-band commercial services between Europe, the United States, and South America. In addition to the commercial payload, Hispasat included an X-band payload for the sole use of the Spanish MOD. Hispasat 1A was followed in 1993 by Hispasat 1B, which had a similar payload configuration to 1A.

In 2001, the Spanish Ministry of Defense decided to move away from the policy of adding defense transponders to the commercial Hispasat satellites and explored the option of procuring their own military satellite. In July 2001, Loral Space and Communications entered into a joint venture agreement with HISDESAT, a Spanish company owned by HISPASAT, INTA (a Spanish government organization), and a number of Spanish aerospace companies to found a joint venture company, XTAR LLC, to lease X-band communications satellite services to the US government and its allies (Fig. 3).

As a result of the joint venture, two satellites were launched. XTAR-EUR[23] was launched in April 2005 into 29°EL and is wholly owned by the joint venture. SPAINSAT was launched in March 2006 into 30°WL and is wholly owned by the Spanish government with one of the onboard payloads serving as their dedicated military satcom resource and the other payload leased to the joint venture under the name of XTAR-LANT.[24] XTAR's capacity on these two satellites is available to be leased to the US government and its allies.

---

[21] Basic guidelines for space development and use of space, www.mod.go.jp

[22] Hispasat satellite fleet information, http://www.hispasat.com/Detail.aspx?SectionsId=67&lang=en

[23] XTAR EUR satellite information, http://www.xtarllc.com/xtar-eur.html

[24] XTAR LANT satellite information, http://www.xtarllc.com/xtar-lant.html

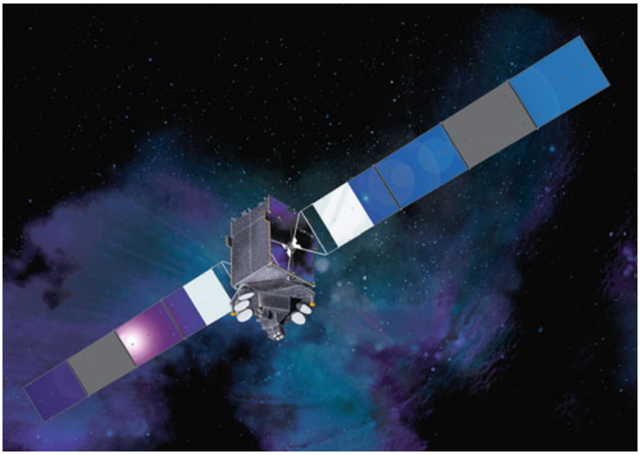


Fig. 3 XTAR-EUR satellite (Image provided by XTAR LLC)

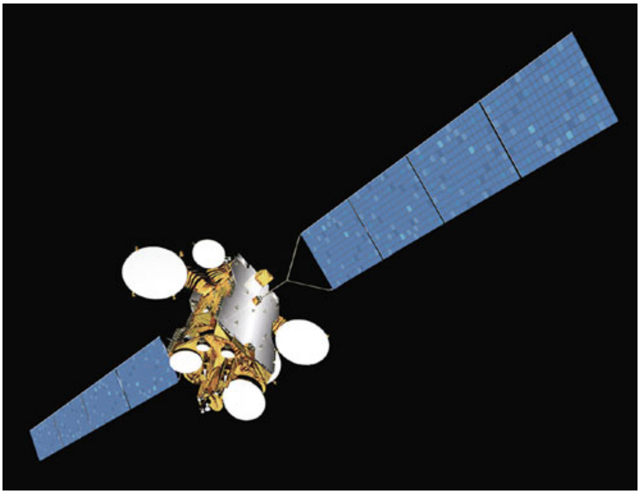


Fig. 4 Yahsat hybrid satellite for the United Arab Emirates (Image provided by Astrium Ltd)

UAE

In August 2007, Al Yah Satellite Communications Company (Yahsat) signed an agreement with a European consortium comprising of EADS Astrium/Thales Alenia Space to manufacture a state-of-the-art dual satellite communications system (Fig. 4).[25]

---

[25] Yahsat program information, http://www.yahsat.ae/yahsecure.htm

The Yahsat program will result in two hybrid satellites, launched within months of one another in 2011, allowing Yahsat to provide civil and military customers with broadcast services, Internet trunking via satellite, and corporate data networks. The system is designed to accommodate the trends of emerging applications in the satellite industry like HDTV and other broadband satellite services using C-band, Ku-band, and Ka-band commercial frequencies as well as to provide a capability that provides for the move of military satellite communications into the military portion of the Ka-band.

Intergovernmental Agreements

Countries that do not have the resources for dedicated or hybrid systems often rely on intergovernmental agreements to obtain shared resources for defense-related communications.

European Nations

Most European nations, other than the ones mentioned above, have neither the budget nor the depth and breadth of requirements to justify investment in dedicated or hybrid satellite capability. These nations have typically used intergovernmental agreements with their allies to gain access to protected communications (Germany did this with France for many years prior to launching the SatcomBW program). When intergovernmental agreements are not possible, then long- or short-term lease contracts with commercial operators or service providers have often proved to be the vehicle of choice.

Nearly every nation has now leased one or more services from Inmarsat to include within its military portfolio for maritime or airborne communications, and this has been augmented over the last 5–10 years with leases of Intelsat, SES, or Eutelsat capacity and more recently with commercial X-band communications leased from either Paradigm or XTAR. These nations include Belgium, Czech Republic, Denmark, The Netherlands, Poland, Portugal, and Slovenia. The launch of the Inmarsat Vs, Intelsat Xpress, and new Inmarsat VI satellites under procurement will likely increase this usage.

NATO/France/Italy/United Kingdom

In May 2004, the NATO Consultation, Command and Control Agency (NC3A) decided to move away from owning and operating its own fleet of satellites and selected a multinational proposal to provide SHF and UHF communications through to 2020. This program, entitled the NATO Satcom Post-2000 (NSP2K) program[26,27]

---

[26] DISA overview of NSP2K program, csse.usc.edu/gsaw/gsaw2005/s9f/stoops.pdf

[27] NATO overview of NSP2K program, http://www.nato.int/issues/satcom/index.html

requires the French, Italian, and British governments to provide NATO with access to the military segment of their national satellite communications systems – Syracuse 3, SICRAL, and Skynet 5, respectively – under a memorandum of understanding. This lease contract replaced the previous constellation of NATO IV satellites (discussed earlier) which were owned by NATO and operated by the UK MOD under an MOU with NATO.

NATO member nations are able to use the NSP2K capacity for their forces' communications needs whenever they are on a NATO exercise or operational deployment. The use of these satellites for national requirements, albeit on an ad hoc basis, has contributed in no small part to the perceived reticence of the nations with smaller requirements to procure long-term commercial satcom solutions for their national satcom needs. As NATO capacity requirements are increasing, spare capacity within NATO's allocation across the three fleets is often not available for individual member nations to use to satisfy their national requirements, and therefore it seems likely that NATO nations will increasingly seek alternative sources of military satcom capacity.

Australia/United States/United Kingdom

The Australian Defence Force (ADF) followed up the launch of Optus C1 with studies into the potential to enhance its capability by procuring capability in the Indian Ocean Region (IOR) and eventually procuring a more capable system for the Pacific Ocean Region (POR).

For the Indian Ocean Region, the ADF signed an MOU with the United Kingdom to allow access to the Skynet 5 capabilities and then followed this with an exclusive contract with Paradigm to secure UHF capacity on the Skynet 5B satellite.

For the POR and worldwide coverage, the ADF opted for access to the WGS constellation by signing an MOU with the US government in 2007.[28] This allows Australian forces access to the full constellation of five satellites and permitted the United States to expand the WGS constellation to six on-orbit spacecraft as, under the terms of the aforementioned MOU with the United States, the ADF agreed to provide sufficient funds to procure the sixth satellite. WGS1 was launched into the POR in late 2007 and became the ADF's primary satellite.

Italy/France

Athena-Fidus is a French-Italian geosynchronous military and governmental EHF/Ka-band wideband communications satellite capable of data transfer rates of up to 3 Gbps. Jointly procured by the French and Italian space agencies and defense procurement agencies, the system is intended to be used by the French, Belgian,

---

[28] Space daily report on ADF entering the WGS program, http://www.spacedaily.com/reports/Australia_To_Join_With_United_States_In_Defence_Global_Satellite_Communications_Capability_999.html

and Italian armed forces as well as the civil protection services of France and Italy. Athena-Fidus is now launched and supporting the military communications needs of its three sponsors.

Sicral 2, as described previously, is a joint Italian-French military satcom program that operates in the UHF and SHF bands. It augments both the Sicral and Syracuse systems. The Sicral 2 program primarily supports satellite communications for the two countries' armed forces and is designed to meet the needs expected to develop in the near future. Like its predecessors, the new satellite and ground segment provides strategic and tactical communications links for both domestic operations and foreign deployments. It supports all military, terrestrial, naval, and aerial platforms, operating in a single integrated network.

Guaranteed, Long-Term Leases and Ad Hoc Leases of Capacity

A lease contract is normally designed to deliver the twin objectives of managing both a nation's operational effectiveness and its defense budgets by ensuring that the nation has access to the required amount of guaranteed communications support without jeopardizing the ability to execute national deployments. Nations with only limited budgets or very small commercial augmentation requirements tend not to rely on a specific solution for their commercial satcom and simply procure what they need, when they need it on an as-available basis. Sometimes this is because their national defense system satisfies everything they might need, and sometimes it is because it is simply not cost-effective to preorder commercial capacity.

Obviously, in an environment where the supply of commercial capacity is limited, guaranteed access is not possible without placing precommitment contracts or reservations with the operator, so a necessary prerequisite of an ad hoc approach is that assured access is not a mandatory requirement for the users. This approach is often taken by nations first testing whether they need satellite communications before assigning specific budget lines for it. Any leases for commercial capacity that are subsequently entered into tend to be short term because these governments normally cannot predict their future demand, and this prevents them from committing to lower-cost long-term contracts.

Several NATO nations with smaller military satcom requirements have opted to augment capacity provided under MOU with commercial leases to support their national operations. Often this approach splits out the procurement of military X-band capacity from commercial capacity but equally it can group all satcom requirements under a single commercial contract. Although Canada and Norway have entered into X-band lease agreements with Paradigm or XTAR as a precursor to a dedicated national solution in the future, more often a commercial satcom lease agreement is intended to be the long-term solution for a nation's military satcom needs.

Nations who can predict their future requirements with a little more certainty or who have more flexibility enter into long-term leases of capacity based on their best analysis of long-term requirements. Belgium, Denmark, The Netherlands, and

Portugal procure their X-band military communications from either Paradigm or XTAR under fixed-term contracts with either a fixed amount of capacity assigned to them for their dedicated use or with the ability to call up capacity as they need it. The Netherlands have expanded on this approach for their C-band and Ku-band having entered into a multiyear contract in 2005 with New Skies (now part of SES World Skies).

These types of contract result in assurances that the amount of contracted capacity will be available for use by the military when it needs it but does require some risk taking on the part of the customer that the capacity will be available where it needs it throughout the contract period as the requirements analysis is normally done before entering into the contractual arrangement.

Government-Industry Partnership

The sixth and newest option for governments, as highlighted above, has been implemented by the United Kingdom. The United Kingdom has had a military satcom program since Skynet 1 was launched in late 1969, making it the third country to launch its own national military satcom system. The United Kingdom has continued its involvement in military satcom up to the present day with the Skynet 4 series (now retired) and Skynet 5 series of satellites.[29] Three new Skynet 5 satellites were launched between March 2007 and June 2008. A fourth satellite, Skynet 5D,[30] was launched in 2013. The United Kingdom uses the Skynet constellation for its protected communications and employs both UHF and X-band frequencies (Fig. 5).

Since 2003, the UK military satcom system has been owned and operated by a commercial company, Paradigm Secure Communications Ltd, which is solely responsible for providing the national critical communications to the UK MOD using the Skynet space and ground systems. This is done under a contract vehicle called the "Skynet 5 Private Finance Initiative (PFI) program." Paradigm took ownership of the existing Skynet 4 military satcom system (now retired) and then obtained loans from the worldwide finance community in order to design, build, and implement the next generation Skynet 5 system.

Not only does Paradigm provide the UK MOD with guaranteed access to the state-of-the-art Skynet 5 satellite system for its highly protected, dedicated military satcom requirements, it also guarantees to supply all of the UK MOD beyond line of sight (BLOS) communications requirements. The UK MOD therefore specifies the service characteristics of any communications link that it needs and Paradigm defines the system solution that it is best able to supply from an operational and

---

[29] Overview of Skynet 5 program, http://www.army-technology.com/projects/skynet/, http://www.astrium.eads.net/en/programme/skynet-5-.html

[30] Announcement of Paradigm's fourth Skynet 5 satellite, http://www.astrium.eads.net/en/press_centre/paradigm-agrees-deal-with-uk-ministry-of-defence-mod-for-fourth-skynet-5.html

Fig. 5 Skynet 5 series of satellites (Image supplied by Paradigm Secure Communications Ltd)

cost-effective perspective to meet that requirement. This may be achieved using military or commercial satellite capacity or ground-based fiber or GSM technology – whichever is most appropriate and meets the UK MOD's requirements.

Paradigm has to maintain the extremely high availability of the military satellites that it owns and operates and ensure that the ground systems are fully operational at all times. However, to execute its responsibility to supply all beyond line of sight communications, Paradigm also has to ensure that it has access to sufficient commercial satcom in a variety of frequency bands, teleport assets in remotely diverse locations and enough fiber leased lines to connect all locations and customer sites. This is done through a variety of long-term leases and Paradigm conducts frequent recompetitions to ensure cost-effectiveness. There is an agreement between Paradigm and the MOD on an incentive scheme to ensure that everything possible is done to provide the capacity needed.

This solution is only successful because the traditional roles of supplier customer are deliberately blurred in the PFI approach. While it is true that there is a comprehensive and detailed contract in place between the UK MOD and Paradigm, a large part of the relationship has to be based on trust. The Skynet 5 contract duration is for a minimum duration of 19 years and the MOD requirements must be met in 2022 just as they were in 2003. Because the very nature of communications requirements is that they are constantly and rapidly evolving and expanding, it is not enough to simply say "the capacity will be there when you want it." Paradigm and MOD therefore work very closely, at both the working level and the management level, to ensure that new developments in requirements are shared as soon as possible. In this way, Paradigm can make sensible investment decisions because it understands that the users will be there once the capacity is available and MOD can rely on the capacity being available for future platforms and applications because it worked closely with Paradigm to ensure those requirements have been taken into account within the joint planning process that they share.

To date, this symbiotic relationship between the UK MOD and its industry partner is unique but is being closely monitored and reviewed by many other nations, as can be observed by France's investigation into the outsourcing of Syracuse during 2009 and 2010 (as discussed later in this chapter).

Commercial Satellite Communications Augmentation

Europe's military and defense forces now procure an increasing percentage of their satellite communications capabilities from commercial sources, with some nations approaching 40 % through commercial leasing. In contrast, the US DoD procures as much as 80 % of its total satcom capability commercially through long- and short-term leases and has at times even exceeded that level. While originally this capacity was procured as ad hoc leasing of commercial capacity for urgent requirements and to cover shortfalls and "gaps," there is a growing tendency among all nations to look toward a more centralized procurement model. An overarching contract vehicle goes some way to alleviating some of the problems associated with an ad hoc commercial satcom requirement. Terms and conditions are pre-agreed with one or more suppliers, ensuring that if capacity is actually available when needed, there is no delay in activating the capacity because of protracted contract negotiations. There is also more likelihood that the contract will be flexible enough to grow and change with the customer's requirements.

In 2001, to enable capacity from the commercial satellite operators to be procured to augment the increasing military requirements, the US DoD's Defense Information and Systems Agency (DISA), issued the Defense Information Systems Network Satellite Transmission Services-Global (DSTS-G) contract. It was an Indefinite Delivery Indefinite Quantity (IDIQ) contract, allowing the military communications users to procure as much or as little commercial services and capacity for as long or as short a time as it wants under an overarching set of contractual terms.

Although some commercial capacity and services are still leased via other means by some DoD elements, the majority from 2001 to 2011 were procured through this contract vehicle. The DSTS-G contract has been replaced with a new program jointly administered by the Government Services Administration (GSA) and DISA. This program, entitled the Future COMSATCOM Services Acquisition (FCSA) program, commenced in early 2011 and is discussed later in this chapter.

The US model of using one overarching contract vehicle and then procuring each element of commercial capacity underneath this "umbrella" has proved to be extremely cost-effective and, while not necessarily being focused on delivering value for money or operational effectiveness, is becoming more popular with allies. In Europe, the procurement of both military and commercial satellite communications is characterized by smaller procurement budgets and, historically, a mistrust of national consolidation. Therefore any method of reducing procurement costs is welcomed and embraced. European nations are increasingly looking at more innovative ways of satisfying their commercial satcom needs.

The French ASTEL-S contract (awarded in 2005) aims to go one step further than DSTS-G by providing fixed tariff sheets for commercial capacity over a wide coverage area and for a fixed period of time. Capacity is still provided on an "as-available" basis by the contractor (Astrium Services Ltd), but the French Navy can plan its budgets in advance due to the surety of the pricing and the contract terms. Astrium Services takes the risk of providing the capacity for the price specified in its fixed-term contract.

The European Defense Agency (EDA) is currently setting up the EDA Satellite Communications Procurement Cell (ESCPC) to fulfill a similar function to DSTS-G for European nations. However, the ESCPC is designed to not only provide a contract vehicle for nations to buy commercial capacity under but also to pool the demand for all European nations through a central procurement body. This allows a lower cost per Megahertz to be negotiated by the procurement cell and for those savings to be passed on to the member nations. Current estimates put European governments' total expenditure on commercial satellite capacity leases in the region of 50 million euros (which is about US$72 million) per year. This program is expected to save participating governments as much as 30–50 % on spot market spending.

Finally, there is the concept of an end-to-end service contract whereby the military procurement agency estimates its long-term needs and then contracts with an industry partner to guarantee this capability throughout the contract lifetime without the customer needing to specify the technical solution to be used. This is precisely the situation with the UK MOD and its contract with Paradigm, discussed earlier in this chapter.

A fundamental remaining question is at what point does "augmentation" capacity become "core" capacity, critical to the warfighter's capability? In the case of the US DoD, as previously stated, over 80 % of the required military satcom capacity is now procured commercially rather than using dedicated US satellites. It is therefore difficult not to believe that the commercial capacity is as much "core" capacity as the dedicated capacity – a situation that would have been impossible to imagine even 10 years ago. This condition is strongly shaping the future of military satellite communications procurement and policy.

The Future

There are historically two major components to military communications traffic: strategic and tactical. Both have an impact in shaping the way the future looks for commercial and military satellite communications:

- Strategic traffic tends to be high data rate, fixed location to fixed location, and relatively easy to predict for a significant period of time. This enables solutions to be deployed using fixed coverage beams and for capacity to be committed over a longer period of time.

- Tactical traffic is characterized by the use of smaller ground terminals in dispersed locations. While data rates for tactical traffic can still be high (and are growing all the time), mobility and flexibility are of paramount importance. The solution calls for rapid redeployment and reconfiguration of assets both in space and on the ground. It is often very difficult to make long range predictions about the precise location of the deployments and the total capacity needed.

In the future, military planning units will continue to see an increase in theaters of conflict being engaged on multiple fronts in disparate locations. This will lead to a shift toward an increase in tactical traffic and, potentially, a decrease in strategic traffic. It is also foreseen as more likely that strategic communications will switch more to other forms of communications and be less dependent on satellites. Since tactical traffic is by its nature harder to predict, this will put greater emphasis on more flexible and capable communications solutions able to respond to an ever changing military environment.

The military satellites currently in production for launch within the 2012–2015 time frame are already starting to incorporate more and more transponder power to support the increased throughput requirements and more flexibility in the shaping of the spot beams in order to satisfy these more intensive "tactical communications" needs. However, it is apparent that some of these needs are overstretching the industry with the quantum leaps in capability and the pressure being put on design and implementation schedules. The US DoD has decided to split the WGS program into two phases to allow phase 2 to be modified for requirements which were not apparent when phase 1 was completed. In Europe, Syracuse 3C was canceled in favor of investigating an outsourcing approach coupled with the joint approach with Italy on Sicral 2. NATO chose to procure its satcom through MOU rather than to replace the NATO IV satellites with a more capable NATO V series.

Therefore, one might assume that with a number of dedicated military satcom programs being merged, changed, or canceled, there is an opportunity for commercial satcom to become an integral part of the military warfighter's arsenal instead of always being referred to as an add-on, augmenting the critical national infrastructure. However, in conflict with the need to replace or augment military capacity is the US DoD's increasing need for flexibility in support of its current theaters of operations. The existing commercial satellites can only partially satisfy these types of requirements and this has been at least partly responsible for the world shortage in commercial satcom capacity, especially within the Middle East and Asia.

Commercial satellite operators have been unable to procure additional satellites with more flexibility and capacity optimized to defense-related needs in order to meet growing military requirements. This is because an operator has to present a viable business case to its shareholders showing that revenue will be recovered over the lifetime of the satellite to offset its investments. The US DoD (and other MoDs around the world) often have difficulty defining a core or fixed requirement in terms that will allow an operator to take a risk on the revenues that it will receive. Dialogue between US DoD and industry on this topic has been steadily increasing over the last few years, mirroring that which has been taking place in Europe over the last decade.

What, therefore, are the defense procurement agencies and satcom industry focusing their efforts on and what trends will be increasingly apparent over the next 3–5 years?

1. More outsourcing of critical and noncritical communications from the military operators into industry
2. New and improved contract vehicles will be introduced designed specifically to improve flexibility
3. Increased investment in Hosted Payloads by governments around the world
4. International partnerships between allied nations

Each of these four trends is described below using a specific example to illustrate the trend:

1. Syracuse outsourcing

   France has recently issued requests for proposals to outsource its Syracuse system to an industry partner and lease back communications services for the lifetime of the satellites. This contract will be similar to the German and UK programs, and will leverage the lessons learned by these nations, while retaining a French national independence. Interestingly, the plans look set to include the future Sicral 2 satellite, which means that the French government is planning from the outset to have an element of not only protection from the future growth in capacity requirements but also international collaboration to maintain value for money.
2. Future COMSATCOM Services Acquisition (FCSA) program

   The DSTS-G contract for the procurement of commercial satellite communications services by DISA was due to expire in February 2011 but was extended until its replacement contract, FCSA, is fully in place. Although there remains a range of different contract vehicles for procuring capability in place across DoD, FCSA will be the main vehicle for DISA and DISA customers for the foreseeable future.

   The FCSA program consists of a set of acquisition parts that replace three expired DISA and GSA contracts, including DSTS-G, Inmarsat, and SATCOM II. Previously under the GSA schedule 70 and the DSTS-G contract there were a limited number of firms leasing capacity directly, and only three firms, ARTEL, CapRock, and DRS, were permitted to sell satcom services directly to the US government. The implementation of the FCSA program will bring a major change to how satellite capacity and services are procured.

   The FCSA program has two new General Service Administration (GSA) Schedule Item Numbers (SINs) under the GSA IT Schedule 70. These new SINs, 132–54 transponded capacity and 132–55 subscription services, are open to bids on a continual basis. They will provide specific satellite services requiring no development or systems integration activities. The FCSA program will also have two IDIQ contracts for providing end-to-end communications satellite solutions.

In addition to allowing several more organizations to lease transponded capacity directly to the government, there will be several new service providers entering the market in addition to the three original DSTS-G providers. This will include the commercial satellite operators as well as those service providers without their own satellites who will be able to purchase satellite capacity from the satellite operators and resell it to the US government.

From the government's perspective, this maintains the current situation of allowing the warfighter to obtain the required capacity and services at lower costs through ongoing competition while enhancing the scope and nature of the marketplace. The government will be able to select services from a much wider range of competitors and technologies on an ongoing basis as the requirements evolve, while still ensuring that government assurance and protection requirements can be met. It is intended that this contract vehicle will be so all-encompassing that a communications procurer will be able to procure services from a few kilobits all the way up to a full payload capability for multiple users. It will take some time before this can become a reality, but it is destined to change both the way in which the procurement authority thinks about its requirements as well as the way in which industry sets itself up to address the evolving and ever more flexible requirements.

3. Hosted Payloads

The Hosted Payloads concept has gained significant popularity within both government and industry. Satellite companies, recognizing the opportunity to further monetize their capital investments, have created new divisions focused specifically on Hosted Payloads. Government agencies, facing new budgetary realities, have issued solicitations and held special invited "industry days" to investigate the cost and feasibility of various commercial solutions, including Hosted Payloads, as a way of fulfilling their mission requirements.

For government agencies, a key challenge to developing and launching a Hosted Payload is the ability to meet the rapid pace of commercial satellite development. Satellite operators have hard, fixed deadlines for launching their spacecraft in order to meet the huge commercial demand for communications. In many cases, the satellites being launched are replacing older ones that have degraded performance or are reaching the end of their useful life. Communications satellite companies cannot afford to delay replenishment satellites to accommodate developmental problems that can often occur with government payloads.

In August 2009, the Office of Space Commercialization, FAA's Office of Commercial Space Transportation, and Futron organized the first government-industry workshop on Hosted Payloads to share lessons learned and develop a common approach to facilitate governmental use of Hosted Payloads. Futron organized follow-on workshops in April and July of 2010 to develop approaches, recommendations, and options for moving forward. It will be interesting to follow how the Hosted Payload concept evolves.

While nations like Australia have agreed to add operational payloads to commercial satellites as they are doing with Intelsat 22, to date the US government has not followed suit and has only placed research and development payloads on commercial satellites. Interestingly, the US National Space Policy published in June 2010 contains specific language encouraging the US military to obtain space capabilities using more innovative approaches including Hosted Payloads. As stated the US government should "work jointly to acquire space launch services and Hosted Payload arrangements that are reliable, responsive to United States Government needs, and cost-effective."

In Canada, Telesat took the approach in 2010 of installing a three channel X-band payload, exclusively for government use, on board its new Anik G1 commercial C-band and Ku-band satellite.[31,32] This satellite, which will be located at 107.3°W when it goes into service in the second half of 2012 is a multimission spacecraft predominantly for direct-to-home (DTH) television broadcasting in Canada and broadband, voice, data, and video services in South America. However, in a move viewed as daring by industry experts at the time, Telesat decided that there was sufficient latent need for government users in the Continental United States and Pacific regions that it would initiate a Hosted Payload program at its own risk. Within only a few months after Telesat's announcement, the full portion of the X-band capacity was purchased by Paradigm to augment the coverage provided by its Skynet 5 fleet and satisfy the needs of Paradigm's existing customers which are not served by Skynet today.

Intelsat is following on from the success of its involvement in the Internet Router in Space (IRIS) Hosted Payload and its Australian Defence Force UHF payload on board IS22 with its Intelsat 27 satellite.[33] This satellite will carry a UHF payload identical to that on board Intelsat 22 but, similarly to Telesat when it announced the Anik G1 contract without a customer for its X-band payload, does not yet have a committed customer to take the UHF capacity.

Iridium has also proceeded with its Aerion for a Hosted Payload initiative that will provide for increased air traffic control, navigation, and management.[34] The Iridium Next constellation of 66 satellites, which is now scheduled for launch in 2017/ 2018 to replace the original Iridium constellation, has Hosted Payloads at the heart of its vision. The deployment of this system has been delayed due to launch failure reviews associated with the Space X Falcon 9 launcher.

4. More national alliances

---

[31] Announcement of Anik G1 satellite by Telesat and Loral, http://www.spaceref.com/news/viewpr.html?pid=30941

[32] Announcement of Paradigm's leasing of Anik G-1 X-band capacity, http://www.spacenews.com/satellite_telecom/101013-paradigm-xband-anik.html

[33] Announcement by Intelsat of its intent to launch Intelsat 27 with a UHF hosted payload, http://satellite.tmcnet.com/topics/satellite/articles/95425-intelsat-is-27-satellite-launch-2012.htm

[34] Iridium NEXT program will include opportunities for hosted payloads, http://www.iridium.com/about/IridiumNEXT/HostedPayloads.aspx

As well as the alliances referenced above for the use of WGS by Australia and the partnership within AEHF between the United States, Canada, Netherlands, and the United Kingdom, there is a growing interest in national alliances in Europe. This is led by France and Italy who are already collaborating on the Sicral 2 satellite and the Athena program as referenced above, but there are also senior level discussions between the United Kingdom and France on a whole range of defense topics.

The Norwegian government has recently decided to launch a partnership with Spain for the purchase of a military communications satellite to contribute to the stability and effective monitoring of Norwegian interests and to support the increasing armed forces' participation in operations abroad where there is no necessary communications infrastructure. The project, which is called HisNorSat, is designed in cooperation with the company HISDESAT and will become operational around the 2014 time frame. The satellite will be partly owned by the Spanish MOD. The partnership will give Norway ownership of a defined part of a joint communications satellite with full control of the Norwegian-owned portion of the satellite which will operate in both X-band and Ka-band. The partnership will give the Norwegian MOD access to a capability far more sophisticated than if it were to procure a stand-alone satellite or even a Hosted Payload on board a commercial satellite. It seems logical that more nations will opt for this approach in the future to exploit synergies in military communications requirements and allied operations in the same geographical regions while increasing cost-effectiveness.

Conclusion

The development and use of communications satellites followed shortly after the launch of Sputnik. Initially in the United States and the United Kingdom commercial and military satellites were on different paths with specific satellites used by each for their own missions with very little use by the military of commercial assets. In other countries, with fewer requirements and smaller budgets, different commercial and military payloads were placed on the same satellite usually with a small military payload on a commercial satellite. In the modern day, development of commercial and military satellite communications programs is not only converging, but these programs and satcom assets are increasingly critical as a joint solution to satisfy a nation's communications needs.

As has been repeatedly noted, some 80 % of the US core military satellite communications requirements objectives are provided by commercial satellites and they can no longer be considered a supplement to dedicated military systems. This has, of course, raised a number of issues concerning the suitability of commercial satellites to carry sometimes quite sensitive traffic. The current pressing operational needs, and the cancellation of future military satcom programs, have forced US military organizations to utilize commercial satellites with no increased enhancements except satellite command encryption and encryption of the traffic being transmitted. Conversely, in Europe the majority of requirements for those nations with access to national infrastructure are still supported on the national military systems with commercial satcom playing a growing part in national infrastructure but still fulfilling an augmentation role. With defense budgets being increasingly constrained, European innovation has focused on increasing value for money without sacrificing operational effectiveness of the warfighter.

From all public predictions, the future over the next 5 years implies a growth in military satcom requirements. This implies the introduction of new platforms requiring more and more data transfer capabilities. This trend suggests that the number of dedicated satellites in orbit will only grow steadily. With the growth in number and sophistication of the UAVs being used worldwide, the need for large amounts of bandwidth to transmit the UAV data to processing facilities appears inevitable. However, there has been no indication on the part of the US or UK governments or others of a need for military Ka-band from commercial satellite firms. Consequently, the satellite companies are unlikely to plan for or launch military Ka-band capability without some indication of probable use. In addition, there is a dearth of military Ka-band terminals to support the reception of the data. While initial Ka-band users will be forced to use dedicated military satellites for their capacity, there will therefore be a continued reliance by the military on other commercial satellites to satisfy a significant portion of their extant and future communications requirements. With an ever increasing number of commercial companies entering the business of providing communications services to governments, it is clear the landscape will be more uncertain and competitive.

The continuing question for the military and industry alike is how to best provide for the capacity and guarantee value for money without sacrificing military effectiveness.

Governments have already developed a variety of financing techniques to access the required capacity. These range from overarching lease contracts that encourage innovation and competition within industry through to solutions such as the United Kingdom's Private Finance Initiative (PFI) where the UK MOD sold its military satellite assets to a private company that operates the system for them and provides both dedicated and growth capacity.

The focus on Hosted Payloads provided by, among others, Intelsat, Telesat, and the Australian Defence Force provides a great opportunity for industry and military to work together to provide not only adjunct capacity but core capacity in a timely fashion. Developing a business case for a Hosted Payload that not only meets the military's needs but also the schedules of the commercial operators is a fundamental challenge that has only just started to be investigated.

The business of providing commercial capacity to governments for military or other government uses has truly become an international business with many different players on both the provider and the customer side. National boundaries are becoming increasingly blurred as coalition forces are increasingly being deployed across the world and interoperability between those forces becomes a given rather than an option. Ensuring that national security requirements can continue to be satisfied in an ever increasing international environment will continue to be a challenge. France/Italy and Spain/Norway are leading the international collaboration developments on their future military communications satellites and will be working hard over the next decade to ensure that national requirements can continue to be met while sharing physical assets.

This chapter has allowed us to present what has happened in the past with the use of military satellite communications and how this history has shaped the present day environment and the increasing usage of commercial communications by the military. The military satellite communications world has always been dynamic and innovative. The next 10 years will see great changes in the area of defense and strategic satellite communication systems. These changes will come not only in specific military and dual-use commercial technologies but also in the creation of yet more innovative business models that have never been seen in the industry before. The objective, however, will remain the same. This is to provide the military users with the communications they need, when and where they need them.

Cross-References

- ► Fixed Satellite Communications: Market Dynamics and Trends
- ► Future of Military Satellite Systems
- ► Mobile Satellite Communications Markets: Dynamics and Trends
- ► Satellite Communications Video Markets: Dynamics and Trends

References

Annual report to congress: military and security developments involving the people’s Republic of China 2010, http://www.defense.gov/pubs/pdfs/2010_CMPR_Final.pdf

DISA conference proceedings 2009 for the commercial satcom session, http://www.disa.mil/conferences/2009/briefings/satcom/Commercial_SATCOM_DISA_Conference_2009.ppt (slide 33)

Long March launch of Chinese Military Satellite: November 2010, http://www.space.com/9606-chinese-military-communications-satellite-reaches-orbit.html

National communications system fiscal year 2007 report, http://www.ncs.gov/library/reports/ncs_fy2007.pdf, p. 26

Satellite 2001 daily news: military bandwidth migration path leads to Ka-, X-band satellite offerings, http://www.satellitetoday.com/eletters/satellite2011_daily/2011-03-11/36343.html

Economics and Financing of Communications Satellites

Henry R. Hertzfeld

Contents

Abstract

The economics and financing of satellite communications is a very large and complex topic. It ranges from normal business planning, analysis, and investment financing, to issues of government policy, dual-use technologies, and national security and defense. Commercial satellite systems represent a special case of economic analysis since such systems are heavily dependent on a government market that is focused on political considerations of budgeting and regulation. Today, satellite telecommunications systems are critical to almost all nations of the world, and they are especially important in approximately 60 nations that have domestic launch and/or satellite operations capabilities. This chapter will specifically focus on four topics: (1) a summary of the economic characteristics of the industry and a review of major trends in the industry, (2) a summary of the elements of a business plan for satellite telecommunications, (3) an analysis of

---

H.R. Hertzfeld (✉)
Space Policy Institute, Elliott School of International Affairs, The George Washington University, Washington, DC, USA
e-mail: hrh@gwu.edu

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_9

issues in the manufacturing productivity for satellites and an analysis of commercial satellite manufacturing compared to government satellites, and (4) a brief discussion of future cost considerations including the increasing risk of space sustainability, insurance, and rules concerning disposal of satellites after their useful lifetime.

Keywords

Auction of spectrum • Commercial satellite systems • “Dual use” of satellite networks • Economics • Insurance • Investment financing • Launch costs • Manufacturing • Market sectors • Operating and capital costs of satellite networks • Satellite services • Satellites • Size of markets • Telecommunications • Video services

Introduction

In 1876, Alexander Graham Bell invented the telephone. The invention spread rapidly and became the standard mode of remote voice communications during the first half of the twentieth century. Copper wires were strung and these became the major mode for the transmission of voice communications.

In the 1940s, Arthur C. Clarke suggested the possibility of using geostationary satellites to beam telecommunications signals from a point on Earth back to multiple points. However, during the mid-twentieth century, telephone signals were being relayed on land with copper cables and microwave towers and across the oceans with similar cables of limited capability compared to those that are in use today. During the 1960s, the first telecommunications satellites were launched successfully to low Earth orbit (LEO).

Although communications satellites had greater capacities than ground systems for overseas transmissions when first deployed in the 1960s and 1970s, this was no longer true when faster terrestrial communications using fiber-optic cables were developed in the 1980s. Cable transmissions of any type, it should be noted, are best designed for point-to-point communications, while satellites are best for point-to-multipoint uses.

Local commercial television broadcasts over the air were inaugurated in the late 1940s and grew rapidly. The larger selection of stations enabled by subscription cable delivery of television to households gradually became a standard form of TV delivery in most countries of the OECD. By the 1990s, national and international cable TV distribution was widespread and direct broadcast satellite TV to consumers was also beginning to grow, enabled by smaller terrestrial receiving antennas and more powerful satellites (Satellite Industries Association, http://www.sia.org/satellites.html). Satellite delivery is particularly advantageous in remote areas not served by cable but is also a strong competitor to cable in urban areas. As mentioned above, it is also the only system that can effectively deliver point-to-multipoint signals in nations that are not well wired, such as in developing countries.

There has been a very rapid and dramatic change in developed economies over the past two decades in terms of telecommunications delivery media. This has been a shift from wireless TV and wired telephone to wired (cable) TV and wireless cell phone systems – although direct broadcast satellite services now represent a significant delivery mode in many economically advanced countries as well. This shift is also beginning to occur in developing countries as well, but some of these patterns are less pronounced and some satellite networks (such as O3b) are seeking to provide broadband Internet services and video services directly to consumers in Africa and other parts of the world and thus seeking to bypass terrestrial cable networks.

These shifts are particularly significant for the satellite and space industry. Space capabilities are actually central to all systems today by being a part of the overall telecommunications services delivery chain. Cable TV, although wire-based to connect to homes, is dependent on network uploads via satellites. Cell phones and systems are also linked and coordinated by backhaul precision timing through GPS satellites. And, of course, satellite TV and radio are now ordinary consumer services in a very competitive market sold and distributed by many companies in many advanced economies such as the United States, Canada, Japan, Korea, Australia, and Europe to name only some of the countries where competitive satellite services are sold.

Satellite Telecommunications Services

Satellite communications networks are separated into several types as noted in earlier chapters and particularly chapter "▶ Space Telecommunications Services and Applications." Fixed satellite services (FSS) typically transmit between a GEO satellite and one or more fixed locations terrestrially. The various types of telecommunications services that FSS networks provide include television distribution, broadband data, and voice communications. Some higher-powered FSS networks provide direct to home video and audio services.

So-called broadcast satellite services (BSS), as defined by the International Telecommunication Union (ITU), provide direct transmission to the home via very small dishes. There are also what are called direct audio broadcast services (DABS) or satellite radio services. Although different frequencies are allocated by the ITU for FSS, BSS, and DABS services, the distinctions between these services are not always very clear to consumers. This is because higher-powered FSS satellites can provide services that look and act like direct broadcast satellites services for either television or radio services. Overall revenues from these various types of satellite networks are heavily weighted toward video services. Television distribution and direct television satellite services generate over 75 % of the revenues, with data transmission representing about 20 % and voice services representing only some 5 % – at least for satellites serving the most economically developed countries.

Mobile satellite services (MSS) transmissions are conducted from satellites to receivers that are not fixed in any one point terrestrially. They include maritime, aeronautical, land mobile services, and other transportation-related uses, including emergency and police communications. Many of these networks are private.

In addition, there are many support services included in the satellite industry. As also described below, they include the manufacturing of satellites, launching services to get the satellites into proper orbits, ground receiving equipment, and many financial, insurance, and other business services.

All components of the industry have been growing, even during the recent economic recession beginning in 2008. Worldwide, the overall telecommunications satellite industry now generates the largest amount of revenues – by far – of any commercial space industry. The largest component of the communications satellite industry, in terms of revenues, is for provision of telecommunications services; this represented 58 % of total industry revenues in 2010. This was followed by revenues from the sales of ground equipment (31 %), manufacturing (8 %), and launches (3 %). (Satellite Industries Association, http://www.sia.org/satellites.html).

There are many satellite applications that provide opportunities for very useful and profitable businesses. Figure 1 in chapter "▶ Satellite Applications Handbook: The Complete Guide to Satellite Communications, Remote Sensing, Navigation, and Meteorology" of this handbook provides a good overview of the many types of satellite applications that affect our daily lives. Over time, it is clear that many of these applications have become a critical part of the economic infrastructure. Of the various market sectors of the commercial satellite world, communications satellites predominate in terms of total revenues, number of users around the world, and direct impact on people's lives. Despite this predominance of satellite communications, the other services, such as remote sensing and space navigation, are still greatly important. Satellite meteorology is typically not a commercial service, but it is nevertheless vital to public safety.

The Business of Satellite Communications

Satellite telecommunication, in light of its huge revenue stream and the billions of consumers that use this service, is clearly the most mature of space applications. In fact, some economists would say that the 50-year-old communications satellite industry virtually represents the only example of a mature commercial use of space.

Planning, financing, building, launching, and operating a communications satellite or satellite system has become routine business. It is a long-term investment, and fits a standard business model. Satellite systems require a high up-front capital expenditure, a reasonably long manufacturing and start-up period (over 2 years), and face a number of high investment risk factors. In spite of the obstacles, satellite telecommunications has proven to be a space application that can generate a long-term multibillion dollar (US) revenue stream and profitable returns.

These systems are in some ways very different from most industries and in other ways identical. The differences are centered on the large government presence in the technological developments as well as the role of governments as a purchaser of these services. The similarities are like those with any other regulated infrastructure or utility that requires complex and expensive investments in equipment and/or distribution systems and that provide essential services to a large number of people. Because of these similarities to any private investment, this chapter will focus on specific topics unique to satellites and space businesses and will not attempt to describe normal business and economic issues that can easily be found in any basic management or economics textbook.

Early space telecommunication systems were not standard business ventures. They were built from a combination of public and private research and development (R&D) investments and required access to space. This vital launch service, in the first decades of communications satellite service, could only be provided through a government launch vehicle. Until the late 1970s, the only vehicles capable of performing the launch services were either in the United States or the USSR. And, the USSR was not in the commercial launch business and did not launch private satellite payloads. The US government's involvement in technological development and regulation, government purchase and use of the services (both military and civilian), and government policy were integral to any corporate telecommunications business plan.

The US government's role has dramatically changed over time. But it is still very important today. Although the US Department of Defense (DoD) has its own satellite telecommunications system, it also purchases a large amount of commercial capacity to fulfill its total communications needs.

In addition, in recent years the DoD has dedicated transponders and instruments on commercial satellite platforms. The use of these "hosted payloads" (both in the case of the United States and Europe) is currently growing and is projected to grow even more. This combination, which creates a new and profitable business opportunity for private satellite operators, also potentially enables defense agencies to save money by requiring fewer dedicated expensive satellites within its own fleet. But it adds an interesting dimension to the relationship between government and industry both in the United States and Europe and raises numerous questions about the role of private business with security-related space assets. This subject of dual use of communications satellites was addressed in chapter "▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications."

The government is also a regulator. In the United States, the Federal Communications Commission allocates the available spectrum and is the US interface with the International Telecommunications Union. Most other countries have a governmental agency or ministry that oversees the use of radio frequencies including those for satellite communications. Often, this is the entity that participates in the International Telecommunication Union (ITU) processes and international conferences (As noted in chapter "▶ Space Telecommunications Services and Applications" the ITU oversees international spectrum issues and defines different types of satellite services and the associated frequencies for that service. The ITU also oversees the assignment of valuable locations in the geostationary orbit, which is where the largest telecommunications satellites are placed. Constellations that operate in low Earth orbit and medium Earth orbit are also under the purview of the ITU in terms of international regulatory processes.).

Governments fund and perform R&D that supports the technological development of the industry. It is also the province of governments to issue licenses for launching payloads into space (In the United States, the Department of Transportation, Federal Aviation Administration, is responsible for licensing launches. In most other countries this is a ministry that addresses space, but in Europe, the European Aviation Safety Agency (EASA) is assuming authority for some suborbital flights for the emerging industry known as "space tourism" or "space adventures."). These licenses require companies to demonstrate a set level of financial responsibility for their space activities and mandate that they follow detailed safety procedures and take a number of steps to avoid the creation of space debris (All FAA regulations can be accessed at: http://www.faa.gov/about/office_org/headquarters_offices/ast/regulations/).

What has changed over time is that any company can now purchase a launch and obtain access to space for a legitimate business purpose. There is competition for these services and they are not limited to the United States. Over ten nations now have launch capabilities and publically sell these launch services. Furthermore, many nations also now have the ability to manufacture satellites, and strong international competition now exists in the satellite manufacturing arena.

Figure 1 illustrates how widespread these satellite manufacturing capabilities are as demonstrated through orders for new commercial satellites. This data as compiled by the Futron Corporation on behalf of the Satellite Industry Association for 2010 shows for this year United States companies had 54 % of the market share, European companies 27 %, Russia 12 %, and India 8 %. Manufacturing capability also exists

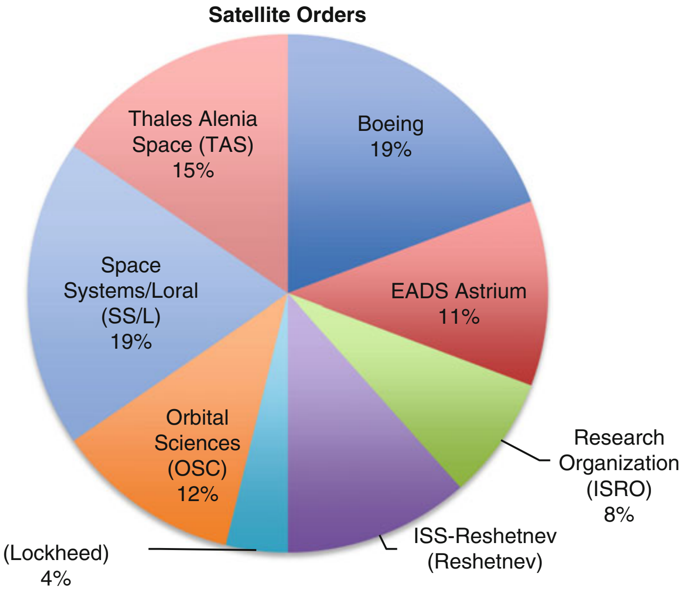


Fig. 1 Commercial GEO satellite orders in 2010 (Futron Corporation: 2010a Year-End Summary)

in China and Japan, and these could expand in international importance in future years.

Initially international telecommunications via satellite were the preserve of a consortium that was set up as an international organization with open-ended membership of the different governments involved. Financing for this entity called Intelsat was provided through the many governments that participated in the consortium. It began with just a dozen members but expanded to over 100 countries. Beginning in the 1980s, this international monopoly approach began to be questioned and there were efforts to create competitive international systems. As the communications satellite sector developed to a more mature, reliable, and developed stage, private companies emerged and increasingly sought to offer GEO-based telecommunications services competitive with Intelsat. In the late 1990s, there were also plans for private LEO telecommunications satellites (These included the proposed Teledesic system for FSS-type offerings and the Globalstar, and Iridium for mobile satellite services, for example).

Although the satellite telecommunications industry has matured and produced many profitable and long-lived systems, it was not without risks and business failures. With the technology sector collapse in 2000, the Teledesic system was never completed, Globalstar had to file for bankruptcy protection; Iridium (owned by Motorola) also went into bankruptcy. Today, nearly 20 years later, large-scale low Earth orbit satellites such as Skybox, OneWeb and SpaceX constellations are being deployed or seriously planned. It is too soon to assess their profitability and long-run success but three things have radically changed in recent years: (1) the ability to develop small satellites with large capabilities and the ability for the satellites in these constellations to communicate with each other, (2) the total cost of manufacturing and launching a small satellite is far less than that for launching a large satellite (even though the cost per kilogram of launching remains very high), and (3) the regulatory structure in nations is adopting to accommodate and incentivize private sector ventures in space.

This is now an extension of the trend that began in the mid-1980s toward the privatization of the satellite communications industry. At that time the industry was also threatened by competition with fiber-optic cables. By 2001, not only had Intelsat been privatized, so had Inmarsat (in 1999) as well as Eutelsat (also in 2001). This was a time of enormous change in the business and economic structure of the satellite communications industry. Along with the privatization was the influx of new companies, the creation of substantial debt equity, the purchase of Comsat by Lockheed, the end of the technology "bubble," and the "dot com" collapse (Mechanick 2011).

This radical and relatively fast change in the way the satellite communications industry was organized and structured created economic benefits as well as added business risks. Bankruptcies (e.g., Iridium, Globalstar, Teledesic, and ICO), new mergers, and international telecommunications conglomerates all became part of a constant flux in companies and market positions. These mergers and acquisitions and the transfer of ownership from governments to equity finance institutions has led to dominance in the international FSS business by Intelsat (headquartered in Bermuda), SES (headquartered in Luxembourg), and Eutelsat (headquartered in Paris). International mobile satellite service is dominated by Inmarsat (headquartered in London), but the reconstituted Iridium and Globalstar still offer key services around the world. The largest industry in terms of revenues is the direct broadcast satellite television services and these are spread around the world. Many smaller satellite companies also continue to operate but with a small part of the global revenues. Most of the satellite commercial business is concentrated in about a dozen satellite operators around the world. There are changes in satellite technology occurring that seem likely to change the economics of satellite communications services moving forward. These are the new high throughput satellites and the new Internet-optimized satellite constellations with a large number of smaller satellites that are now planned to be deployed in low Earth orbit. As mentioned above, it is still too early to assess the impact that these two new trends will ultimately make. (The appendices at the end of the handbook provides information with regard to the many different commercial satellite communications entities providing domestic, regional, or international services.)

Trends in Access to Space and in Manufacturing Satellites

Access to Space

There has been a long-standing expectation in the space community that the cost of access to space (i.e., launch vehicles) will drop exponentially. A major breakthrough in launch technologies has been a goal of numerous unsuccessful R&D programs such as the NASA/Lockheed-Martin X-33 effort that was canceled in 2001 as well as other similar efforts such as the X-34, the X-37, the X-38, and the X-43 (Pelton and Marshall 2006). The corollary of a dramatic drop in cost of access (and prices for launching) is that a floodgate will open and that markets will suddenly develop for new uses of the space environment. Profitable private ventures will flourish and government agencies will be able to purchase inexpensive launches for research satellites and payloads.

This call for inexpensive access illustrates the dramatic way in which economic factors could influence the demand for space. To date, their influence can only be found in the negative hypothesis: that expensive access to space has capped the demand for space activities and created a barrier to entry that is virtually insurmountable for most activities. This is an example of a “technology push” where the emphasis is on the supply side – providing access cheaply. There really is no economic reason to develop the technology unless there is a sufficient market demand to do something of value in space. There may be other reasons – social, political, or security – to go to space often and cheaply which could provide a public-goods stimulus for additional investments in cheap access technologies.

The assumption that cheaper access to space is the key to the future growth of space activities should be subjected to a closer analysis. The results may be very mixed, that is, cheaper access to space will clearly benefit both suppliers of space products and services as well as consumers of those products. But there are current and future space activities that will exist whether or not the cost of access is significantly decreased. Telecommunications is one of these.

First, consider the types of space applications that are not very price sensitive to launch costs. Essentially they require very large up-front investments that are recovered and exceeded relatively quickly over time from project revenues. Typically the services are sold to end users and therefore have a large mass market where the stream of revenues is relatively easily foreseen. In telecommunications, the existing large demand for voice and other transmissions existed before communications satellites were developed. Once a space satellite or facility is launched and placed into the desired orbit, it can have an expected lifetime of 15 years. This is a life expectancy double that of a telecommunications satellite built just a couple of decades ago, which will result in a noticeable decrease in the demand for future satellite manufacturing and launches. Of course other unpredictable factors such as the future demand for telecommunication services and the crowding of the most profitable geostationary orbits and spectrum bands will also affect future launch demand.

Also future in-orbit servicing and re-fueling could also have an impact on the industry. If successful, these types of services will enable the extension of the life of many operating satellites (through refueling techniques) as well as enable the monitoring and repair of some satellites by using advanced maneuvering techniques along with cameras to diagnose problems. Farther into the future, if satellites are designed with docking technologies, true “plug and play” upgrades may be possible and economically feasible.

Numerous business cases have demonstrated that even the high cost of building the satellite system and the high cost of launching it are relatively small percentages of the total revenue over its operational lifetime. Cheaper access to space might mean higher profits for the owner or operator of the system, but today’s profits are sufficiently large that expensive up-front costs have not deterred companies from making these investments. These costs have made it more difficult for satellites to compete with high-capacity and high-cost-efficiency fiber-optic networks.

Second, consider the types of activities have the best opportunity to grow if there is cheap access. The largest opportunity in this respect might be activities that require multiple and regular trips to space and return to Earth. This implies one of three things:

1. That there is something to do in space itself (e.g., manufacturing or transporting people to space and providing for their return)
2. That point-to-point Earth transportation through space at high speeds could be proven to be technologically feasible and safe
3. That a true market for space adventurism or tourism exists and, as above, people will hopefully have something useful to do there

Third, consider the opportunities related to private research and development (R&D) involving space activities. Such activities are presently far too expensive for most companies or universities (The availability of direct government subsidies and other incentives for space research has been the standard practice for many years. With the current budget deficit coupled with the increasing complexity and cost of research equipment, future government aid is not likely to match the demand for this type of research effort). Private capital markets for high-risk R&D funds are often not large enough for a space project, since the cost of a launch is usually included in the cost of a corporate or university research program. In today's environment, an expensive launch can be the deterrent to proceed with the project.

Fourth, consider government programs or project activities that are subject to major budget pressures. There is, of course, a difference between an agency's budget and the project's budget. Many government project managers are advocating cheaper access in order to carry out their projects on a cost-effective basis since their individual funds are constrained. The agency-level huge capital requirements to fund a technology program that might lead to reducing launch costs are outside of the scope and capability of the project offices that are generally most concerned with current operating costs. Even though they are within the same government organizations, the role of a project manager is more similar to that of any final demand consumer.

Fifth, consider that there is a limit to how much launch costs can be lowered. Even if the cost of the launch vehicle is reduced dramatically, a number of other economic factors are not likely to change. Among them are:

- The high costs of launch facilities, payload integration, storage, testing, etc.
- For the foreseeable future, only launches from a coastal location or a very sparsely populated and remote area will be permitted because of safety considerations. This will make it necessary to transport, at considerable cost, payloads a significant distance from the point where the business is located or the product is manufactured to the launch site. The same delay will also exist at the delivery site. These costs will not be reflected in the launch price itself but are real costs in time and transportation to the customer (Sea launch operations or manufacture and launch from the state of California where many manufacturers are located, however, could possibly mitigate these considerations.).
- Delays in launches will frequently occur and add to launch costs.
  - Launch vehicles are complex machines, and mechanical problems will occur with some frequency.
  - Weather will delay launches as it does today for both space launches and even normal airline traffic.
  - Security issues may cause delays.
  - Regulatory issues (safety, financial, environmental, etc.) will also likely continue to be complex and costly.

Payloads bound for space will need to carry very valuable commodities where the speed of delivery is of the highest priority. This means that a launch schedule has to have a high degree of reliability with little variance, otherwise alternatives will be financially more attractive. The time value of money, therefore, becomes a large expense, unrelated to the hardware costs of physically getting to space and returning. Export control issues will continue to dominate launches and will become particularly difficult if landings and relaunch occur in different nations. The demand for launches may never reach a level where economies of scale in manufacturing and launching will be realized. Insurance and liability issues will continue to be problems, particularly since the cost of insurance is related not only to the safety record of launch vehicles but also to the general level of claims payouts of all insurance and reinsurance policies. And, finally there is always the probability of an accident and the risks of suspended operations for a long period of time until the cause of the accident is determined and fixed.

Sixth, and last, one must consider the economics of the cost of developing a new and cheaper launch system. What will the government or private organization pay for the very expensive development of a new system? Who will bear the risks? Will the costs be amortized over the lifetime of the vehicles (and result in higher launch prices) or will a government underwrite the costs? Who benefits from such a system, and will taxpayers be willing to assume the burden of the cost?

The answers to these questions are not just an academic exercise. They go to the root of the linkages between economic and social motivations for future space activities, and how they are answered will shape much of future space development.

It is interesting to note that a 1975 study of the next 200 years in space made an assumption that access to space would be much cheaper by the year 2000 (Brown and Kahn 1977). The study analyzed many scenarios for the future using a variety of different assumptions. One of these assumptions stands out prominently. By extrapolating the rapid trend in technological improvements, most noticeably in integrated chips and computers, and transferring that to launch vehicle improvements during the 1950s and 1960s, the report concluded that this trend of increased productivity and efficiency coupled with rapid decreases in prices would continue. Clearly, it has not. Space access is nearly as expensive today as it was in 1975.

A common thread of the literature on space commercialization is that cheap access is key to the future development of space. Given the above-mentioned parameters and the very difficult hurdles that will have to be overcome in many more areas than simply new launch technologies, this assumption comes into question and thus should be studied much more closely. It very well may be that some important activities will occur if launches are dramatically cheaper. But, history has already demonstrated that profitable space activities, particularly in telecommunications and related services that have large and mature terrestrial markets, are possible even with expensive launches. Likewise, it is possible that new launch systems using tethers, so-called space elevators, rail guns, or nuclear or electrical propulsion (as opposed to chemical propulsion) may be developed in future years, but such alternatives are not near-term prospects, and the implications of such alternative launch systems are not clear at this time.

In addition, telecommunications companies are experimenting with new technologies, cost, and operating structures. Intelsat General, for example, has announced a new capability with its Epic^{ng} satellites due to be launched in 2016. Technological improvements will enable multi-spot beams that will enable a pricing structure to customers that reflects their actual use of mobile equipment rather than the current systems that lease a fixed amount of transponder capacity to a geographic area. Not only will this allow pricing to reflect use and likely result in lower prices to consumers, but it will also free up bandwidth so that the company can better serve surge requirements or geographic shifts in demand.

Comparisons of Productivity in Manufacturing Satellites

Satellites have become more efficient for companies to manufacture by using state-of-the-art production techniques and by a steady demand enabling the realization of economies of scale. Satellites have also become larger, more powerful, and longer-lived (An exception to this is the development of microsatellites and nanosatellites for LEO applications. This discussion is primarily focused on the large GEO telecommunications satellites. In the future, it is possible that some telecommunications applications will be possible with very small satellites).

As noted earlier, the expected lifetime of a new GEO satellite is now more than double what it was 20 years ago. Not all satellites are the same, and the following discussion documents the important differences between manufacturing a government satellite from those made for commercial purposes. Such satellites are different products: commercial satellites are produced relatively quickly and efficiently in response to for-profit pressures, while the government satellites are often pushing the new technology edge and are also subject to government-mandated oversight and audits. Many military satellites have special requirements for radiation hardening, encryption capabilities, and redundancy or protective switches. Often the same companies produce both types of satellites. A comparison, discussed below, of manufacturing productivity and efficiency has documented that the commercial satellites are made faster and more efficiently (Coonce et al. 2010). But the study also highlights a number of important financial and economic characteristics of the manufacturing process and concludes that improvements in the efficiency of producing government satellites would also be possible without major systemic changes.

Three types of satellite systems are compared. First are commercial telecommunications satellites manufactured for private customers. Second are civilian government telecommunications and research scientific satellites. Last are the military satellites for communications and Earth observations.

To compare systems of similar content or classes across agencies, a normalizing metric is necessary. Although some simple metrics, such as cost per kilogram, can be used to compare different systems, such a metric does not provide an assessment of the overall capability and complexity of a system. To assess the relative efficiency of different systems, the Complexity Based Risk Assessment (CoBRA) approach was chosen to assess a “dollar per unit complexity” metric.

The CoBRA complexity index is based on the order of 50 different system parameters, including mass, power, data rate, the number and type of instruments, solar array size, etc., and is used to determine the relative ranking of a system

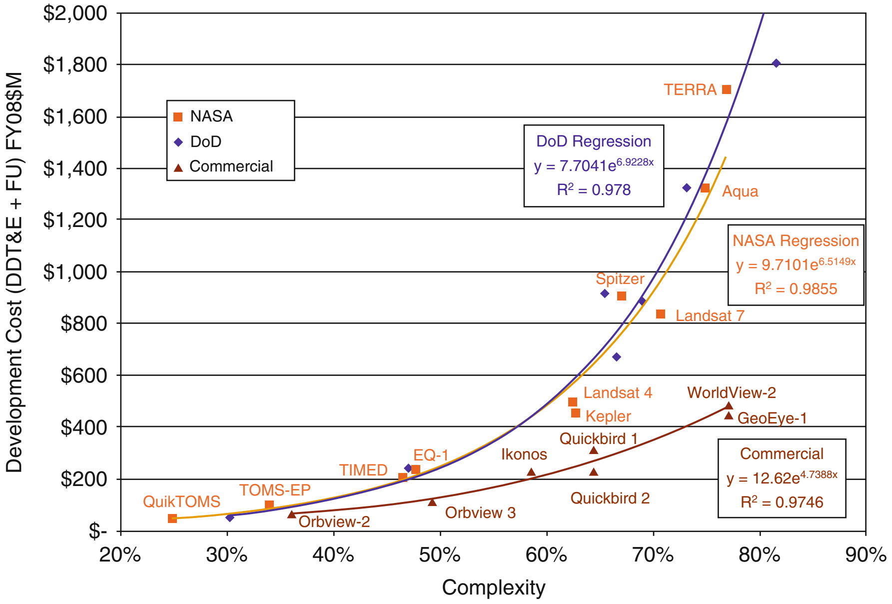


Fig. 2 Efficiencies in DoD and NASA production are similar but less than commercial

compared to over 120 other satellites. The measure is on a scale of 0–1.0, with the low values having the least capability relative to all of the spacecraft in the database, and a high value representing the most capable system (Bitten et al. 2005).

A regression of complexity versus cost for different customers reveals insights into relative efficiencies. Figure 2, shows the plot of a regression of complexity versus development cost for the DoD, NASA, and commercial imaging. Figure 2 above shows a substantially higher cost for a given level of complexity, relative to similar commercial systems.

A potential explanation for such a trend can be shown when looking at a similar regression for the same systems relative to the time schedule of production as shown in Figure 3 above. The regression for NASA and DoD missions are similar to the cost regression shown previously where schedule increases as the complexity increases. This makes intuitive sense as the development cost typically increases as schedule increases and both are greater with higher levels of complexity (Fig. 3).

This trend, however, is not the same for commercial imaging systems. As shown, the regression for schedule relative to the increasing complexity for commercial systems is similar regardless of the level of complexity.

Commercial systems show cheaper costs and shorter manufacturing times because they rely on the same payload and spacecraft bus for each successive satellite. The commercial satellite manufacturers tend to develop “platforms” that can be used with a series of progressively larger satellites. This is in some ways comparable to the “platforms” that automobile manufacturers now use. Establishing

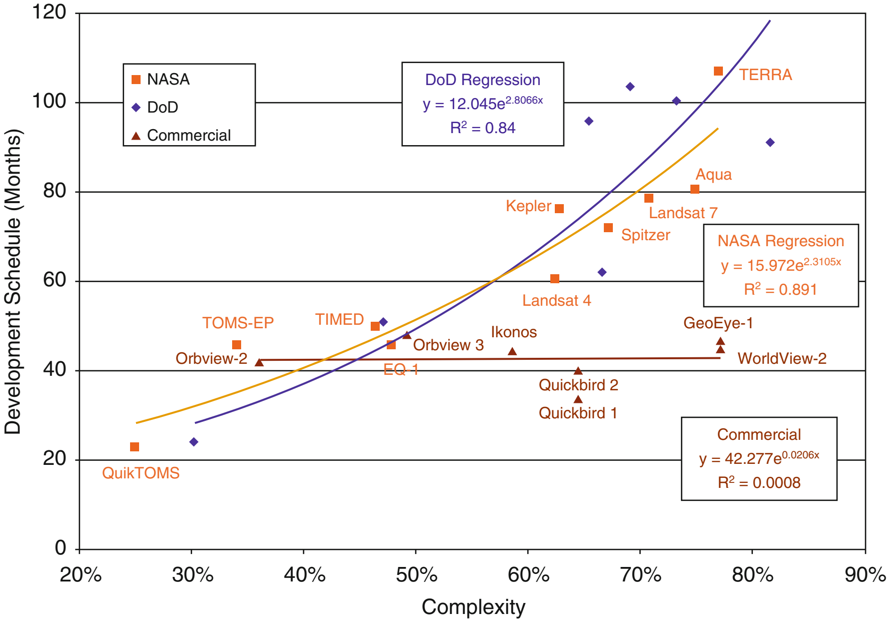


Fig. 3 Schedule increases with complexity for government systems

a long-term commitment and partnership with industry providers enables this evolutionary approach where teams can build upon past experience to become more efficient for more complex future systems (An example of such a system is the QuickBird, WorldView 1, and WorldView 2 evolutionary approaches that migrated a 0.6 m imaging system on a standard bus [BCP 2000 for QuickBird] to a more capable 0.6 m imaging system on a larger, standard bus from the same provider [BCP 5000 for WorldView 1] to a 1.1 m imaging system [WorldView 2] on the same bus as used for WorldView 1. This evolutionary approach minimized risk and maximized team efficiencies).

The evolutionary approach was developed because commercial satellites are able to take advantage of a number of options not usually applicable to government satellites, including:

- A fixed-cost tight time schedule
- A much less complex design and instrumentation than government research and development satellites
- An evolutionary approach to cutting-edge technology
- A ground system that is an integral component
- An unambiguous set of technical requirements which demand lower skill levels
- A payment schedule based on progress and in some cases incentives for efficient, timely, and reliable performance

The analysis in the study as well as comments from industry and government reviewers had the following conclusions:

- NASA and DoD spacecraft require about twice the systems engineering staff relative to what is done on commercial ventures.
- NASA and DoD spacecraft require about 1.5 times longer for assembly, integration, and testing than commercial ventures.
- On a cost per pound basis, DoD projects cost about four times as much as commercial projects, while NASA projects cost about twice as much as commercial projects.
- On hours per drawing basis, DoD projects require more than three times as much time as commercial projects while NASA projects two times as much time.
- At least one company ranks productivity as follows (highest to lowest): commercial, performance-based government contracts, cost-plus-award-fee government contracts, and DoD classified projects.
- Government-sponsored special communications satellites take more than twice as long to manufacture than comparable commercial communications satellites (over 5 years, compared to 2.5 years) and are even less efficient to manufacture because:
  - They have more reporting requirements that need formal approval.
  - They have to undergo more testing.
  - They have more on-site government and other personnel.
  - They are designed to carry out more functions.

Manufacturers of commercial satellites use proven component parts. They assemble and test them rather than invest in new technological development. They also noted the use of standardized processes and that customers do not change requirements once a contract is signed. This allows them to produce their satellites within a 24–36-month timeframe, which, as mentioned above, is approximately half the amount of time required for government projects. They also noted that commercial fixed-price contracts are easier to finance and administer than cost-plus government contracts since contract and payment schedules are negotiated between client and customer without restrictions imposed by complex government procurement regulations. Up-front and progress payments are scheduled according to milestones that are generated to encourage schedule compliance. Commercial entities also purchase risk insurance because they have incentives to deliver on time. They noted that government programs impose far more technical reviews, changes in design, and oversight than commercial customers.

Another conclusion is that there is not that much difference in productivity among different government programs (although some of the data presented suggest that unclassified projects are more “productive” [in terms of cost efficiency] than the classified ones). Table 1 in chapter “▶ Introduction to Satellite Navigation Systems,” summarizes the full set of differences in producing satellites for the government compared to private customers. Although this study only examined the US experience, there is a reasonable expectation that similar results would be found if commercial, governmental, and military satellite projects were compared in other regions such as Europe (Table 1).

Table 1 Commercial satellite and government-sponsored projects compared

Category	Commercial	Government
Development trend	Evolutionary	Revolutionary
Production	Standardization and reuse of building blocks. Build multiple units	Unique designs. Build one of a kind
Requirement definition	Well understood before project start	Not well understood at project start
Requirement stability	Stable	Unstable
Stakeholders	Single customer/stakeholder	Many stakeholders
Performance specification	Specifies only performance requirements	Specifies performance requirements and methods
Design incentive	Profit driven	Science driven
Design approach	Satellite buses viewed as product line and are a known entity	Changes, especially after the start of the project, drives the design of the satellite bus
Cost and schedule	Based on known similar historical data (buy mode)	Cost and schedule estimates are optimistic (sell mode)
Funding stability	Stable	Potential annual changes (often a result of budget pressures)
Portfolio management	If a project gets into trouble, it typically gets canceled	Projects allowed to continue and usually cause collateral damage to the portfolio (a very inefficient outcome)
Procurement process	Streamlined	Long and complicated
Contract type	Incentives for early delivery and late delivery penalties	Cost-plus-type contracts
Oversight and reporting	Minimal oversight of subcontractors	Extensive oversight of primes and subcontractors
Test philosophy	Deletes non-value-added processes (profit driven)	Tends to avoid seeking waivers. Success valued on success of mission, not cost or schedule overruns

In short, although both commercial satellites and government satellites share many technologies, commercial products are built and operated with profit as the objective. Government satellites, and in particular NASA satellites, are for research and development purposes and are often typically designed as first of a kind, using cutting-edge technology and with knowledge or support of other government programs as the goal.

Because of the above reasons, there is no a priori reason to conclude that a comparison of commercial satellites and government satellites, even though they may be designed for similar purposes, should or will result in equivalent costs and performance.

Conclusion

The economics of communications satellite has evolved from a government-controlled, privately operated system to a heavily regulated, oligopolistic, somewhat competitive essential part of our economic infrastructure. From the early voice and data transmissions, there are now a wide variety of satellite services ranging from direct broadcast television to the rapid transmission of data and information for the global financial network.

The space system continues to be expensive and risky. Only relatively large companies can effectively compete for manufacturing satellites, launch services, and operations. The large number of mergers over the past 20 years is strong evidence of this, coupled with the emergence of only a few dominant firms. However, terrestrial services using the satellite-based relay and transmission are spread over many different sectors, many different companies, and many different end users. It is truly competitive and is the fastest growing part of the satellite communications business.

Also evidenced by the maturity of the industry is the international and global dimension of the industry. One of the main advantages of using satellites for communications centers on their global or at least broad regional coverage and their ability to broadcast information simultaneously from one point to many points on Earth. In the 1960s, the United States had developed the technology and had the ability to launch these satellites. This was matched only by the Soviet Union, mainly by their launch capabilities, not their advanced technology in telecommunications. Because of the strategic importance of this capability, the United States dominated the industry. Today that has changed. The United States still has many capabilities in terms of advanced telecommunications satellite technology, but very capable and competitive systems, particularly for civilian purposes, can be bought from commercial suppliers in many parts of the world and launched by many other countries as can be seen in Fig. 1 in chapter "▶ Introduction to Satellite Navigation Systems" and in the Appendices to this handbook.

From an economic perspective, there are a number of challenges facing the future of the industry. First, there are competing forms of transmission such as fiber-optic cables that did not exist when satellites were first deployed. Second, the available spectrum is limited and scarce.

Allocating spectrum for communications purposes is both an international and diplomatic exercise as well as an economic one. Nationally, it is handled differently in each nation, some using sophisticated economic means such as auctions and others using more political and less market-driven allocation schemes.

Space itself is more crowded with human-made objects. Some are controllable and working and others are older abandoned satellites or debris. These represent potential hazards to orbiting satellites. Furthermore, there are still no agreed-upon effective means of controlling the growth of the debris or ensuring that there will be a sustainable and secure future for satellite operations.

One of the more daunting and important issue facing the satellite communications industry is this increasing risk of serious damage and consequent service interruptions and the liability due to a collision due to debris or a derelict satellite. There are costs associated with developing better space sustainability. First, hardening a spacecraft when it is being built to minimize damage is expensive and adds weight to the launch payload, which entails additional expense. Second, while the spacecraft is in orbit, fuel must be reserved for additional maneuvers to avoid a collision with oncoming uncontrolled objects. Third, additional fuel must be reserved for end-of-life deorbiting or boosting to a graveyard orbit (Or, in the future, servicing satellites may provide alternatives for end-of-life maneuvers. But at present, the cost of these still-to-be-developed services is undetermined. These types of in-space services will also face major regulatory issues, and the combination of expense and administrative hurdles may not produce a viable economic business). Fourth, additional personnel must be dedicated to minimizing debris during manufacture, operations, and possibly even when the satellite is no longer in use.

The manufacturing and operating firms would largely be the entities to incur these costs. In addition, governments now face monitoring, mitigation, regulatory costs associated with satellite communications and satellite applications. In the future, if technology permits, they may face cleanup costs. These can range from relatively trivial routine monitoring to very expensive in-orbit activities. The funds for these activities may come from a combination of governmental funds, insurance companies, and the owner/operator firm's themselves.

Although economics – the allocation of resources and the opportunity to make a profit from an investment in satellite communications businesses – will drive many aspects of this business, the involvement of governments will continue to add cost, risk, and political dimensions to any private sector activity in space. However, government's involvement has diminished somewhat as the industry has matured, and current trends indicate that the industry will continue to grow rapidly, and the degree of influence governments have over private satellite communications activities may thus also continue to diminish.

It is likely that current and future developments in this industry will be apparent on both the supply side and the demand side. On the supply side, private firms will develop new technologies and operating systems that will reduce costs. It is likely that prices to consumers will also decrease. Smaller but more numerous smaller satellites will be developed that will also contribute to cost efficiencies, particularly for launching into space.

On the demand side, both the expansion of markets into developing nations and the growth and merging of the earth observations, telecommunications, and navigation satellite services into the overall information sector, coupled with the advent of "big data" systems, will increase demand for space-based services as well as provide both government and private customers measurable improvements in productivity and in services.

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ History of Satellite Communications
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Satellite Applications Handbook: The Complete Guide to Satellite Communications, Remote Sensing, Navigation, and Meteorology
- ▶ Satellite Communications Overview
- ▶ Satellite Communications Video Markets: Dynamics and Trends
- ▶ Satellite Orbits for Communications Satellites
- ▶ Space Telecommunications Services and Applications

References

R.E. Bitten, D.A. Bearden, D.L. Emmons, A quantitative assessment of complexity, cost, and schedule: achieving a balanced approach for program success. in *Sixth IAA International Conference on Low-Cost Planetary Missions*, Kyoto, 11–13 Oct 2005

W.M. Brown, H. Kahn, in *Long-Term Prospects for Developments in Space* (A Scenario Approach) (Hudson Institute, New York, 1977). NASW-2924, 30 Oct 1977

T. Coonce, J. Hamaker, H. Hertzfeld, R. Bitten, NASA productivity. J. Cost Anal. Parametric. **3**(1), 59–73 (2010). Society of Cost Estimating and Analysia – International Society of Parametric Analysis

Futron Corporation, Futron corporation state of the satellite industry report 2010 (2010a), sponsored by the Satellite Industries Association, http://www.futron.com/resources.xml#tabs-4. Accessed 29 May 2011

Futron Corporation, Futron forecast of global satellite services demand – overview (2010b), http://www.futron.com/resources.xml#tabs-4. Accessed 29 May 2011

Futron Corporation, Telecommunications report (2010c), http://www.futron.com/resources.xml#tabs-4. Accessed 29 May 2011

M.J. Mechanick, *The Politics of the Establishment and the Eventual Privatization of the Three Major International Satellite Organizations.* (White & Case, LLP, 2011), (unpublished manuscript)

J.N. Pelton, P. Marshall, *NASA's Unsuccessful X-Projects, Space Exploration and Astronaut Safety* (AIAA, Reston, 2006), pp. 149–178

Satellite Industries Association: Satellites 101, http://www.sia.org/satellites.html. Accessed 20 May 2011

Further Reading

D. Cavosa, Satellite Industries Association, COMSTAC presentation (2004), http://www.sia.org/present.html. Accessed 29 May 2011

European Satellite Operators Association, Economics of satellites (2010), http://www.esoa.net/Economics_of_satellites.htm. Accessed June 2011

B. Henoch, Satellite technology basics, March (2007), http://www.sia.org/present.html. Accessed 29 May 2011

Z. Szajnfarber, M. Stringfellow, A. Weigel, The impact of customer–contractor interactions on spacecraft innovation: insights from communication satellite history. Acta Astronom. **67**, 1306–1317 (2010)

Satellite Communications and Space Telecommunication Frequencies

Michel Bousquet

Contents

M. Bousquet (✉)
Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), Toulouse, France
e-mail: michel.bousquet@isae.fr

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_13

Abstract

Radio frequencies allow information to be transmitted over large distances by radio waves. The essential element to high-quality satellite communications is the assignment of radio-frequency spectrum to various types of services. Only a limited amount of such spectra is assigned to Earth-space radio links, and thus the available bandwidth must be used with a high degree of efficiency. There are many technical elements associated with the efficient use of RF spectra for satellite communications and navigation, and these elements are addressed in some detail in this chapter.

The basic properties of electromagnetic waves are first discussed, together with an overview of the basic electromagnetic phenomena such as reflection, refraction, polarization, diffraction, and absorption useful to define how radio waves travel in free space and in atmosphere. The basic parameters used to characterize the antennas responsible for generating and receiving these waves are introduced.

A survey of the propagation impairments (gas and rain attenuation, scintillation, etc.) due to the nonionized lower layers of the atmosphere from Ku- to Ka- and V-bands is presented. On the other hand, radio waves of Global Navigation Satellite Systems (GNSS) interact with the free electrons of the upper atmosphere ionized layers on their path to the receiver, changing their speed and direction of travel.

Keywords

Attenuation effects • Antenna characteristics • Ionospheric effects • Polarization and depolarization effects • Path loss • Propogation mechanisms • Radio frequency • Radio regulation • Scintillation • Space radiocommunications services • Spectrum • Spectrum allocation • Tropospheric effects • Wave characteristics • Wavelength

Introduction

Radio frequencies allow information (images, sound, and data) to be transmitted over large distances by radio waves. They are the basis of satellite communications. They are a portion of the "electromagnetic spectrum" which is the term used to describe the range of possible frequencies of electromagnetic radiations. Indeed, electromagnetic radiation, a form of energy exhibiting wavelike behavior as it travels through space, is classified according to the frequency of its wave. This elemental consideration of the physics of electromagnetic phenomena is discussed in detail in chapter "▶ Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing" by Prof. Rycroft. Just a few of the key concepts are reiterated here to explain the exploitation of electromagnetic spectra explicitly used for space communications.

Radio Waves

As discussed in the next section, any variation in time of a charge or of a magnetic moment creates coupled electric and magnetic fields characterizing an electromagnetic field. The variation in time of the electromagnetic field is the image of the time variation of the generating sources. Moreover, the electromagnetic field exhibits a space variation at the same pace: the electromagnetic wave propagates without requiring any physical support and carries energy. If the sources are periodically moving in time, i.e., oscillating at a given frequency, the electromagnetic field oscillates at the same frequency. In homogeneous free space, the electromagnetic field is also periodic in space, and the electromagnetic wave propagates with a constant velocity, the speed of light in vacuum. The wavelength is the period in space of the electromagnetic field, that is, the replica of the period in time domain (i.e., proportional to the inverse of frequency), and is proportional to the wave velocity. Thus, either the frequency or the wavelength characterizes an electromagnetic wave.

Possible values of the wave frequency constitute a continuum called "electromagnetic spectrum." The electromagnetic spectrum, in order of increasing frequency and decreasing wavelength, includes radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays, and gamma rays (Fig. 1). Although, the propagation phenomenon is independent of the value of the frequency, the ability of a material to interact, namely, absorb the energy of an electromagnetic radiation, is strongly frequency dependent. Depending on frequency, some materials could be "seen" by the wave that is reflected or absorbed; others are "transparent" to the wave if no interaction occurs. This is the case with atmosphere that is transparent (or with limited attenuation) or opaque to the electromagnetic waves depending on wavelength.

Radio waves are a type of electromagnetic radiation with wavelengths in the electromagnetic spectrum longer than infrared light. More specifically, the radio spectrum includes radio waves with frequencies between 3 kHz and 300 GHz, corresponding to radio wavelengths from thousands of kilometers to under 1 mm. Naturally occurring radio waves are made by lightning or by astronomical objects. Artificially generated radio waves are used for fixed and mobile radio terrestrial and satellite communication, broadcasting, radar and navigation systems, etc. When considering terrestrial communications, different frequencies of radio waves have different propagation characteristics in the Earth's atmosphere; long waves may cover a part of the Earth very consistently, and shorter waves can reflect off the ionosphere and travel around the world. Earth-space links (satellite communication and navigation) are using shorter wavelengths (centimeter to tens of centimeters) that bend or reflect very little and travel mainly on a line of sight.

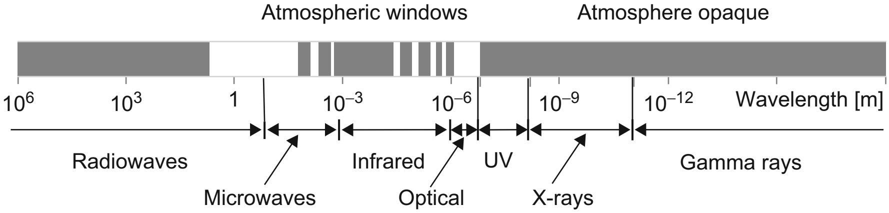


Fig. 1 Electromagnetic spectrum

Although valid in all the electromagnetic spectrum, electromagnetic field theory has been established by Maxwell in the radio-frequency range (Maxwell et al. 1865). Maxwell noticed wavelike properties of light and similarities in electrical and magnetic observations and proposed equations that described light waves and radio waves as electromagnetic waves that travel in space. In 1887, Heinrich Hertz demonstrated the reality of Maxwell's electromagnetic waves by experimentally generating radio waves in his laboratory. Many inventions followed, making practical the use of radio waves to transfer information through space.

The study of electromagnetic phenomena such as reflection, refraction, polarization, diffraction, and absorption is of critical importance in the study of how radio waves move in free space and over the surface of the Earth. This is the rationale for the overview on these phenomena presented in section "Electromagnetic Waves."

Need for Radio Regulations

Radio spectrum is an essential resource underpinning one of world's most dynamic sectors: wireless communications. As well as telecommunications, wireless technologies support services in areas as diverse as transport, security, and environmental protection. But the spectrum is a finite resource, so its allocation requires effective and efficient coordination at global level. Radio regulations are necessary to ensure an efficient and economical use of the radio-frequency spectrum by all communications systems, both terrestrial and satellite. While so doing, the sovereign right of each state to regulate its telecommunication must be preserved. It is the role of the International Telecommunication Union (ITU) to promote, coordinate, and harmonize the efforts of its members to fulfill these possibly conflicting objectives. These issues are discussed in greater details in the radio regulations that refer to the various space radiocommunication services, defined as transmission and/or reception of radio waves for specific telecommunication applications. These are listed in Table 1 in chapter "▶ Space Telecommunications Services and Applications."

Frequency bands are allocated to the above radiocommunication services to allow compatible use. The allocated bands can be either exclusive for a given service or shared among several services. Regarding allocations, the world is divided into three regions. Frequency allocations are revised regularly at the World Administrative Radio Conference (WARC).

For example, the fixed satellite service (FSS) makes use of the following bands:

(a) Around 6 GHz for the uplink and around 4 GHz for the downlink (systems described as 6/4 GHz or C-band). These bands are occupied by the oldest systems (such as INTELSAT, American domestic systems, etc.) and tend to be saturated.

(b) Around 8 GHz for the uplink and around 7 GHz for the downlink (systems described as 8/7 GHz or X-band). These bands are reserved, by agreement between administrations, for government use.

Table 1 Frequency allocations

Radiocommunication service	Typical frequency bands for uplink/downlink	Usual terminology
Fixed satellite service (FSS)	6/4 GHz	C-band
	8/7 GHz	X-band
	14/12–11 GHz	Ku-band
	30/20 GHz	Ka-band
	50/40 GHz	V-band
Mobile satellite service (MSS)	1.6/1.5 GHz	L-band
	30/20 GHz	Ka-band
Broadcasting satellite service (BSS)	2/2.2 GHz	S-band
	12 GHz	Ku-band
	2.6/2.5 GHz	S-band
Radionavigation satellite service (RNSS)	1.164–1.3 GHz (down)	Lower L-band
	1.559–1.617 GHz (down)	Upper L-band
	5.000–5.010 GHz (up)	C-band
	5.010–5.030 GHz (down)	C-band

(c) Around 14 GHz for the uplink and around 12 GHz for the downlink (systems described as 14/12 GHz or Ku-band). This corresponds to current operational developments (such as EUTELSAT, SES, etc.).
(d) Around 30 GHz for the uplink and around 20 GHz for the downlink (systems described as 30/20 GHz or Ka-band). These bands are raising interest due to large available bandwidth and little interference due to present rather limited use. This corresponds to the new developments of high-capacity systems for Internet access (such as VIASAT, KA-SAT, etc.).

The bands above 30 GHz (Q- and V-band, possibly W) will be used eventually in accordance with developing requirements and technology.

The mobile satellite service (MSS) makes use of the following bands:

(a) VHF (very high frequency, 137–138 MHz downlink and 148–150 MHz uplink) and UHF (ultrahigh frequency, 400–401 MHz downlink and 454–460 MHz uplink). These bands are for non-geostationary systems only.
(b) About 1.6 GHz for uplinks and 1.5 GHz for downlinks, mostly used by geostationary systems such as INMARSAT, and 1,610–1,626.5 MHz for the uplink of non-geostationary systems such as GLOBALSTAR.
(c) About 2.2 GHz for downlinks and 2 GHz for uplinks for the satellite component of IMT2000 (International Mobile Telecommunications).
(d) About 2.6 GHz for uplinks and 2.5 GHz for downlinks.
(e) Frequency bands have also been allocated at higher frequencies such as Ka-band.

The broadcasting satellite service (BSS) makes use of downlinks at about 12 GHz. The uplink is operated in the FSS bands and is called a *feeder link*.

The radio navigation satellite service (RNSS) makes use of frequencies in L-band, mainly from space to Earth (downlink). Before WARC-2000, RNSS allocations were divided in two parts:

(a) Lower L-band, from 1,215 to 1,260 MHZ, called L2
(b) Upper L-band, from 1,559 to 1,610 MHz, called L1

These bands have been split between the GPS and GLONASS systems, for instance, 1,563–1,587 MHz for GPS and 1,597–1,617 MHz for GLONASS.

At the WARC-2000, new frequency bands have been added to accommodate the needs of new GNSS (global navigation satellite systems) such as GALILEO.

(a) 1,164–1,215 MHZ to be shared between new GPS signals (called L5 from 1,164 to 1,191.795 MHz, carrier frequency 1,176.45 MHz), GALILEO signals (called E5 from 1,164 to 1,215 MHz, carrier frequency 1,191.795 MHz, further subdivided into E5a and E5b with carrier frequencies 176.45 MHz and 1,207.14 MHz, respectively), new GLONASS signals (called L3 from 1,164 to 1,215 MHz), and others (COMPASS B2, China)
(b) 1,260–1,300 MHz called E6, to be shared between GALILEO signals (called E6, with carrier frequency 1,278.75 MHz) and others (COMPASS B3)
(c) 5,010–5,030 MHz at C-band

In addition to frequency bands used by the satellites to transmit navigation signals (downlink), bands 1,300–1,350 MHz and 5,000–5,010 MHz could be used to uplink dedicated signals to the satellites.

It should be noted that the new proposed GNSS systems are planning to share the upper L1-band (called E1 1,559–1,591 MHz with Galileo, B1 with COMPASS, etc.) (Table 1).

The following table summarizes the above discussion.

Nomenclature of the Frequency and Wavelength Bands

The ITU Radiocommunication Assembly, in the Recommendation ITU-R V.431, has defined a nomenclature to be used for the description of frequency and wavelength bands, given in Table 2.

Certain frequency bands are sometimes designated by letter other than the symbols and abbreviations recommended in the above Table. The symbols in question consist of capital letters which may be accompanied by an index (usually a small letter). There is at present no standard correspondence between the letters and the frequency bands concerned, and the same letter may be used to designate a number of different bands. For information, letter designations used by some authors in the field of space communications are indicated in Table 1.

Table 2 Nomenclature of frequency and wavelength bands

Band number	Symbols	Frequency range (lower limit exclusive, upper limit inclusive)	Corresponding metric subdivision
3	ULF	300–3,000 Hz	Hectokilometric waves
4	VLF	3–30 kHz	Myriametric waves
5	LF	30–300 kHz	Kilometric waves
6	MF	300–3,000 kHz	Hectometric waves
7	HF	3–30 MHz	Decametric waves
8	VHF	30–300 MHz	Metric waves
9	UHF	300–3,000 MHz	Decimetric waves
10	SHF	3–30 GHz	Centimetric waves
11	EHF	30–300 GHz	Millimetric waves

Electromagnetic Waves

The basic properties of electromagnetic waves traveling in free space and in atmosphere considered as a uniform medium are provided in the following section and can be specified by a small set of descriptive parameters used to define the behavior of waves.

Radio-Frequency Wave Characteristics and Maxwell's Equations

The *Maxwell equations* (Maxwell et al. 1865) specify the relationships between the variations of the vector electric field E and the vector magnetic field H in time and space within a medium which characterize propagating electromagnetic waves.

Both Faraday's and Ampere's laws are included in Maxwell's equations, but the decisive contribution of Maxwell was to complement Faraday's law in linking the time variation of the electric induction to the variation in space of the magnetic field.

Maxwell definitely establishes that any time variation in electric induction results in the apparition of time-variant magnetic field and conversely any time variation in magnetic induction results in time-varying electric field.

An oscillating electric field produces a magnetic field, which itself oscillates to recreate an electric field and so on. This phenomenon is represented by the so-called curl equations in the set of Maxwell's equations. Indeed Maxwell describes the way electric and magnetic field lines spread out (<< *diverge* >>) and circle around (<< *curl* >>). Without entering in a rigorous derivation and too many details, the Maxwell equations could be explained as follows:

An electric field is produced by a time-varying magnetic field.
A magnetic field is produced by a time-varying electric field or by a current.

The interplay between the two fields stores energy and hence carries power.

The E-field strength is measured in volts per meter and is generated by either a time-varying magnetic field or by a free charge.

The H field is measured in amperes per meter and is generated by either a time-varying electric field or by a current.

The properties of the propagating medium are incorporated in Maxwell equations through the constants ε, the *permittivity* of the medium, and μ, the *permeability* of the medium, which relate electric and magnetic fields $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{H}}$ to electric and magnetic inductions $\vec{\boldsymbol{D}}$ and $\vec{\boldsymbol{B}}$:

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \text{ and } \boldsymbol{B} = \mu \boldsymbol{H}$$

The scalars ε and μ (tensors in the general case of anisotropic media) model the electric and magnetic properties of the medium as it reacts to the field-producing induction. The permittivity of the medium is expressed in Farad per meter and the permeability of the medium in Henry per meter.

These constants are expressed relative to the values $\mu_0 = 4\pi 10^{-7}\mathrm{Hm}^{-1}$ and $\varepsilon_0 = 8.854\ 10^{-12} = 10^{-9}/36\,\pi\ \mathrm{Fm}^{-1}$ in free space as

$$\mu = \mu_r \mu_0 \text{ and } \varepsilon = \varepsilon_r \varepsilon_0$$

where

μ_r and ε_r are the relative values ($\varepsilon_r = \mu_r = 1$ in free space).

Strictly speaking, *free space* provides a reference to vacuum, but the same values can be used as good approximations when performance is compared to dry air at “typical” temperature and pressure.

Maxwell’s equations are first-order evolution equations in time and space, symmetrically coupled in an open homogenous space. This symmetry, in free space, allows deriving, for both electric and magnetic fields, uncoupled second-order harmonic equations or wave equations which sustain independent oscillatory or harmonic solutions in both space and time. In complex notation, electric and magnetic field amplitudes read as

$$E = \overline{E} e^{j\omega t} e^{j2\pi\frac{z}{\lambda}} \text{and} H = \overline{H} e^{j\omega t} e^{j2\pi\frac{z}{\lambda}}$$

where E and H are the complex amplitudes of the field. The pulsation ω is related to the frequency f of the electromagnetic field by

$$\omega = 2\pi f$$

And λ, the period in space, is the wavelength that is related to the frequency by

$$\lambda = \frac{v}{f}$$

With v the phase velocity which takes the value of the speed of light in the considered medium given by

$$v = \frac{1}{\sqrt{\mu\varepsilon}} = \frac{c}{\sqrt{\mu_r\varepsilon_r}}$$

In free space, the speed of the light is given by

$$c = \frac{1}{\sqrt{\mu_0\varepsilon_0}} = 3.10^8 m.s^{-1}$$

It is common practice to introduce the *wave number* k or the *propagation constant* $\gamma = jk$, related to the wavelength by $k = \frac{2\pi}{\lambda} = \omega\sqrt{\mu\varepsilon}$.

When considering the range of radio frequencies for satellite communications, say from around 300 MHz to 30 GHz, the free-space wavelength varies from 1 m to 1 cm.

The properties of the wave are characterized by a given value of amplitude, frequency, and phase. When one (or a combination) of these parameters is varied according to the amplitude of an information signal, this allows information to be carried in the wave between its source and destination. This process, called carrier *modulation*, is a key concept in satellite communications. Modulation is discussed in some detail in chapter "▶ Satellite Radio Communications Fundamentals and Link Budgets" by Daniel Glover.

Field Structure and Plane Wave

Far from the source and from any obstacle, as in free space, fields of the radio waves may be considered having a spherical structure and dispersing energy in the radial direction.

The amplitudes of the fields $\overline{E}$ and $\overline{H}$ in the above expressions of the fields are decaying as function of the inverse of the distance r from the source:

$$\overline{E} = \overline{E}(r) \propto \frac{E_0}{r} \text{ and } \overline{H} = \overline{H}(r) \propto \frac{H_0}{r}$$

This radial variation translates into a bounded value of the integration of the power density over the three-dimensional space, which is equal to the power delivered by the source.

The *power density* at a distance r is related to the amplitude of the electric and magnetic field components $E_{0\perp}$ and $H_{0\perp}$ on the plane tangent to the sphere of radius

r centered on the sources, which represents the wave front. In the plane of the wave front, electric and magnetic field components are perpendicular to each other.

The amplitudes $E_{0\perp}$ and $H_{0\perp}$ are related through the *wave impedance* ζ such as

$$E_{0\perp} = \zeta H_{0\perp}$$

The wave impedance is a function of the permittivity and the permeability of the medium:

$$\zeta = \sqrt{\frac{\mu}{\varepsilon}}.$$

In vacuum, $\zeta = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 120\pi = 377\boldsymbol{\Omega}$

As far as ε_r and μ_r are real and positive constants, the medium is said to be *lossless*, and the amplitudes $E_{0\perp}$ and $H_{0\perp}$ are in phase. The wave front is "locally plane" (*plane wave*).

When the wave interacts with the medium (*lossy media*), energy is removed from the wave and converted to heat. Absorption by the media is taken into account considering a complex permittivity and/or a complex permeability, the imaginary part of which yields to an imaginary part in the wave number, introducing an exponential attenuation α in the spatial variation of fields:

$$E = \overline{E}e^{-\alpha z}e^{j\omega t}e^{j2\pi\frac{z}{\lambda}} \text{ and } H = \overline{H}e{-}^{\alpha z}e^{j\omega t}e^{j2\pi\frac{z}{\lambda}}$$

It should be noted that, as the wave impedance is complex, the amplitudes $E_{0\perp}$ and $H_{0\perp}$ are no longer in phase.

The constant α is known as the *attenuation constant*, with units of per meter (m^{-1}), which depends on the permeability and permittivity of the medium, the frequency of the wave, and the *conductivity* of the medium, σ, measured in Siemens or ohm per meter ($\Omega\ \mathrm{m}^{-1}$). Together σ, μ, and ε are known as the *constitutive parameters* of the medium. In nonlinear media, ε_r and μ_r depend on the field's intensity.

Poynting Vector and Power Density

The Poynting vector $\boldsymbol{S}$ measured in watts per square meter describes the magnitude and direction of the power flow carried by the wave per unit of surface perpendicular to the direction of propagation z, i.e., the *power density* of the wave. The instantaneous power density is represented by the instantaneous Poynting vector $\boldsymbol{S}$ defined as the vector product between electric and magnetic field:

$$\boldsymbol{S}(t) = \boldsymbol{E}(t) \wedge \boldsymbol{H}(t)$$

The total power leaving a surface A is given by the flux of the Poynting vector through A. It should be noted that the power flux is carried out by the field components tangential to the surface. Field components normal to the surface do not contribute to this flux.

In the case of harmonic variation in time domain, the fields may be written as $\boldsymbol{E} = \boldsymbol{E}(r)e^{j\omega t}$ and $\boldsymbol{H} = \boldsymbol{H}(r)e^{j\omega t}$, and therefore the expression of the instantaneous Poynting vector S reads as

$$\boldsymbol{S}(t) = \frac{1}{2}\left[\Re e(\boldsymbol{E}(r) \wedge \boldsymbol{H} * (r)) + \Re e(\boldsymbol{E}(r) \wedge \boldsymbol{H}^{*}(r))e^{2j\omega t}\right]$$

where the asterisk (*) denotes the complex conjugate.

In practice, the time average of the power flow over one period is considered:

$$S_{av} = \frac{1}{2}[\Re e(\boldsymbol{E}(r) \wedge \boldsymbol{H}(r))]$$

As mentioned above, only the field component in the plane normal to the propagation direction z contributes to the power flow. Introducing the notation of the transverse field of the locally plane waves introduced in the previous session, the average power density could be written as

$$S_{av} = \frac{1}{2}\left[\Re e\left(E_{\perp}(r) \wedge H_{\perp}^{*}(r)\right)\right] = \frac{1}{2}|E_{\perp}|^{2}\Re e(\zeta)^{-1}$$

$\boldsymbol{E}$, $\boldsymbol{H}$, and $\boldsymbol{S}_{ave}$ form a right-hand set, i.e., $\boldsymbol{S}_{ave}$ is in the direction of movement of a right-hand corkscrew, turned from the $\boldsymbol{E}$-direction to the $\boldsymbol{H}$-direction.

Wave Polarization

The time variation of the amplitude and the orientation of the extremity of the transverse component of the electric field define the *polarization* of the wave. In the more general case, the curves traced by the extremity of this field component along the time at a given position or, similarly, along the direction of propagation at a given time are ellipses. The wave is said to be *elliptically polarized.*

In Fig. 2, the electric field is parallel to the x-axis; the wave is *linearly polarized* (x-polarized). This wave could be generated by a straight wire antenna parallel to the x-axis.

An entirely distinct y-polarized plane wave could be generated with the same direction of propagation and recovered independently of the other wave using pairs

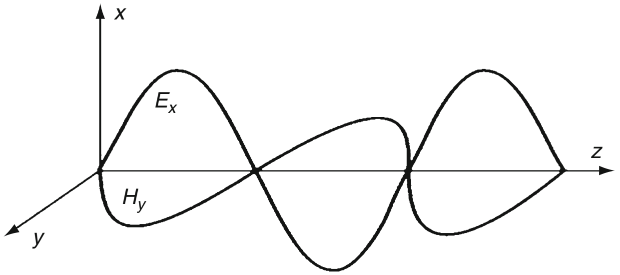


Fig. 2 Linearly polarized time-harmonic plane wave

of transmit and receive antennas with opposite polarization. This principle called "frequency reuse by orthogonal polarizations" is of critical importance in satellite communications in order to provide two independent communication channels in the same portion of allocated bandwidth.

The waves described above are linearly polarized since the electric field vector has a single direction along the whole of the propagation axis. If two plane waves of equal amplitude and orthogonal polarization are combined with a 90° phase shift, the resulting wave is *circularly polarized* (CP), in that the motion of the electric field vector will describe a circle centered on the propagation vector. The field vector will rotate by 360° for every wavelength traveled. Two solutions are possible: either *right-hand circularly polarized* (*RHCP*) or *left-hand circularly polarized* (*LHCP*); RHCP describes a wave with the electric field vector rotating clockwise when looking in the direction of propagation.

LHCP and RHCP waves are orthogonal and can also be considered for *frequency reuse* in satellite communications. Circularly polarized waves are commonly used in satellite communications, since they can be generated and received using antennas that do not require specific orientation along the direction of propagation (in contrast, the use of linear polarization implies that the transmit and receive electric field should match). Another advantage is that circularly polarized communication channels are not influenced by the Faraday rotation (see section "Tropospheric Effects on Satellite Communications").

If the component waves have different amplitudes and arbitrary phase difference, the result is the generic elliptically polarized wave introduced above, where the electric field vector still rotates at the same rate but varies in amplitude with time. In this case, the wave is characterized by the ratio between the maximum and minimum values of the instantaneous electric field, known as the axial ratio AR (Fig. 3):

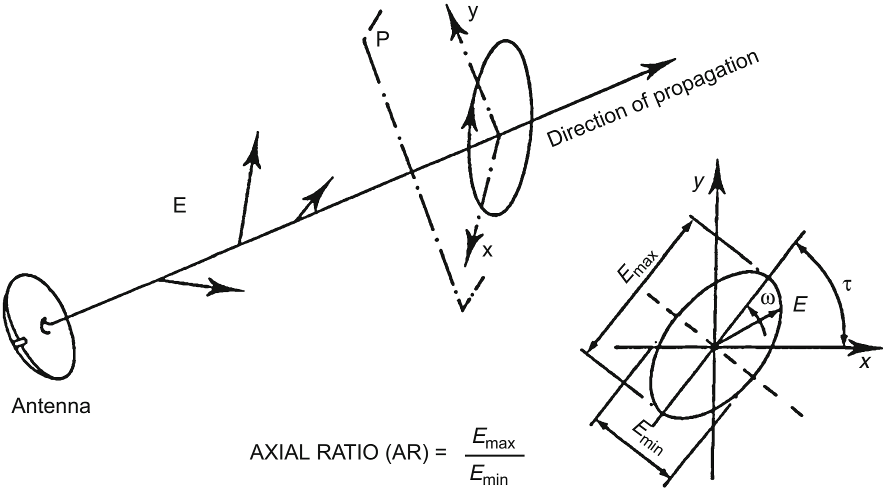


Fig. 3 Elliptical polarization and axial ratio

$$\mathrm{AR} = E_{\max} / E_{\min}$$

AR is defined to be positive for left-hand polarization and negative for right-hand polarization.

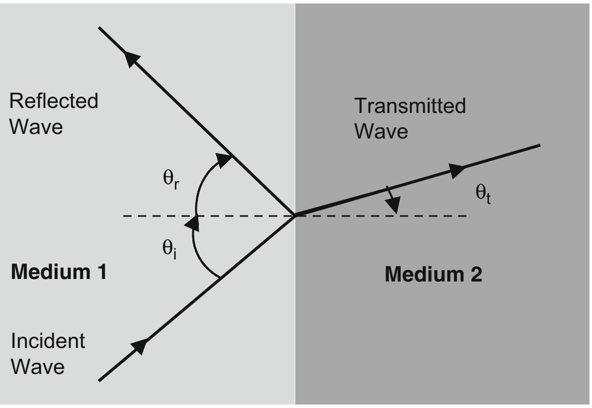
Propagation Mechanisms

In the previous section, the environment in which waves were propagating consisted of media within which the constitutive parameters did not vary in space. In practice, the real environment includes boundaries between media (between air and the ground, from Earth to space, etc.). These boundary effects give rise to changes in the amplitude, phase, and direction of propagating waves. Four main types of propagation mechanisms exist and can be used to describe the interaction of the electromagnetic waves with the environment. These mechanisms include reflection, refraction, scattering, and diffraction, and they all affect the amplitude, direction, and phase of the propagating radio waves.

Reflection, Refraction, and Transmission

Whenever a radio wave impinges to an obstruction having different material parameters than the propagation medium and greater dimensions than the wavelength, reflection and transmission will occur. Applying Maxwell's equations in the case of smooth homogeneous interface and mediums, the interaction behavior of the propagating wave with these interface mediums results in that two new waves are produced, each with the same frequency as the incident wave. Both waves have their Poynting vectors in the plane that contains both the incident propagation vector and the normal to the surface, called the *scattering plane*. The first wave, called *reflected* wave, propagates within medium 1 away from the boundary, making an angle θ_r to the normal. The second, called *transmitted* wave, results from the mechanism of *refraction* and travels on into medium 2 making an angle θ_t to the surface normal (Fig. 4).

Fig. 4 Plane boundary model



Snell's law of reflection states that the angle θ_r of the reflected field is equal to the angle of the incident field, θ_i, to the interface. Snell's law of refraction states that the refracted angle, θ_t, is a function of the incident angle and the materials of the two media:

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{n_2}{n_1}$$

where n_1 and n_2 are the refractive indices of the two media. The *refractive index* is obtained by using the constitutive parameters of the medium:

$$n = \sqrt{\varepsilon_r \mu_r}$$

The *refractive index* is the ratio of the free-space phase velocity, c, to the phase velocity in the medium v. Note that the frequency of the wave is unchanged following reflection and transmission; instead, the ratio $v/\lambda = f$ is maintained everywhere, then the wave within the denser medium (higher permittivity and permeability) has a smaller phase velocity and longer wavelength than in free space.

In addition to the change of direction, the interaction between the wave and the boundary also causes the energy to be split between the reflected and transmitted waves. Several methods can be used to calculate the reflected and the transmitted field. An introduction to the topic is provided considering the simple boundary model. The amplitudes of the reflected and transmitted waves are given relative to the incident wave amplitude by the Fresnel reflection and transmission coefficients. The Fresnel coefficients depend on the constitutive parameters of the materials, the polarization of the incident field, and the angle of incidence. The coefficients are different for parallel or perpendicular polarization.

	Perpendicular polarization	Parallel polarization
Reflection coefficient	$R_\perp = \frac{n_2 \ \cos\theta_i - n_1 \ \cos\theta_t}{n_2 \ \cos\theta_i + n_1 \ \cos\theta_t}$	$R_\| = \frac{n_2 \ \cos\theta_t - n_1 \ \cos\theta_i}{n_2 \ \cos\theta_t + n_1 \ \cos\theta_i}$
Transmission coefficient	$T_\perp = \frac{2n_2 \ \cos\theta_i}{n_2 \ \cos\theta_i + n_1 \ \cos\theta_t}$	$T_\| = \frac{2n_2 \ \cos\theta_i}{n_2 \ \cos\theta_t + n_1 \ \cos\theta_i}$

If E_i is the incident field on the interface of the media, then the reflected and transmitted field can be calculated by using the Fresnel coefficients. Therefore

$$E_r = RE_i \text{ and } E_t = TE_i$$

The formulations described until now describe the case of a *lossless medium*. For *lossy medium*, the same formulations can be used but the electrical permittivity ε that is used must be replaced by the complex permittivity:

$$\varepsilon_{\text{complex}} = \varepsilon' - j\varepsilon''$$

where ε' is the real part of the complex permittivity and ε'' the imaginary part. The *loss tangent* δ of a material or the conductivity σ can also be used.

Rough Surface Scattering

The reflection mechanism discussed so far refers to the case where the surface is considered to be smooth (specular reflection). This condition assumes that the reflection angle is equal to the incidence angle and that all the scattered energy is concentrated in the reflected ray. If the surface is considered to be rougher, the incident ray will be scattered from a large number of positions on the surface. This broadens the scattered energy, which effectively means that the energy in the specular direction is reduced. The degree of scattering depends on the incidence angle and on the roughness of the surface compared to the wavelength of transmission. To account for scattering, the reflection coefficient needs to be multiplied by the surface roughness factor, which is less than unity and depends exponentially on the standard deviation of the surface roughness.

Diffraction

Diffraction occurs when the path between the transmitter and the receiver is obstructed by a physical object. It could be noticed that some field power could be measured in the zone behind the object even though a direct line of sight does not exist between the transmitter and the area behind the obstacle. The effect of diffraction can be explained by Huygens principle, which states that all points on a wave front can be considered as point sources for the production of secondary wavelets and that these wavelets combine to produce a new wave front in the direction of propagation. Diffraction is caused by the propagation of secondary wavelets into the shadow region, giving rise to a bending of waves behind the obstacle. The field strength of a diffracted wave in the shadowed region is the vector sum of the electric field components of all the secondary wavelets in the space around the obstacle.

Antenna Fundamentals

The previous sections described wave interactions with the propagation media. This section introduces the antennas responsible for generating and receiving these waves. The parameters that characterize an antenna and its application are discussed further in chapter "▶ Satellite Radio Communications Fundamentals and Link Budgets."

Concept of Antenna

Most fundamentally, an antenna is a way of converting currents or guided waves traveling in one dimension present in a waveguide, microstrip, or transmission line, into radiating waves traveling in free space, carrying power away from the transmitter in three dimensions into free space (or vice versa). As a consequence of Maxwell's equations, the radiated wave is caused by current charges oscillating in periodic motion due to the excitation by a sinusoidal transmitter. Close to an antenna, the field patterns change rapidly with distance and include both reactive energy (which oscillates toward and away from the antenna, appearing as a reactance which only stores but does not dissipate energy) and radiating energy. Further away, the reactive fields are negligible and only the radiating energy is present, resulting in a variation of power with direction which is independent of distance. These regions are conventionally divided at a radius $R = 2L^2/\lambda$, where L is close to the dimension of the antenna and λ the wavelength. Within that radius is the *near-field* or *Fresnel region*, while beyond it lies the *far-field* or *Fraunhofer region.* Within the far-field region, the wave front behaves as spherical waves, so that only the power radiated in a particular direction is of importance, rather than the particular shape of the antenna. Measurements of the power radiated from an antenna have to be made in the far-field region (or special account has to be taken of the reactive fields if in the near field).

Radiation Pattern

The *radiation pattern* of an antenna is a plot of the far-field radiation from the antenna. It is convenient to work in spherical coordinates (r, θ, φ) rather than Cartesian coordinates, with the antenna placed at the origin. The power radiated from an antenna is typically plotted per unit solid angle, that is, the *radiation intensity U* (watts per unit solid angle). This is obtained by multiplying the power density at a given distance by the square of the distance r, where the power density S (watts per square meter) is given by the magnitude of the time-averaged Poynting vector:

$$U = r^2 S$$

In such a way, the radiation pattern is the same at all distances from the antenna, provided that r is within the far field. For example, in the case of an antenna radiating equally in all directions, an *isotropic* antenna, if the total power radiated by the antenna is P, then the power is spread over a sphere of radius r, so the power density at this distance and in any direction is

$$S = P/4\pi r^2$$

The radiation intensity, independent of r, is

$$U = r^2 S = P/4\pi$$

Radiation pattern is plotted by normalizing the radiation intensity by its maximum value.

Directivity and Gain

The *directivity D* of an antenna, function of the direction, is defined by the ratio of the radiation intensity of the antenna in direction (θ, φ) with respect to the mean radiation intensity in all directions, or in other terms, with respect to the radiation intensity of an isotropic antenna radiating same total power. Sometimes directivity is specified without referring to a direction. In this case, the term "directivity" implies the maximum value of $D(\theta,\ \varphi)\ = D_{\max}$.

The *power gain* G, or simply the *gain*, of an antenna is the ratio of its radiation intensity to that of an isotropic antenna radiating the same total power as accepted by the real antenna. The *gain* differs from the *directivity* by the ratio of the power radiated (which is proportional to the so-called radiation resistance), to the power accepted by the antenna (which is proportional to the sum of the radiation resistance and the *loss resistance*). Indeed, in practice, some power is lost within the antenna itself due to losses (*loss resistance*) in either the conducting or dielectric parts of the antenna.

Also, one should pay attention that transfer of power between microwave components connected together depends on the relative values of source and load impedance ($Z = R + jX$, with R resistance, X reactance).

The maximum of the source power is delivered to the antenna if the source impedance $Z_s(Z_s =\ R_s +\ jX_s)$ and the total antenna impedance Z_a ($Z_a =\ R_r +\ R_l + jX_a$, where R_r is the radiation resistance and R_l is the loss resistance) are complex conjugates.

When antenna manufacturers specify simply the gain of an antenna, they are usually referring to the maximum value of G. The use of the isotropic antenna as a reference is emphasized by giving the gain units of dBi (see end of paragraph):

$$G[\mathrm{dBi}] = 10\log G$$

A radiation pattern plot for a generic directional antenna is shown in Fig. 5, illustrating the *main lobe*, which includes the direction of maximum radiation (sometimes called the *boresight* direction) and several *sidelobes* separated by *nulls* where no radiation occurs. Note that the drawing represents a cut of the actual

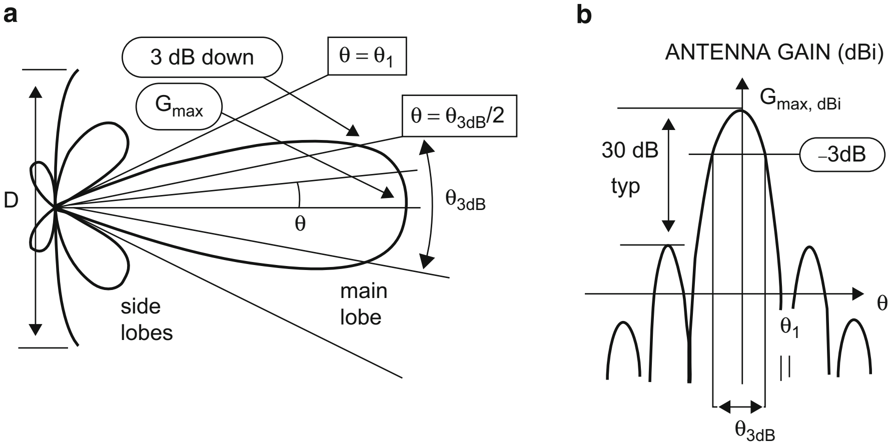


Fig. 5 Antenna radiation pattern

three-dimensional pattern in a plane containing the direction of maximum intensity (or maximum directivity or gain).

Some common parameters used to compare radiation patterns are defined as follows:

- The *half-power beamwidth* (HPBW), or θ_{3dB} *beamwidth*, is the angle subtended by the half-power points of the main lobe.
- The *sidelobe level* is the amplitude of the biggest sidelobe, usually expressed in decibels relative to the peak of the main lobe.
- The decrease of envelope of the peak of sidelobes as a function of the off-axis angle.

The Bel is a logarithmic unit of power ratio, where one Bel corresponds to an increase in power by a factor of 10 relative to a given reference power. It is conventional to express a ratio in decibels (one-tenth of a Bel) relative to some standard reference parameter. This reference is usually indicated by appending an appropriate letter. For instance, the power produced by a transmitter can be expressed in dBW, while here the gain of the antenna is in dBi (referring to the gain of an *isotropic antenna*).

Antenna Aperture

In the previous sections, the antenna has been considered as a transmitting device. It could also be used in receive mode. The performance of a receive antenna can be derived from the *reciprocity theorem:* If a voltage is applied to the ports of an antenna A and the current measured at the ports of another antenna B, then an equal current will be obtained at the ports of antenna A if the same voltage is applied to the ports of antenna B. A useful consequence of this theorem is that the antenna directivity and gain is the same whether used for receiving or transmitting, so all of the gain and pattern characteristics discussed above are fully applicable in receive mode. Another parameter that plays a critical role in the performance of a receiving antenna is the antenna effective aperture $A_{\text{eff}}[\text{m}^2]$.

If an antenna is used to receive a wave with a power density S watts per square meter, it will produce a power in its terminating impedance (receiver input impedance) of P_r watts. The constant of proportionality between P_r and S is the *effective aperture* of the antenna:

$$Pr = A_{\text{eff}} S$$

For some antennas, such as horn or dish antennas, the aperture has some physical interpretation (related to diameter of the physical surface), but the concept is just as valid for all antennas.

It could be shown that the maximum receive antenna gain $\text{G}_{\text{r, max}}$ is related to the effective aperture as follows:

$$\text{Gr, max} = A_{\text{eff}}\left(4\pi/\lambda^2\right)$$

Path Loss in Wireless Communications

The path loss between a pair of antennas is the ratio of the transmitted power to the received power. It includes all of the possible elements of loss associated with interactions between the propagating wave and any objects between transmit and receive antennas.

Free-Space Loss

Consider two antennas separated by a distance r great enough that the antennas are in each other's far field and oriented such as their direction of maximum gain is aligned. If P_t is the power accepted by the transmit antenna and G_{rmax} the antenna gain, then the power density incident on the receive antenna is

$$\Phi = P_t G_{t\,\max} / 4\pi r^2$$

Considering a receiving antenna with maximum gain $G_{r\max}$ and with aperture A_{eff}, the received power expresses as

$$P_r = A_{\text{eff}} P_t G_{t\,\max} / 4\pi r^2$$

Introducing the relationship between A_{eff} and $G_{r\max}$ and expressing the ratio of the received power to the transmitted power:

$$P_r / P_t = G_{t\,\max} G_{r\,\max} (\lambda / 4\pi r)^2$$

This expression defines the term $(4\pi r/\lambda)^2$ which is called the *free-space loss* L_{fs}. Because of square law dependence on both frequency and distance, the free-space loss increases by 6 dB for each doubling in either frequency or distance.

With frequency in gigahertz and distance R in kilometers, the free-space loss in decibels expresses as

$$L_{fs}(\text{dB}) = 92.4 + 20\log R + 20\log f$$

Polarization Mismatch Loss

The expression assumes ideal matching of polarization state of the receive antenna with that of the incoming wave. If not, a *polarization mismatch loss* should be accounted for, defined as the ratio between the power received by the antenna and the power which would be received by an antenna perfectly matched to the incident wave. In free space, the polarization state of the received wave is the same as that of the transmitter antenna. In more complicated media, which may involve polarization-sensitive phenomena such as reflection, refraction, and diffraction, the wave polarization is modified during propagation in accordance with the discussions in the previous sections.

Antenna Gain Fallout Due to Depointing

The calculation above assumed that the directions of maximum gain of the antennas are aligned. In practice, each antenna will introduce some depointing, that is, the angular difference between the direction of maximum gain and the direction of the other antenna. Therefore, the antenna gains may not necessarily be the maximum values.

Although some books introduce the gain fallout due to depointing in the path loss, to describe the propagation medium independently of system component gains, it is preferable that antenna gain values to be used are those corresponding to the direction of the other antenna.

Note that the same concept may apply for the polarization mismatch that could be considered as a reduction of the available antenna gain.

Atmospheric Effects for Earth-Space Radio Links

The concept of free-space loss assumes that the propagation medium between the antennas is lossless. This is a reality in space, for instance, in the case of intersatellite links. However, when considering an Earth-satellite link, the radio-frequency waves travel through the atmosphere and interact with the propagation medium.

Effects of the nonionized atmosphere (troposphere, i.e., lower layers of atmosphere) need to be considered at all frequencies but become critical above a few GHz and for low elevation angles. The propagation loss on an Earth-space path, relative to the free-space loss, is the sum of different contributions as follows:

Attenuation by atmospheric gases
Attenuation by rain and clouds
Scintillation

Effects of the troposphere on a satellite link are discussed in some detail in section "Tropospheric Effects on Satellite Communications."

Around and below 1 GHz, ionospheric effects play a particularly important role in the performances of satellite navigation systems. This is the topic addressed in section "Ionospheric Effects on Satellite Navigation Links."

Effect of Environment with Mobile Communications

The propagation effects in the case of satellite mobile communications are a combination of different phenomena. Indeed in addition to the path loss which is an overall decrease in field strength as the distance between the transmitter and the receiver increases, blockage and multipath effects create *shadowing* and *fast fading* which appear as time-varying processes between the antennas. Shadowing is the result of the varying nature of blockages of the radio-frequency path by trees and buildings in between the satellite and the receiver. Fast fading results from the constructive and destructive interference between multiple waves reaching the mobile after reflections on the local environment of the mobile. All of these processes vary as the relative positions of the transmitter and receiver change and as any contributing objects between the antennas are moved. Fast fading involves variations on the scale of a half-wavelength (30 cm at 1.5 GHz) and introduces variations as large as 20–30 dB. These are complex phenomena, the modeling of which is out of the scope of this chapter, and a detailed analysis can be found in Castanet (2007).

Link Budget

The path loss as discussed above between a pair of antennas is the ratio of the transmitted power to the received power. Therefore, knowing the transmitted power, it is possible to determine how much carrier power will be available at the receiving end input. In practice, the communication link performance is function of the ratio of the amount of available carrier power to the amount of unwanted signals (called noise). Indeed, at the receiver input, the wanted carrier power is not the only contribution. Different unwanted signals contribute to what is called the system noise temperature T. Those sources of noise include the noise captured by the antenna and generated by the feeder, which can actually be measured at the receiver input, plus the noise generated downstream in the receiver, which is modeled as a fictitious source of noise at the receiver input, treating the receiver as noiseless. The system noise temperature T can be used to infer the amount of noise power spectral density No (W/Hz).

The radio-frequency link performance is evaluated as the ratio of the received carrier power, C, to the noise power spectral density, No, and is quoted as the C/No ratio, expressed in hertz (Hz).

The calculation of carrier power, noise power spectral density for a complete communication link is the *link budget*, and the influence of the C/No ratio on the end-user performance (e.g., bit error rate, BER) depending on the type of modulation and coding are discussed in detail in chapter "▶ Satellite Radio Communications Fundamentals and Link Budgets."

Tropospheric Effects on Satellite Communications

Satellite telecommunication links operating at frequencies above 10 GHz are disturbed by tropospheric phenomena that degrade link availability and service performance. This section presents a review of the propagation properties for satellite telecommunication systems operating at Ku-, Ka-, and Q/V-bands.

Introduction

Most of today's satellite systems for fixed communications and broadcast are operating in the conventional Ku-band. But it should be noted that there is a strong interest from satellite operators and the space industry for using bands above Ku, mainly thanks to the wider bandwidths (up to 3 GHz at V-band) allocated to fixed satellite services (FSS) that make design of very high-capacity (up to hundreds Gbit/s) satellites possible. In addition, high-frequency bands offer significant technological advantages over conventional ones: reduced RF equipment size, which is particularly important considering the limited room available on board satellites, higher antenna gains, and narrower beams for a given antenna size.

However, a major limitation of high-frequency bands for Earth-space radio links is the effect of radio wave propagation through the lowest layers of the atmosphere, i.e., the troposphere. Indeed, as the operating frequency is increasing, the attenuation and scintillation effects of atmospheric gas, clouds, and rain become more severe, together with some depolarization phenomena. Ionospheric effects can be considered as negligible for these frequency bands. Each of these contributions has its own characteristics as a function of frequency, geographic location, and elevation angle. For non-GSO systems, the variation in elevation angle should also be considered.

Attenuation Effects

Attenuation is the main propagation effect affecting satellite links at high-frequency bands. Attenuation is made up with several contributions due to atmospheric gases (oxygen and water vapor) in clear-sky conditions, clouds, and precipitation. At Ku-band, only rain attenuation has to be considered, whereas the other contributions already not negligible at Ka-band can affect strongly system performance at Q/V-band.

Attenuation Due to Atmospheric Gases

Attenuation by atmospheric gases caused by absorption depends mainly on frequency, elevation angle, altitude above sea level, and water vapor density (absolute humidity). At frequencies below 10 GHz, it may normally be neglected. Its importance increases with frequency above 10 GHz, especially for low elevation angles. Recommendation ITU-R P.676 gives a complete method for calculating gaseous attenuation.

At a given frequency, the oxygen contribution to atmospheric absorption is relatively constant. However, both water vapor density and its vertical profile are quite variable. Typically, the maximum gaseous attenuation occurs during the season of maximum rainfall.

Cloud Attenuation

Clouds are constituted of suspended water drops which are clearly smaller in size than satcom wavelengths considered in this document. Cloud attenuation depends directly on the amount of liquid water, which is a function of the type of cloud present on the propagation path. Cloud attenuation increases with the integrated or total liquid water content (*ILWC* in kg/m^2) of the cloud and becomes noticeable from the Ka-band. Values of ILWC can be obtained everywhere in the world from maps given in Recommendation ITU-R P.840.

While cloud attenuation can be neglected at Ku-band, it has to be considered at Ka-band especially in tropical or equatorial climates or for low elevation links. At V-band, cloud attenuation can reach large values especially on the uplink which can degrade strongly the link quality.

Rain Attenuation

Attenuation due to rain is the most important propagation effect to be considered in the design of satellite telecommunication systems because it represents high level of attenuation and may affect communication systems for significant periods of time.

Spatial characteristics of precipitation influence the total attenuation on a link. Two major types of rain may be identified for temperate regions: stratiform and convective. Two more specific types have to be considered for tropical and equatorial regions and refer to tropical storms (hurricanes, typhoons) and monsoon precipitation. The dimensions of the rainy volume and the rain intensity depend on the type of rain: stratiform precipitation is composed of large areas with low rainfall rates, and convective precipitation consists of small areas (horizontal extent lower than 10 km) with intense rainfall (up to 50 mm/h for 0.01 % of an average year in temperate areas).

The rainfall rate (mm/h) plays a detrimental role in the amount of attenuation due to rain and is a key parameter in the various models that have been developed to estimate the amount of attenuation that could be expected on an Earth-space link. As rain presents a strong variability in space and time, the rainfall rate is a function of the percentage of time and depends on the climatic area. The rainfall rate conditions the specific attenuation (dB/km) of rain that depends on the temperature, size distribution, and shape of the raindrops. Specific attenuation may vary from fractions of dBs to tens of dBs as frequency increases (Ku- to V-bands) depending strongly on the rainfall (from a few mm/h to 100 mm/h in tropical areas). Global information on the cumulative distribution of rainfall rate is available in Recommendation ITU-R P.837 (Fig. 6).

The rain height is therefore another important parameter for calculating the influence of rain and is assimilated to the height of the −2 °C isotherm in order to take into account the influence of the melting layer (ITU-R Recommendation P.839).

A lot of propagation models for prediction of rain attenuation have been proposed in the literature for more than 30 years. Empirical models rely on the calculation of the rain-specific attenuation (expressed in dB/km) and on the estimation of a reduction factor in order to take into account the fact that rain cells have a limited spatial extension (Recommendation ITU-R P.618).

The rain-specific attenuation γ_R(dB/km) is obtained from the rainfall rate R (mm/h) using the power-law relationship:

$$\gamma_R = kR^{\alpha}$$

Values for the coefficients k and α are determined as functions of frequency, f (GHz), in the range from 1 to 1,000 GHz (Recommendation ITU-R P.838).

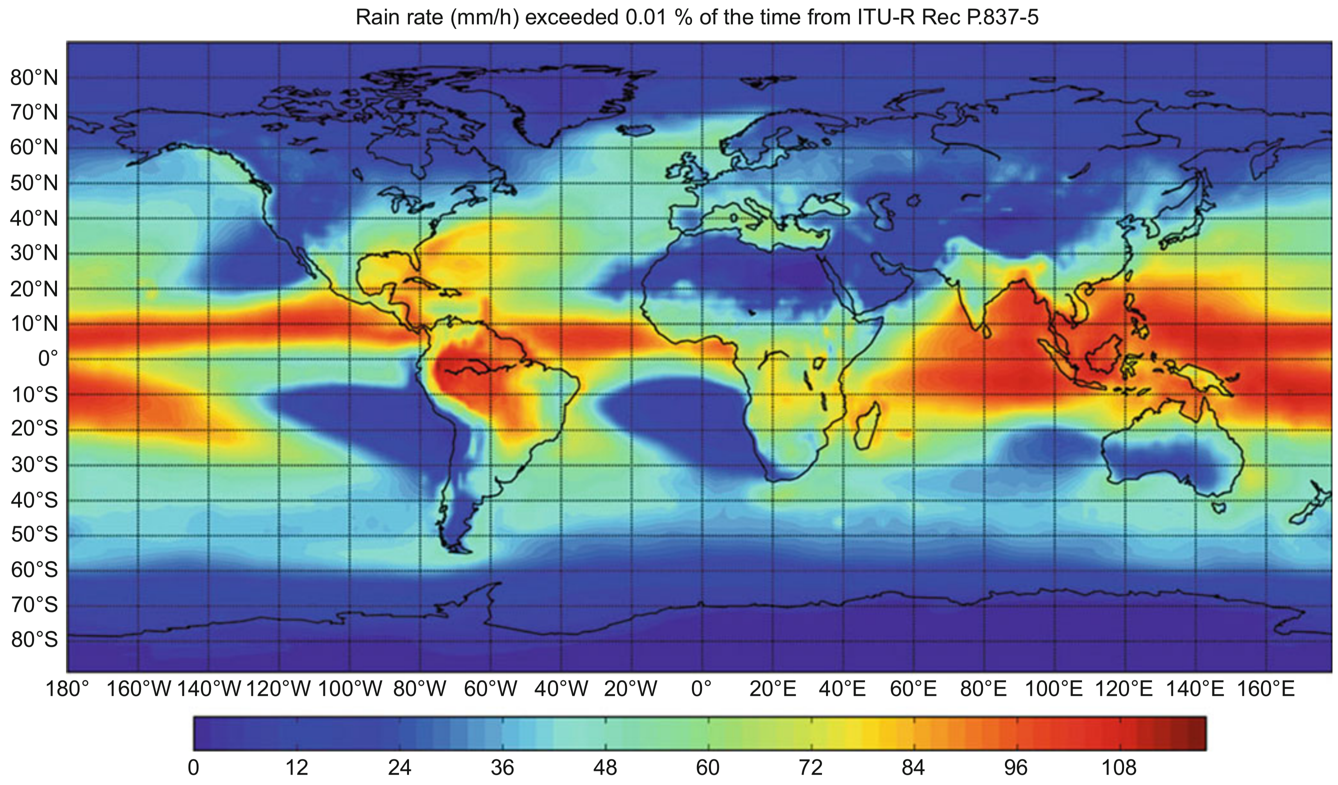


Fig. 6 Rain rate (mm/h) exceeded for 0.01 % of an average year

Estimation of the long-term statistics of the slant-path rain attenuation at a given location for frequencies up to 55 GHz for a given percentage of time can be obtained combining information available in Recommendations P.838, P.839, and P.618. Information could be synthesized in a step-by-step procedure that allows the calculation of rain attenuation that is overstepped at a given location for a given percentage of time (Maral and Bousquet 2009, chapter "▶ Space Telecommunications Services and Applications"). That percentage of time corresponds typically to the accepted unavailability of the link, i.e., one minus the required link availability (the percentage of the time where the link should be able to convey information with a specified quality).

At Ku-band, rain attenuation in a range of a few dBs can be overcome through the use of static margins which allow the required availability (say around 99.95 %) to be obtained for the Earth-space radio-frequency links. Due to technology limitations, a static system margin can no longer be considered as the sole means of compensating propagation disturbances for small percentage of time as the frequency increases to Ka-band and above. Fade mitigation techniques based on adaptive techniques (power control, adaptive coding and modulation, terminal diversity) have to be implemented in order to provide the required availability.

Scintillation Effects

Scintillation corresponds to rapid fluctuations of the received signal characteristics (amplitude, phase, angle of arrival). On satellite links, ionospheric scintillation and tropospheric scintillation can occur, depending on the frequency band.

Tropospheric scintillation is caused by small-scale refractive index inhomogeneities induced by atmospheric turbulence along the propagation path. Above about 10 GHz, tropospheric scintillation will be the unique phenomenon responsible for signal rapid fluctuations. Indeed ionospheric scintillation decreases with frequency, and tropospheric scintillation increases with frequency.

Depolarization Effects

Dual polarization transmission is a way to double the system capacity in a given frequency band, which allows the use of the frequency resource to be improved and then to reduce the transmission cost. However, some atmospheric effects limit system performances in terms of isolation between polarizations.

These atmospheric effects regarding wave polarization are different in the ionosphere and in the troposphere. For frequencies above 10 GHz, ionospheric effects (such as Faraday rotation which affects mainly system behavior in linear polarization) can be neglected. Tropospheric effects are caused by the presence of nonspherical hydrometeors (rain drops, ice crystals, snowflakes) with a non-vertical falling direction. It leads to different attenuations and phase shifts on each of the main orthogonal polarization directions, induces a rotation of the polarization plane, and then an increase of the coupling between both orthogonal polarizations.

Recommendation ITU-R P.618 provides a method to calculate long-term statistics of depolarization from rain attenuation statistics.

Conclusion

A survey of the propagation properties of the atmosphere from Ku- to Ka- and V-bands has been presented in this section. The main propagation impairments limiting the performance of satellite communications systems are:

- Attenuation effects due to gas, cloud, rain, and melting layer
- Scintillation in clear sky and inside clouds
- Depolarization due to rain and ice

At Ku-band, the propagation impairments are mainly governed by rain attenuation, which depends on the climatic zone. In temperate climate, typical values of a few dBs are encountered that could be easily mitigated using a static margin (in excess power) in the design of the radio-frequency link.

Systems working at Ka-band are submitted to stronger impairments, and particularly rain attenuation can reach high levels such as 20 or 30 dB in temperate climate at 30 GHz for percentages lower than 0.1 % of an average year, translating in a poor link availability. Fade mitigation techniques have to be implemented into the system in order to improve the link availability and the service performance.

When considering systems operating at V-band, very strong impairments can be obtained, up to several tens of dB at 50 GHz in terms of rain attenuation for low percentages of time. At these frequencies, other effects such as oxygen, water vapor, and cloud attenuation can degrade the system performance even in clear-sky conditions (especially for low elevation). The design of such system has to take into account these effects.

Ionospheric Effects on Satellite Navigation Links

The ionosphere is a layer of the upper atmosphere roughly between 50 and 1,500 km above the Earth, which has been ionized by the rays from the sun and, as a result, contains free electrons. Electromagnetic radio waves transmitted from satellites of Global Navigation Satellite Systems (GNSS) such as GPS, GLONASS, COMPASS, and GALILEO interact with the ionospheric plasma on their path to the receiver, changing their speed and direction of travel. The ionosphere has three main effects on satellite to ground signals at the frequencies used for satellite navigation: group delay, scintillations, and Faraday rotation. Since GNSS signals are circularly polarized, Faraday rotation will not be considered here.

It should be noted that troposphere introduces also some delay variation on the radio-frequency link. Models are used to mitigate these effects. The availability of GNSS systems results in a rapidly growing number of civilian users who are taking advantage of the freely available signals in space for a wide range of applications in navigation, surveying, tracking, etc., opening a broad market for developers of receivers and applications. Terrestrial overlay systems are being put in place to augment the navigation signals so that they can be used with higher precision or to provide integrity for safety-of-life critical operations such as landing a commercial aircraft (GBAS, ground-based augmentation systems, or SBAS, satellite-based augmentation systems).

Basic Concepts

The presence of these free electrons causes a propagation delay indicating a travel distance larger than the real one in the propagation of Global Navigation Satellite Systems (GNSS) signals from satellite to receiver. The amount of delay affecting a particular signal is proportional to the total number of free electrons along the propagation path. At the commonly used L-band frequencies (between 1.2 and 1.6 GHz), the ionosphere causes signal delays that correspond to a range of errors of up to 100 m. As the determination of the terminal position is based on the combination of range estimation to 4 or more satellites, signal delays causing range errors are detrimental to the performance of GNSS systems.

In a first-order approximation, the range error is proportional to the integral of the electron density along the ray path (total electron content – *TEC*). It is called vertical *TEC* when the propagation path has an elevation angle of 90°.

Within the frequency range of GNSS signals, the amount of delay affecting a particular signal is inversely proportional to the square of its frequency ($\sim 1/f^2$). Due to the inertia of the free electrons, this propagation error decreases with increasing radio frequency called ionospheric dispersion and has a measurable impact on radio signals up to frequencies of about 10 GHz.

Irregularities in the distribution of the free electrons along the propagation path can cause rapid fluctuations in the amplitude and phase of received signal, a phenomenon known as scintillation. The dispersive interaction of the navigation signal with the ionospheric plasma impinges space-based radio systems working at L-band frequencies. Scintillation can cause GNSS receivers to temporarily lose lock on one or more of the satellite signals, depending on the number of signals affected by scintillation and on the intensity of the resulting amplitude and phase fluctuations.

Satellite-Based Navigation Technique and Influence of Ionosphere

Satellite navigation relies on precise measurement of the time it takes for a signal to propagate from a spacecraft to the antenna of the navigation receiver. If there is no refractive medium on the path, the location can be determined by calculating the distance to a number of visible satellites by multiplying the time observed by the vacuum speed of light. With the position and clock offsets of the satellites precisely known, the user of the navigation receiver can easily establish his own position in a given coordinate system.

In reality there are some additional propagation delays, which have to be taken into account. These delays are caused by the interaction of the electromagnetic wave field with the charged particles of the ionospheric plasma, the neutral but polarized particles in the troposphere and with obstacles. The resulting observation equation can be written for the code and the carrier phase ф and ф, respectively, as

$$\begin{aligned} \phi &= \rho + c(dt - dT) + dI + dT + dMP + dq + dQ + \varepsilon \\ \phi &= \rho + c(dt - dT) - dI + dT + dMP + dq + dQ + N\lambda + \varepsilon \end{aligned}$$

where ρ is the true geometrical range between GPS satellite and receiver, c is the vacuum speed of light, dt and dT are the satellite and receiver clock errors, d_I is the ionospheric delay along the ray path s, d_T is the delay caused by the troposphere, d_{MP} is the multipath error, dq and dQ are the instrumental satellite and receiver biases, λ is the wavelength of the carrier, N is the phase ambiguity number (integer), and ε is the residual error.

The space weather-sensitive ionospheric propagation term d_I is a function of the refraction index and related effects such as diffraction and scattering.

In a first-order approximation of the refraction index, the ionospheric delay is proportional to the integral of the electron density along the ray path according to:

$$d_I = \frac{K}{f^2} \int_S^R n_e ds$$

with $K = 40.3\ \mathrm{m^3 s^{-2}}$.

The integral of the local electron density n_e along the ray path between satellite S and receiver R is called total electron content (*TEC*).

Typical vertical *TEC* values range from a few *TEC* units (1 TECU = 10^{16} electrons/m^2) for nighttime conditions at high latitudes to well above 200 TECU at equatorial latitudes and during ionospheric storm conditions.

At the L_1 GPS frequency of 1.575 GHz, 1 TECU is equivalent to a path length increase of 0.162 m indicating maximum ionospheric vertical delays of up to about 30 m under heavily perturbed conditions. The delay may even increase by a factor of more than 2 at low elevation angles, i.e., when the path length through the ionosphere increases considerably.

The first-order ionospheric range error (*IRE*) typically varies from 1 to 100 m at the GNSS frequencies.

Ionospheric Effects as a Function of Latitude

The world can be divided basically into three regions as the ionosphere is concerned.

They are:

- The low latitudes which include the equatorial and equatorial “anomaly” regions
- The midlatitudes
- The high latitudes which include auroral and polar caps

In practice, each of these major ionospheric regions can be further broken down into subregions. During geomagnetically disturbed periods, the auroral regions expand and move equatorward, thus expanding the size of the auroral regions and shrinking the size of the midlatitude regions. Thus, these regions are called transitional regions.

The largest region is the equatorial and so-called equatorial anomaly region, the effects of which can be measured up to at least ±20° from the magnetic equator. Note also that most of the continent of South America is located in the equatorial region, as is much of Africa, all of India, and portions of many other countries. Also, ground stations in countries that are located in the lower midlatitudes can view the ionosphere in the equatorial region or into the auroral region when monitoring GNSS satellites at relatively low elevation angle. Thus, the nature of the ionospheric effects affecting a GNSS receiver is not a simple function of the location of that receiver but of the geographic extent of the regions crossed by the lines of sight to the satellites in view.

Depending on the ionospheric regions, range delay characteristics and scintillation effects display different characteristics. In midlatitude regions, equivalent vertical delays form fairly smooth planar surfaces. Scintillation effects are in general insignificant. Strong phase scintillation occurs only during severe geomagnetic and ionospheric storms under solar maximum conditions. The midlatitude ionosphere normally has the smallest temporal and spatial gradients and has a rare scintillation occurrence, and the 5° by 5° grid of ionospheric corrections that is discussed into the next section has already been shown to work well in this region. The other world regions, however, have larger range delays as well as larger spatial and temporal gradients, a large occurrence of scintillation effects, and, in some cases, a larger day-to-day variability in range delay than the midlatitude ionosphere.

Mitigation Techniques

Single-Frequency Receivers

Using an ionospheric model, single-frequency users can reduce the ionospheric range error. Basic GNSS receivers obtain ionospheric corrections from simple models that were derived from analyses of historical data. One such model is the single-frequency GPS ionospheric model. This model is a vertical *TEC* model that uses a mapping function and a thin layer ionosphere assumption to convert vertical to slant *TEC*. It has been shown to statistically correct approximately 50 % rms of the actual vertical ionospheric delays. This estimate was done using Faraday rotation *TEC* data from stations at different geographical locations (Klobuchar 1987). The equations for the model are implemented in the receiver, but they operate on a small set of model coefficients broadcast by GPS satellites. These coefficients are regularly updated by the GPS ground control segment based on observations of the state of the ionosphere during the previous few days. The accuracy of the corrections is limited by the fact that the model is simple and therefore can only account for first-order variations in the ionosphere, not being able to reproduce basic spatial variations like those found at low latitudes. In particular, the model cannot account for possible irregularities, whether small or large, that may exist at the time the corrections are applied. This model performs best when the ionosphere is a quiet state. It tends to underestimate the magnitudes of delays during ionospheric storms.

The single-frequency ionospheric correction algorithm proposed for the Galileo system is based on the 3D NeQuick model (Radicella and Leitinger 2001) driven by an "effective ionization level" Az, valid for the whole world and applicable for a period of typically 24 h. The global Az is given in terms of three coefficients, function of the modified dip latitude or "MODIP" (that depends on the true magnetic dip and the geographic latitude). The Az coefficients are transmitted in the navigation message to the user to allow Az calculation at any wanted location. It must be noted that this model calculates the slant *TEC* by means of integration along the ray path. Therefore, no mapping function or thin layer approximation for the ionosphere is applied.

Residual range errors resulting from the limited accuracy of ionospheric delay corrections with single-frequency receivers have to be accounted for when evaluating the accuracy, integrity, availability, and continuity performance of GNSS navigation solutions. The main causes of large-scale variations in vertical *TEC* are related to the 11-year solar cycle, seasonal changes, day-to-day changes, and time of day.

Multiple-Frequency Receivers

The frequency dependency characterizes the ionosphere as a dispersive medium as already mentioned above. Since the amount of delay affecting a particular GNSS signal is inversely proportional to the square of the frequency of that signal, accurate calculations of *TEC* (or ionospheric range delay) along the line of sight between a receiver and a satellite are possible using two signals with different frequencies. However, the accuracy of such calculations depends on the calibration technique used and has to take into account effects like multipath.

The differential code phase $\Delta\Phi = \Phi_2 - \Phi_1$ measured at the two frequencies L_1 and L_2 reveals a first-order estimation of *TEC*:

$$TEC = \frac{\Delta\Phi \cdot f_1^2 f_2^2}{K \cdot \left(f_1^2 - f_2^2\right)} + TEC_{\mathrm{cal}}$$

Besides random errors, the calibration term TEC_{cal} includes specific satellite and receiver code phase delays that do not cancel out in $\Delta\Phi$.

The availability of a third or more carrier frequencies improves the performance. In case of a third frequency, the simplest approach is the linear combination of three different frequency pairs for computing ionosphere-free solutions.

Real-Time Ionospheric Measurements

To achieve high-precision positioning, propagation errors have to be mitigated as accurately as possible. Terrestrial overlay systems are used to augment the navigation signals so that they can be used with higher precision or to provide integrity for safety-of-life critical operations such as landing a commercial aircraft (GBAS, ground-based augmentation systems, or SBAS, satellite-based augmentation systems).

Indeed, errors of several tens of meters with partially corrected ionospheric delays cannot be tolerated for approach operations during which vertical guidance is provided to the aircraft such as approach with vertical guidance (APV) and precision approach (PA) operations.

SBAS receivers can correct for ionospheric delays more accurately than basic GNSS receivers because they use information derived from real-time ionospheric delay measurements (see following section). These measurements are collected by a network of reference stations and used by the SBAS ground system to estimate and broadcast vertical delays at the nodes of a standardized ionospheric grid defined on the thin shell. For each line of sight to a satellite, the receiver interpolates among the nearest grid nodes to the location of the ionospheric pierce point (IPP), then converts the interpolated vertical delay to a range (or slant) delay by applying a standardized "obliquity factor" that accounts for the angle at which the line of sight pierces the thin shell. With this type of augmentation, the accuracy of the corrections is limited by:

The relatively sparse sampling of the ionosphere available to the SBAS ionospheric delay estimation process

The reduction of the ionosphere (in which free electrons are distributed over a wide range of altitudes) to a two-dimensional model (ionospheric grid) associated with a fixed one-to-one mapping between vertical delays and range (slant) delays

Time delays between the collection of ionospheric delay measurements by the SBAS ground infrastructure, the broadcast of ionospheric grid information, and the application of the corrections by the SBAS receiver

The monitoring of ionosphere over a region of interest is a key component of SBAS systems. Several techniques have been considered for ionospheric grid estimation. Some techniques proceed on the basis of local models that are separately estimated at each ionospheric grid point (IGP). These techniques rely on low-degree polynomials such as a simple constant, a planar surface, or a quadratic surface to represent variations in vertical ionospheric delays in the local area of each IGP. Other techniques model the ionosphere over the entire service area with a unique model such as high-degree polynomial surfaces or spherical harmonics.

Current SBAS implementations rely on low-degree polynomial surfaces that are separately estimated at each IGP. Indeed, the accuracy of these simple models is good, they are easy to implement (and therefore also verify and certify), and formal error formulas corresponding to a prescribed level of confidence can be easily computed.

However, techniques that work well in midlatitude regions such as the planar fit approach to estimating vertical grid delays, and the simple obliquity factor which converts slant to vertical delays, then back from vertical to slant delays, may not perform adequately well to ensure a high availability of APV service in the equatorial region. Indeed, spatial and temporal changes in the ionosphere that cannot be adequately represented by simple models, and with the possible presence of narrow ionospheric structures that may develop between measurements, the accuracy of the ionospheric delay corrections broadcast to the users will be limited.

Ionospheric scintillation can cause many receivers in the region affected by it to lose lock on one or several satellite signals. Receiver design is the primary source of mitigation against scintillation effects.

GBAS provides pseudorange measurement corrections that include ionospheric corrections. These corrections are derived from real-time measurements and correct for the combined effects of various sources of errors including ionospheric delays. The main limitation to the accuracy of these corrections is the spatial separation between the GBAS ground station and the GBAS receiver.

Ground-Based Ionosphere Monitoring

As outlined in the previous sections, dual-frequency GNSS measurements enable the estimation of the total electron content (*TEC*) along the observed ray paths providing valuable information on the ionospheric state.

To obtain absolute *TEC* values, the ionospheric travel time delay or range error has to be calculated from differential code phases. Since the link-related travel times are biased by the instrumental delays at the satellite and at the receiver, the derived *TEC* data must be calibrated.

To convert the numerous slant *TEC* measurements into the vertical for reference, the ionosphere is assumed to be compressed in a thin single layer at the height h_I. The location of the piercing point of the radio link with the ionospheric layer is considered to be the "point" of measurement. Assuming a single-layer ionospheric shell at a fixed height of about 350–400 km (single-layer approximation), the slant *TEC* values are mapped to the vertical at the piercing points by a geometrical mapping function taking into account the elevation angle of the ray path from the satellite, the single-layer height, and the radius of the Earth (see Arbesser-Rastburg and Jakowski 2007).

Vice versa, this function is also used to compute the slant ionospheric propagation error if grid-based vertical *TEC* information is provided.

For ionospheric imaging, the availability of a sufficient number of measurements is required.

Acknowledgments The works of the European COST 296 Group and of the International SBAS-IONO group (White paper) represent important source contribution to this chapter section. I would also like to acknowledge the material provided by Bertram Arbesser-Rastburg, Head of the Propagation Division at ESA/ESTEC and Chair of the ITU-R Group 3. Similarly, the efforts of the ITU and its members and the quality of their publications, from which a lot of material for the Tropospheric section has been extracted, are fully acknowledged by the author. All ITU References are available on the web at www.itu.int/publications.

Cross-References

- ▶ History of Satellite Communications
- ▶ Regulatory Process for Communications Satellite Frequency Allocations

- ► Satellite Applications Handbook: The Complete Guide to Satellite Communications, Remote Sensing, Navigation, and Meteorology
- ► Satellite Communications Overview
- ► Satellite Radio Communications Fundamentals and Link Budgets
- ► Space Telecommunications Services and Applications

References

B. Arbesser-Rastburg, N. Jakowski, Effects on satellite navigation, Chapter 13, in *Space Weather – Physics and Effects*, ed. by V. Bothmer, I. Daglis (Springer, Berlin, 2007)

L. Castanet, *Influence of the Propagation Channel on Satellite Communications – Channel Dynamic Effects on Mobile, Fixed and Optical Multimedia Applications* (Satnex E-book, Shaker, 2007)

J.A. Klobuchar, Ionospheric time-delay algorithm for single-frequency GPS users. IEEE Trans. Aerosp. Electron. Syst. AES **23**(3), 325–331 (1987)

G. Maral, M. Bousquet, *Satellite Communications Systems*, 5th edn. (Wiley, Chichester, 2009)

J. Maxwell, A dynamical theory of the electromagnetic field. R. Soc. Trans. **CLV**, 459 (1865)

S.M. Radicella, R. Leitinger, The evolution of the DGR approach to model electron density profiles. Adv. Space Res. **27**(1), 35–40 (2001)

Recommendation ITU-R P.618–9:2007, Propagation data and prediction methods required for the design of Earth-space telecommunication systems

Recommendation ITU-R P.676–8:2009, Attenuation by atmospheric gase

Recommendation ITU-R P.836–4:2009, Water vapour: surface density and total columnar content

Recommendation ITU-R P.837–5:2007, Characteristics of precipitation for propagation modelling

Recommendation ITU-R P.838–3:2005, Specific attenuation model for rain for use in prediction method

Recommendation ITU-R P.839–3:2001, Rain height model for prediction method

Recommendation ITU-R P.840–5:2012, Attenuation due to clouds and fog

Recommendation ITU-R P.841–4:2005, Conversion of annual statistics to worst-month statistics

Recommendation ITU-R V.431–7:2000, Nomenclature of the frequency and wavelength bands used in telecommunications

Regulatory Process for Communications Satellite Frequency Allocations

Ram S. Jakhu

Contents

Abstract

Ready access to radio frequencies with limited interference and appropriate orbital positions are indispensable and highly valuable tools for all satellite communications. However, radio frequencies are limited, natural, and international resources. Furthermore, the global demand for radio spectrum has been increasing exponentially. Acting primarily through the International Telecommunication Union (ITU), the international community has developed a very complex regulatory regime that provides detailed rules and processes that govern the international allocation and allotment of radio frequencies and orbital positions.

R.S. Jakhu (✉)
Institute of Air and Space Law, McGill University, Montreal, QC, Canada
e-mail: ram.jakhu@mcgill.ca

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_14

This chapter briefly describes those regulatory processes as well as the manner in which they are created as part of the functioning of the ITU. This chapter thus provides the basics of ITU procedures for frequency allocations. The immediately following chapter provides the status of the ITU World Radio Conference held in Geneva, Switzerland in the fall of 2015 and the many key new outcomes that occurred at this conference with regard to satellite communications.

Keywords
Coordination and registration processes • Frequency allocations • Frequency allotments • Geostationary or geosynchronous • Orbit • International telecommunication union • ITU administrative regulations • ITU constitution • ITU convention • Master international frequency register • National frequency assignments • Plenipotentiary conferences • Radio communications bureau • Radio frequencies • Radio regulations board, radio stations • Satellite communications • Spectrum • Table of frequency allocations • World radio communications conference

Introduction

Radio frequencies and orbital positions are indispensable tools for satellite communications, i.e., in order to function properly, all satellites need appropriate orbits (paths in outer space) and interference-free radio frequencies. However, if they are not properly used, harmful interference could occur which, in turn, might reduce the quality of communications.

The starting point for a satellite operator is always to meet the mandatory requirements for a national radio license under national regulatory procedures. National assignment[1] of radio frequencies and orbital positions are carried out through applicable national licensing systems and procedures, which vary from country to country. However, once radio frequencies and orbital positions are secured by a satellite operator under a national license, the initiation of and follow-up for international coordination and registration processes to ensure interference-free use of radio frequencies and orbital positions through the International Telecommunication Union (ITU) are undertaken on behalf of that operator by the respective State (i.e., the national communications regulatory administration). It is interesting to note that in 2009, the ProtoStar communication satellite system (worth about $500 million), which was established to provide direct-to-home television and broadband internet services in Asia using Ku-band and C-band radio frequencies, had to declare bankruptcy because it could not find any State

---

[1] ITU Radio Regulations, Article 1.18, (as amended by World Radiocommunication Conference in 2007; hereinafter referred to as the ITU Radio Regulations) define “assignment” (of a radio frequency or radio frequency channel) as an “Authorization given by an Administration (State) for a radio station to use a radio frequency or radio frequency channel under specified conditions.”

(Administration) to properly register with ITU the radio frequencies and orbital positions it intended to use.[2]

The general perception is that the need for telecommunications (especially those involving the use of satellites and radio frequencies) will expand rapidly. However, the level of that expansion will be greatly determined by the availability of radio frequencies and orbital positions. All satellite operators will try to secure as many radio frequencies and orbital positions as possible in order to maintain and expand their market share. According to Adrian Ballintine, chief executive of NewSat a company that recently acquired seven geostationary orbital slots for its new satellites, "[t]he slots are extremely valuable assets with senior filing status, outstanding geographic footprint and certainly enough capacity to see NewSat's long-term future assured. We are now in a position to launch multiple satellites, each of which could generate in excess of $100 million of (earnings) per year."[3] However, radio frequencies and orbital positions are limited natural resources of an international character that must be shared among several radio services and all countries. Therefore, a very complex and extensive international regulatory regime has been established through an intergovernmental organization, i.e., the ITU. New challenges in accessing appropriate radio frequencies arise due to increasing privatization, competition, and globalization and even abuse of the ITU regulatory processes. One can expect more rigid and extensive international regulations and procedures in the not too distant future.

This chapter discusses: (1) the structure and functioning of the ITU in order to explain how rules and procedures are adopted and (2) the processes for obtaining radio frequencies and orbital slots through coordination within, and registration with, the ITU.

---

[2] S. Nadgir, UPDATE 2-ProtoStar files for bankruptcy, to sell satellites (2009), http://www.reuters.com/article/2009/07/29/protostar-idUSBNG6434120090729. Venture Capital Dispatch, VC-Backed ProtoStar falls out of orbit and into bankruptcy (2009), http://blogs.wsj.com/venturecapital/2009/07/29/vc-backed-protostar-falls-out-of-orbit-and-into-bankruptcy/. The ITU Radiocommunication Bureau expressed its concern about the operation of ProtoStar satellite network without its proper registration with the ITU, particularly due to possible harmful interference as well as "risk of physical collision between the 'Thuraya-3' satellite – operating on the assigned frequencies of the Emarsat-4S satellite network, which is located at 98.5° East – and the ProtoStar-1 satellite." The Bureau was of the opinion that there was "no information concerning the associated ITU satellite network filing, or the responsible administration under which the ProtoStar-1 satellite would be brought into use and operated, the Bureau was extremely concerned and alarmed to be the witness of a situation in which a satellite, in this case, ProtoStar-1, could be operated in contravention of the ITU Constitution, particularly No. 196 and No. 18.1 of Article 18 on Licences of the Radio Regulations, and this without a responsible Administration and by an unknown operating agency not duly authorized by an ITU Member State." See Director, Radiocommunication Bureau, Report to the 48th meeting of the Radio Regulations Board, ITU Document: RRB08-3/3-E, page 4, of 4 August 2008.

[3] NewSat price, turnover soar on satellite news, http://www.smh.com.au/business/newsat-price-turnover-soar-on-satellite-news-20110201-1abv4.html.

Structure and Functioning of the ITU

Purposes of the ITU

The purposes of the ITU, inter alia, are: (a) to maintain and extend international cooperation between all members for the improvement and rational use of telecommunications of all kinds; (b) to promote and to offer technical assistance to developing countries in the field of telecommunications; (c) to promote the development of technical facilities and their most efficient operation with a view to improving the efficiency of telecommunication services; and (d) to promote, at the international level, the adoption of a broader approach to the issues of telecommunications in the global information economy and society by cooperating with other world and regional intergovernmental organizations and nongovernmental organizations concerned with telecommunications.[4] Pursuant to these purposes, in the field of satellite telecommunications, the ITU: (a) effects the allocation[5] of bands of the radio frequency spectrum, the allotment[6] of radio frequencies, and registration of radio frequency assignments, and any associated orbital positions in the geostationary-satellite orbit in order to avoid harmful interference between radio stations; (b) coordinates efforts to eliminate harmful interference between radio stations and to improve the use made of the radio frequency spectrum and of the geostationary-satellite orbit; (c) facilitates the worldwide standardization of telecommunications for a satisfactory quality of service; and (d) fosters international cooperation in the delivery of technical assistance to developing countries and the creation, development, and improvement of telecommunication equipment and networks in developing countries.[7]

Legal Instruments Governing the ITU

The ITU functions under three basic international legal instruments (i.e., treaties). They are: (a) the Constitution of the International Telecommunication Union (hereinafter referred to as the ITU Constitution), (b) the Convention of the International

---

[4] Article 1.2 of the Constitution of the International Telecommunication Union (Geneva, 1992) as amended by the Plenipotentiary Conferences in 1994, 1998, 2002, 2006, and 2010 (hereinafter referred to as the ITU Constitution).

[5] ITU Radio Regulations, Article 1.16, define "allocation" (of a frequency band) as an "Entry in the Table of Frequency Allocations of a given frequency band for the purpose of its use by one or more terrestrial or space radiocommunication services or the radio astronomy service under specified conditions. This term shall also be applied to the frequency band concerned."

[6] ITU Radio Regulations, Article 1.17 define "allotment" (of a radio frequency or radio frequency channel) as an "Entry of a designated frequency channel in an agreed plan, adopted by a competent conference, for use by one or more administrations for a terrestrial or space radiocommunication service in one or more identified countries or geographical areas and under specified conditions."

[7] ITU Constitution, Article 1.3.

Telecommunication Union (hereinafter referred to as the ITU Convention), and (c) the Administrative Regulations.[8] The Administrative Regulations, which regulate the use of telecommunications and are binding on all Members of ITU,[9] are: (a) the International Telecommunication Regulations and (b) the Radio Regulations.[10] In the event of an inconsistency between a provision of the Constitution and a provision of the Convention or of the Administrative Regulations, the Constitution prevails.[11] In the case of inconsistency between a provision of the Convention and a provision of the Administrative Regulations, the Convention prevails.[12] Ratification, acceptance, approval, or accession of the Constitution and the Convention, also constitute consent to be bound by the Administrative Regulations adopted prior to the date of signature of the Constitution and the Convention.[13] A Member State must notify its consent to be bound by a partial or complete revision of the Administrative Regulations to the ITU Secretary-General.

The Members of ITU are bound to abide by the provisions of the Constitution, the Convention, and the Administrative Regulations in all telecommunication offices and stations established or operated by them which engage in international services or which are capable of causing harmful interference to radio services of other Administrations (countries).[14] The Members are also bound to take the necessary steps to ensure the observance of the provisions of the Constitution, the Convention, and the Administrative Regulations by operating agencies (including private companies) authorized by them to establish and operate telecommunications and which engage in international services or operate stations capable of causing harmful interference to the radio services of other countries.[15] No communication transmitting station ought to be established or operated by a private person without a license issued in conformity with the provisions of the Radio Regulations by the government of the Administration to which the station in question is subject.[16] Therefore, Member States must require their operating agencies to obtain radio licenses from the appropriate national regulatory authorities (e.g., the Federal Communications Commission in the United States). However, under Article 48 of the ITU Constitution, Member States are exempted from the application of ITU Radio Regulations with respect to their military radio installations, although these installations must observe regulatory provisions relative to giving assistance in case of distress and to the measures to be taken to prevent harmful interference. In addition, when these installations are used for the provision of public correspondence service or other

---

[8] ITU Constitution, Article 4.1.

[9] ITU Constitution, Article 54.1.

[10] ITU Constitution, Article 54.1.

[11] ITU Constitution, Article 4.4.

[12] ITU Constitution, Article 4.4.

[13] ITU Constitution, Article 54.2.

[14] ITU Constitution, Article 6.1.

[15] ITU Constitution, Article 6.2.

[16] ITU Radio Regulations, Article 18.1.

services governed by the Radio Regulations, they must, in general, comply with the regulatory provisions for the conduct of such services.

ITU Membership

Membership in the ITU, the oldest specialized agency of the United Nations (UN), is based on the principle of universality, i.e., the rule is: once a member, always a member. Thus, membership cannot be canceled or suspended. However, a Member may be barred from attending a particular meeting or all meetings or conferences of the ITU. There are two types of members of ITU, i.e., State Members and Sector Members-Associates. As of February 2011, there were 193 Member States and over 700 Sector Members of the ITU.[17]

State membership: A State that is a Member of the UN becomes State Member of the ITU by acceding to its Constitution and Convention.[18] If a State is not a Member of the UN, its application for membership needs to be approved by two-thirds of the Member States of the ITU.[19] An "instrument of accession" covering both the Constitution and the Convention must be deposited with the Secretary-General of the ITU in order for a State to become a Member State of the ITU.

Sector membership: Global communication industry representatives are allowed to join the ITU as Sector Members. This supports the purposes and activities of the Sector of which they become Members and contributes to defraying the expenses of the Sector concerned. The entities that are eligible to become Sector Members include: (a) recognized operating agencies, scientific or industrial organizations, and financial or development institutions which are approved by the Administration of the Member State concerned; (b) other entities dealing with telecommunication matters which are approved by the Administration of the Member State concerned; and (c) regional and other international telecommunication, standardization, and financial or development organizations.[20] With the prior approval of the Administration (Government) of the Member State in which it has its headquarters, an entity may submit an application to the Secretary-General of ITU, on the receipt of which that entity becomes a Sector Member.[21] Sector Members are entitled to participate (a) in the work of Study Groups and subordinate groups and (b) in the process of preparing recommendations and comments before the adoption of recommendations, if any. However, they are not allowed to be involved in voting for, or approval of, recommendations.[22] As from October 2010, academic institutions, universities, and research establishments that are concerned with the development of

---

[17] http://www.itu.int/members/index.html.

[18] ITU Constitution, Article 2.

[19] ITU Constitution, Article 2.

[20] http://www.itu.int/members/sectmem/categories.html.

[21] http://www.itu.int/members/sectmem/categories.html.

[22] http://www.itu.int/members/sectmem/participation.html.

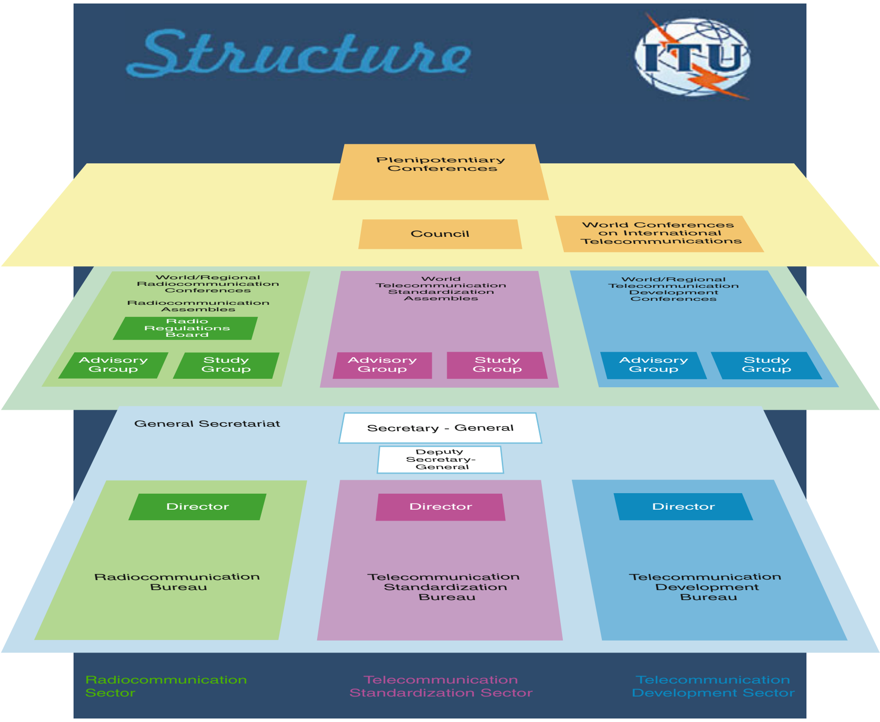


Fig. 1 Structure of International Telecommunication Union (ITU, ▶ http://www.itu.int/net/about/structure.aspx. (Reproduced with the permission of ITU)

communications are allowed to be admitted to participate without voting rights in the work of the three Sectors of ITU for a trial period of 4 years.[23]

Organizational Structure of the ITU

Since its birth in 1865,[24] the ITU has evolved to become a very complex international organization with a multifaceted organizational structure, which currently includes: (a) the plenipotentiary conference, (b) the Council, (c) World Conferences on International Telecommunications, (d) the Telecommunication Standardization Sector, (e) the Telecommunication Development Sector, (f) the Radiocommunication Sector, and (g) the General Secretariat.[25] See Fig. 1.

---

[23] http://www.itu.int/members/academia/index.html.

[24] For details about ITU's history, visit: http://www.itu.int/en/history/Pages/default.aspx.

[25] ITU Constitution, Article 7.

The plenipotentiary conference, the supreme organ of the Union, is composed of delegations representing all Member States.[26] It is convened every 4 years. Exceptionally, extraordinary plenipotentiary conference may be convened with a restricted agenda to deal with specific matters by a decision of the preceding ordinary plenipotentiary conference or if two-thirds of the Members of the Union so request.[27] The plenipotentiary conference, inter alia, determines the general policies of the Union; elects the Members of the ITU Council, the Secretary-General and the Deputy Secretary-General, the Directors of the Bureaux of the three ITU Sectors, and the members of the Radio Regulations Board; considers and adopts, if appropriate, proposals for amendments to the Constitution and the Convention; and concludes or revises, if necessary, agreements between the ITU and other international organizations.[28]

The council is the governing body of the ITU during the intervening periods between the plenipotentiary conferences. The Council considers "broad telecommunication policy issues in accordance with the guidelines given by the plenipotentiary conference to ensure that the Union's policies and strategy fully respond to changes in the telecommunication environment."[29] It drafts the agenda for administrative conferences. Currently, there are 48 Member States represented on the Council and they were elected during the 2010 plenipotentiary conference.[30]

The telecommunication standardization sector (*ITU-T*) studies technical, operating, and tariff questions and adopts recommendations on them with a view to standardizing telecommunications on a worldwide basis.[31] World telecommunication standardization conferences are convened every 4 years; however, an additional conference may be held in accordance with the relevant provisions of the Convention.[32]

The telecommunication development sector (*ITU-D*) facilitates and enhances telecommunications development by offering, organizing, and coordinating technical cooperation and assistance activities. It is required to: (a) raise the level of awareness of decision-makers concerning the important role of telecommunications

---

[26] ITU Constitution, Article 8.1.

[27] ITU Constitution, Article 8.3.

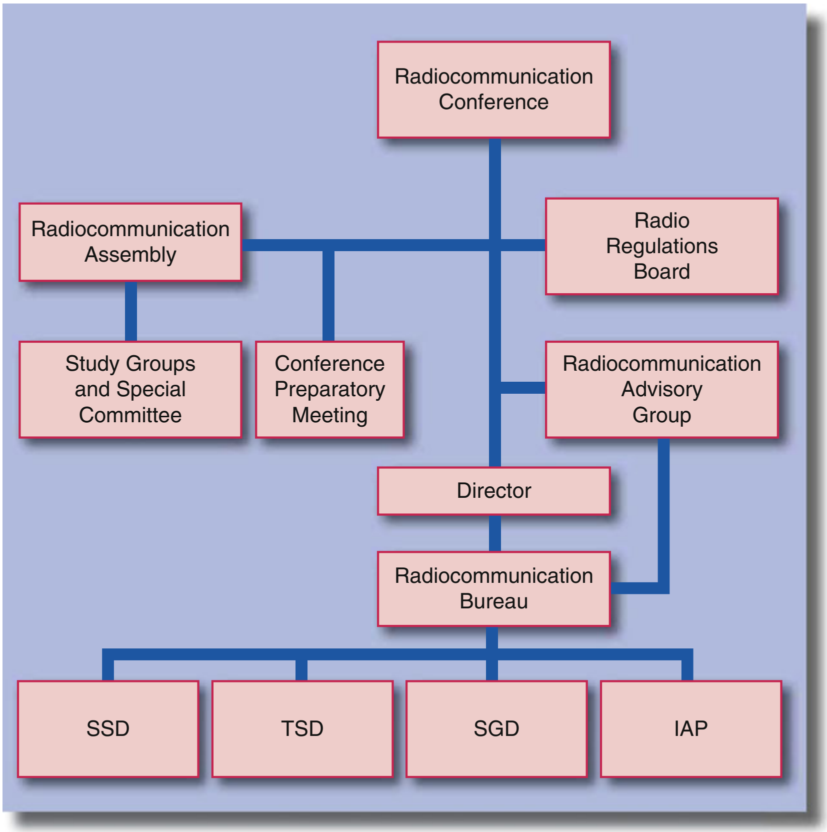
[28] ITU Constitution, Article 8.2.

[29] ITU Constitution, Article 10.4.

[30] They are: Mexico (votes received 143), Brazil (135), Canada (135), Argentina (131), Cuba (125), Venezuela (119), United States (114), Costa Rica (93), Paraguay (91), Spain (138), Italy (136), France (135), Germany (130), Sweden (126), Turkey (125), Greece (109), Russian Federation (123), Bulgaria (116), Romania (114), Poland (107), Czech Republic (93), Egypt (122), Kenya (119), Algeria (114), Morocco (114), Ghana (112), Tunisia (111), South Africa (105), Mali (101), Burkina Faso (97), Nigeria (95), Rwanda (93), Senegal (93), Cameroon (83), Indonesia (135), China (134), Japan (133), Malaysia (127), Korea (Rep. of) (125), Bangladesh (123), Thailand (121), Australia (119), India (119), United Arab Emirates (114), Kuwait (108), Saudi Arabia (105), and Philippines (97).

[31] ITU Constitution, Article 17.1.

[32] ITU Constitution, Article 18.



SSD – Space Services Department
TSD – Terrestrial Services Department
SGD – Study Group Department
IAP – Informatics, Administration and Publications Department

Fig. 2 Structure of radiocommunications sector (ITU, http://www.itu.int/ITU-R/index.asp?category=information&rlink=sector-organization&lang=en) (Reproduced with the kind permission of ITU)

in national economic and social development programs, (b) encourage participation by industry in telecommunication development in developing countries, and (c) offer advice on the choice and transfer of appropriate technology.[33]

The radiocommunication sector (*ITU-R*) is the main unit within the ITU that adopts the detailed rules, procedures, and standards for ensuring interference-free use of radio frequencies and orbital positions and their routine implementation. For this purpose, the Sector functions under a very complex organizational structure and by following intricate decision-making processes. See Fig. 2.

---

[33] ITU Constitution, Article 21.2.

The primary function of the Radiocommunication Sector is to ensure the rational, equitable, efficient, and economical use of the radio frequency spectrum by all radiocommunication services, including those using the geostationary-satellite orbit.[34] It works through: (a) world and regional radiocommunication conferences; (b) the Radio Regulations Board; (c) radiocommunication assemblies, which are associated with world radiocommunication conferences; (d) radiocommunication study groups; and (e) the Radiocommunication Bureau, headed by an elected Director.[35] The members of the Radiocommunication Sector are: (a) of right, the Administrations of all Member States of the Union and (b) entities or organizations that become Sector Members in accordance with the relevant provisions of the Convention.[36]

World radiocommunication conferences (*WRCs*) are exclusively entitled to revise the ITU Radio Regulations, to deal with any question related to matters of worldwide character involving allocation and allotment of radio frequencies and orbital positions,[37] and to instruct the Radio Regulations Board and Radiocommunications Bureau and review their activities. They are normally convened every 3–4 years and are the main bodies of the ITU which establish international regulatory processes related to satellite communications.

The radio regulations board (*RRB*) is constituted by 12 part-time individuals elected by the plenipotentiary conference.[38] The members of the RRB are required to serve as custodians of an international public trust and not as representatives of their respective Member States countries or regions.[39] They must refrain from intervening in decisions directly concerning their own respective Administrations and must not request or receive instructions relating to the exercise of their duties from any government or any public or private organization or person. Similarly, Member States and Sector Members are obligated to respect the exclusively international character of the duties of the members of the RRB.[40] The Board meets four times a year, in stark contrast to the more permanent nature of its predecessor, the International Frequency Registration Board (IFRB). Fewer functions and powers have been entrusted to the RRB as compared to those of the former IFRB.

The RRB develops rules of procedures, which are submitted to the upcoming World Radiocommunication Conference for the necessary modifications to the ITU

---

[34] ITU Constitution, Article 12.1.

[35] ITU Constitution, Article 12.2.

[36] ITU Constitution, Article 14.

[37] ITU Constitution, Article 14.1.

[38] Elected by the 2010 plenipotentiary conference, the current members of the radio regulations board are: Ricardo Luis Terán (Argentine Republic); Julie Napier Zoller (United States); Alfredo Magenta (Italy); Mindaugas Zilinskas (Lithuania); Victor Strelets (Russian Federation); Baiysh Nurmatov (Kyrgyz Republic); Stanley Kaige Kibe (Kenya); Mustapha Bessi (Morocco); Simon Koffi (Côte d'Ivoire); Yasuhiko Ito (Japan); Ali R. Ebadi (Malaysia); and P. K. Garg (India).

[39] ITU Constitution, 13.3.

[40] ITU Constitution, 13.3.

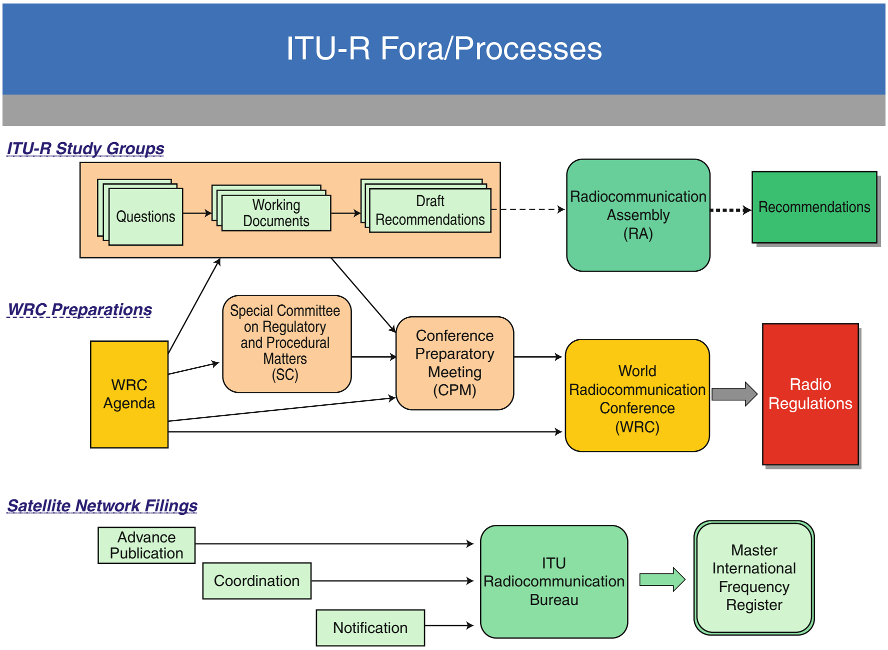


Fig. 3 ITU radiocommunications sector and processes (Reproduced with the kind permission of Boeing)

Radio Regulations. The RRB's Rules of Procedure are used by the Radiocommunications Bureau for the purpose of registering radio frequencies and orbital positions in the Master International Frequency Register (Master Register).[41]

In brief, it should be noted that the organizational structure of and the procedures followed by the ITU Radiocommunication Sector are complex and intricate. See Fig. 3.[42]

Processes for Obtaining Radio Frequencies and Orbital Slots

The main goals of the ITU's international regulatory regime governing satellite communications are to avoid harmful interference and to ensure equitable access to radio frequencies and satellite orbital slots. According to Article 45 of the ITU Constitution:

---

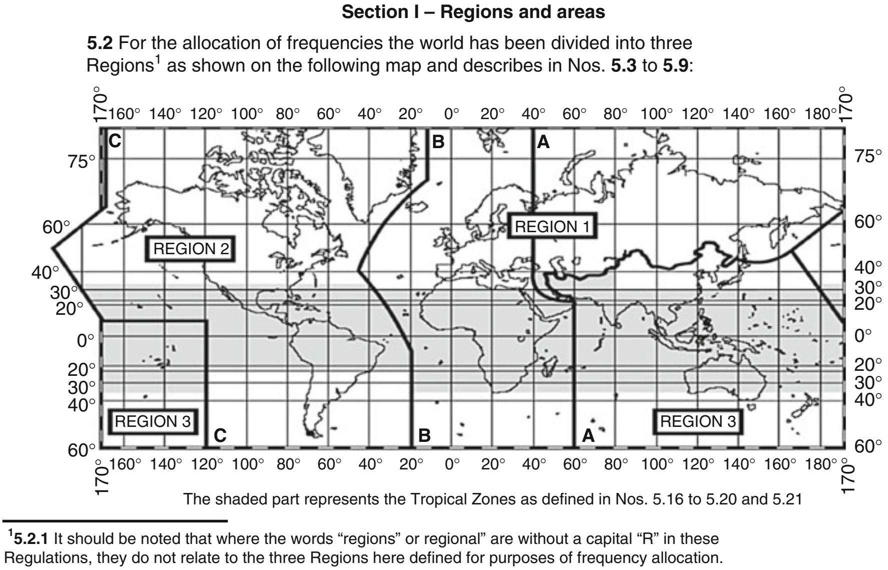
[41] ITU Constitution, Article 14.2.

[42] Audrey Allison, "Latest developments in ITU radiofrequency regulations and procedures including coordination and registration, for interference-free operation of satellites," a lecture given at the Institute of Air and Space Law, McGill University, Montreal, Canada, 26 January 2008.

- All radio stations must be established and operated in such a manner as not to cause harmful interference to the radio services of other Members which are operated in accordance with the provisions of the ITU Radio Regulations.
- The Members are required to take all practicable steps to prevent the operation of electrical apparatus and installations of all kinds from causing harmful interference to the radio services.
- Each Member undertakes to require their recognized operating agencies to respect the above-mentioned obligations.

The avoidance of harmful interference is achieved through processes of allocation and registration of radio frequencies and orbital positions with the ITU. In this context, it is important to keep in mind the earlier mentioned specific definition of “allocation” as included in the ITU Radio Regulations. For the purpose of allocation, the ITU has divided the world into three regions. See Fig. 4.

As determined periodically, World Radiocommunication Conferences add, delete, and modify all allocations in a complex “Table of Frequency Allocations” under Article 5 of the ITU Radio Regulations (Table 1). This Article provides the categories of all radio services as well as their respective rights or priorities to use.



Section I – Regions and areas

5.2 For the allocation of frequencies the world has been divided into three Regions[1] as shown on the following map and describes in Nos. **5.3** to **5.9**:

The shaded part represents the Tropical Zones as defined in Nos. 5.16 to 5.20 and 5.21

---

[1] **5.2.1** It should be noted that where the words “regions” or regional” are without a capital “R” in these Regulations, they do not relate to the three Regions here defined for purposes of frequency allocation.

Fig. 4 Regions and areas of the world for purposes of allocation of radio frequencies (ITU Radio Regulations, Article 5.2.) (Reproduced with the kind permission of ITU)

Table 1[43] (i.e., one page extracted from the Table of Frequency Allocations in Article 5 of Radio Regulations) provides an example of allocations within the 2,520–2,700 MHz band to various services within the three ITU regions. In order

---

[43] ITU Radio Regulations, Article 5, specify that:

- 5.23 *Primary and secondary services*
- 5.24 1) Where, in a box of the Table in Section IV of this Article, a band is indicated as allocated to more than one service, either on a worldwide or regional basis, such services are listed in the following order:
- 5.25 *a*) services the names of which are printed in "capitals" (example: FIXED); these are called "primary" services;
- 5.26 *b*) services the names of which are printed in "normal characters" (example: Mobile); these are called "secondary" services (see Nos. 5.28–5.31).
- 5.27 2) Additional remarks shall be printed in normal characters (example: MOBILE except aeronautical mobile).
- 5.28 3) Stations of a secondary service:
- 5.29 *a*) shall not cause harmful interference to stations of primary services to which frequencies are already assigned or to which frequencies may be assigned at a later date;
- 5.30 *b*) cannot claim protection from harmful interference from stations of a primary service to which frequencies are already assigned or may be assigned at a later date;
- 5.31 *c*) can claim protection, however, from harmful interference from stations of the same or other secondary service(s) to which frequencies may be assigned at a later date.
- 5.34 *Additional allocations*
- 5.35 1) Where a band is indicated in a footnote of the Table as "also allocated" to a service in an area smaller than a Region, or in a particular country, this is an "additional" allocation, i.e., an allocation which is added in this area or in this country to the service or services which are indicated in the Table (see No. 5.36).
- 5.36 2) If the footnote does not include any restriction on the service or services concerned apart from the restriction to operate only in a particular area or country, stations of this service or these services shall have equality of right to operate with stations of the other primary service or services indicated in the Table.
- 5.38 *Alternative allocations*
- 5.39 1) Where a band is indicated in a footnote of the Table as "allocated" to one or more services in an area smaller than a region, or in a particular country, this is an "alternative" allocation, i.e., an allocation which replaces, in this area or in this country, the allocation indicated in the Table (see No. 5.40).
- 5.40 2) If the footnote does not include any restriction on stations of the service or services concerned, apart from the restriction to operate only in a particular area or country, these stations of such a service or services shall have an equality of right to operate with stations of the primary service or services, indicated in the Table, to which the band is allocated in other areas or countries.
- 5.41 3) If restrictions are imposed on stations of a service to which an alternative allocation is made, in addition to the restriction to operate only in a particular country or area, this is indicated in the footnote.
- 5.50 5) The footnote references which appear in the Table below the allocated service or services apply to more than one of the allocated services, or to the whole of the allocation concerned. (WRC-2000)
- 5.51 6) The footnote references which appear to the right of the name of a service are applicable only to that particular service.

Table 1 Table of frequency allocations (Reproduced with the kind permission of ITU)

2,520–2,700 MHz		
Allocation to services		
Region 1	Region 2	Region 3
2,520–2,655 FIXED MOD 5.410 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.413 MOD 5.416	**2,520–2,655** FIXED MOD 5.410 FIXED-SATELLITE (space-to-Earth) 5.415 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.413 MOD 5.416	**2,520–2,535** FIXED MOD 5.410 FIXED-SATELLITE (space-to-Earth) 5.415 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.413 MOD 5.416 5.403 5.415A ADD 5.4A01
		2,535–2,655 FIXED MOD 5.410 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.413 MOD 5.416 5.339 5.417A 5.417B 5.417C
5.339 5.405 5.412 5.417C 5.417D 5.418B 5.4I8C	5.339 5.417C 5.417D 5.418B 5.418C	5.417D MOD 5.418 5.418A 5.418B 5.418C
2,655–2,670 FIXED MOD 5.410 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.347A 5.413 MOD 5.416 Earth exploration-satellite (passive) Radio astronomy Space research (passive)	**2,655–2,670** FIXED MOD 5.410 FIXED-SATELLITE (Earth-to-space) (space-to-Earth) 5.347A 5.415 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.347A 5.413 MOD 5.416 Earth exploration-satellite (passive) Radio astronomy Space research (passive)	**2,655–2,670** FLXED MOD 5.410 FIXED-SATELLITE (Earth-to-space) 5.415 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.347A 5.413 MOD 5.416 Earth exploration-satellite (passive) Radio astronomy Space research (passive)
5.149 5.412	5.149	5.149 5.420

to fully understand the precise nature and scope of specific allocations, one needs to keep in mind the definitions, categories of services and their operational priorities, and hundreds of footnotes mentioned under almost each and every allocation. These footnotes serve as explanations and exceptions to the general rule of allocations made in each category and region (Fig. 5).

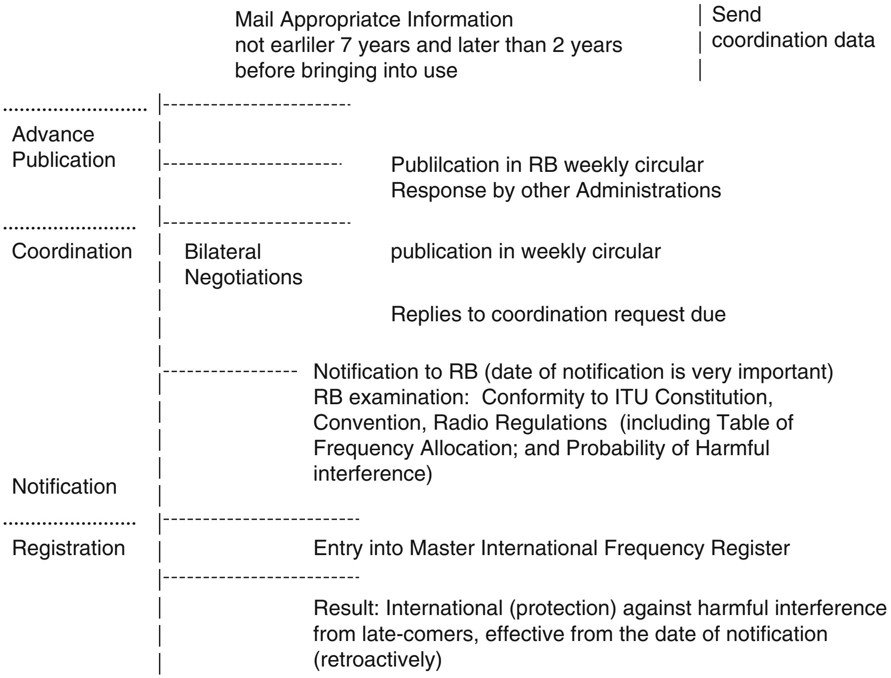


Fig. 5 First-come, first-served procedure (applicable to all space services in all bands except those for which a priori plans are adopted)

International Notification, Coordination, and Registration of Radio Frequencies and Orbital Positions

According to Article 8 of the ITU Radio Regulations, international rights and obligations in respect of radio frequency assignments and orbital positions are derived from the recording of those assignments and positions in the Master International Frequency Register (Master Register or MIFR) in accordance with the provisions of ITU Radio Regulations. Only those radio frequency assignments and associated orbital positions that have been properly recorded in the Master Register are entitled to the right to international recognition, i.e., protection against harmful interference. If harmful interference to any station, whose assignment has been properly registered, is actually caused by the use of a frequency assignment which is not in conformity with the ITU Radio Regulations, the radio station using the latter frequency assignment must, upon receipt of advice thereof, immediately eliminate this harmful interference.[44]

---

[44] ITU Radio Regulations, Article 8.5.

There are two methods for international notification, coordination, and registration of radio frequencies and orbital positions, i.e., the so-called first-come, first-served method and the a priori planning.

First Come: First Served

In order to be entitled to a right of international protection against harmful interference, the assignments must be properly coordinated within ITU and registered with the ITU, inter alia, if: (a) international protection against harmful interference is desired, (b) the assignment will be used for international service, or (c) it is believed that the use of a new assignment will cause harmful interference.[45]

Ensuring Equitable Access: A Priori Planning

In order to ensure equitable access to radio frequencies and orbital positions, Article 44. 2 of the ITU Constitution specifies that "In using frequency bands for radio services, Member States shall bear in mind that radio frequencies and any associated orbits, including the geostationary-satellite orbit, are limited natural resources and that they must be used rationally, efficiently, and economically, in conformity with the provisions of the Radio Regulations, so that countries or groups of countries may have equitable access to those orbits and frequencies, taking into account the special needs of the developing countries and the geographical situation of particular countries." This Article introduced the concept of equity or equitable access with respect to the use or the sharing of the radio frequencies and orbital positions. Giving effect to this concept, the ITU has a priori allotted (distributed) radio frequencies and associated orbital positions among its Member States at World Radiocommunications Conferences. However, so far, it has only been implemented in connection with a limited number of allotment plans, e.g.: (a) provision and associated frequency allotment plan for the aeronautical mobile (OR) service in the bands allocated exclusively to that service[46]; (b) frequency allotment plan for the aeronautical mobile (R) service and related information[47]; (c) broadcasting-satellite service (BSS) operating in 12 GHz band and associated feeder links[48]; and (d) fixed-satellite service (FSS) operating in 6/4 GHz and 14/11 GHz bands,[49] etc. The rarity of such plans is attributable to the desire of ITU State Members to retain their freedom in the use of radio frequencies and orbital positions (Fig. 6).

Decision-Making Process

For the purpose of registration, the notifying Administration is required to send to the ITU the required information that is published by the organization. This process

---

[45] ITU Radio Regulations, Articles 11.1–11.8.

[46] ITU Radio Regulations, Appendix 26 (3,025 and 18,030 KHz).

[47] ITU Radio Regulations, Appendix 27.

[48] ITU Radio Regulations, Appendix 30A.

[49] ITU Radio Regulations, Appendix 30B.

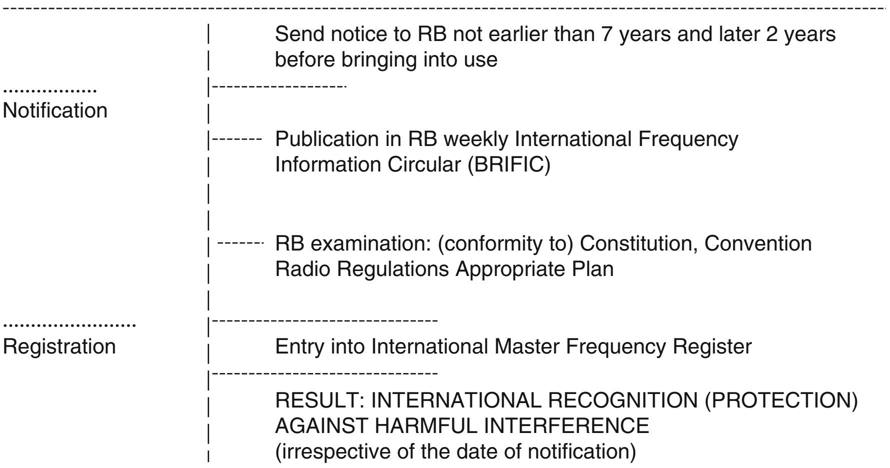


Fig. 6 An a priori plan procedure (applicable to BSS in 12 GHz band and FSS in 6/4 GHz and 14/11 GHz bands)

is called "advance publication." An Administration which considers that the newly notified and published satellite system might interfere with its already registered radio frequencies is entitled to object to the registration of the latter satellite system. In such cases, the notifying Administration may ask for coordination with the objecting Administration. The purpose of this advance publication is to give other Administrations the opportunity to look at the information and to communicate with the publishing Administration within a period of 4 months if there is a risk of interference.[50]

Applying Rules of Procedures, the RB examines the notifications with respect to their conformity to ITU Constitution, Convention, and the Radio Regulations (including Table of Frequency allocations, allotment plans, and probability of harmful interference). In this regard, if the findings of the RB are positive, it registers the notified assignments in the Master Register.

According to Article 14 of the ITU Radio Regulations, any Administration may request to the RB for a review of its finding(s) or any other decision(s) of the RB. If the outcome of the review does not successfully resolve the matter, or if it would adversely affect the interests of other Administrations, the RB is required to prepare a report and send it to the Administration which requested the review and to any others concerned in order to enable them, if they so desire, to address the RRB. The RB then sends the report with all supporting documentation to the RRB. The decision of the RRB on the review, to be taken in accordance with the Convention,[51]

---

[50] ITU Radio Regulations, Article 9.3.

[51] ITU Constitution, Article 14.2.

is regarded as final in so far as the RB and the RRB are concerned. However, if the Administration which requested the review disagrees with the RRB's decision, it may raise the matter at a World Radiocommunication Conference.

In order to reduce the abuse of regulatory processes,[52] the ITU has adopted several mechanisms, the first one of which fixes a period of 7 years for bringing into use a communication satellite system from the date of submission of the advance publication information.[53] "If, after the expiry of the period of 7 years from the date of receipt of the relevant complete information. . ., the administration responsible for the satellite network has not brought the frequency assignments to stations of the network into use, the corresponding (published) information. . ., shall be cancelled, but only after the administration concerned has been informed at least 3 months before the expiry date."[54] However, the ITU Radio Regulations do not place any time limitations upon the right of States to continue occupying radio frequencies and orbital slots after they have started using them. When the notifying Administration suspends the use of a recorded assignment to a communication satellite system for a period not exceeding 18 months, it is obliged to inform the RB about the date on which the suspended assignment will be brought back into regular use.[55] The RB is also entitled to review periodically the Master Register with the aim of maintaining or improving its accuracy.[56] Using this power, the RB can reduce (if not eliminate) the hoarding of recorded radio frequency assignments, which are not actually used. (See ITU Circular Letter CR/301, "Removal of unused frequency assignments (Space Services) from the Master Register," 1 May 2009).

Second: Each notifying Administration is required to provide evidence of seriousness of its intention of establishing a satellite network. ITU Radio Regulations Resolution 49,[57] Annex 2, requires that the following information be provided to the ITU at the time of submission of information of radio frequency assignments:

- Identity of the satellite network: (a) identity of the satellite network, (b) name of the Administration, (c) country symbol, (d) reference to the advance publication information, (e) reference to the request for coordination, (f) frequency band(s), (g) name of the operator, (h) name of the satellite, and (i) orbital characteristics

---

[52] Particularly to reduce the registration of so-called paper or virtual satellites, the notifications for which are filed with the ITU without any serious plans for the acquisition and launch of these satellites.

[53] ITU Radio Regulations, Article 11.44.

[54] ITU Radio Regulations, Article 11.48.

[55] ITU Radio Regulations, Article 11.49.

[56] ITU Radio Regulations, Article 11.50.

[57] Entitled "Administrative due diligence applicable to some satellite radiocommunication services."

- Spacecraft manufacturer: (a) name of the spacecraft manufacturer; (b) date of execution of the contract; (c) contractual "delivery window" (planned period, beginning, and end dates); and (d) number of satellites procured
- Launch services provider: (a) name of the launch vehicle provider, (b) date of execution of the contract, (c) anticipated launch or in-orbit delivery window, (d) name of the launch facility, and (e) location of the launch facility

At the 2007 World Radiocommunication Conference, the filing with the ITU by the notifying Administration of the required documents showing seriousness of its planned satellite system was made mandatory. If this is not completely done in a timely fashion, the satellite notified to ITU for registration "shall no longer be taken into account and shall not be recorded in the MIFR" (Master International Frequency Register).[58]

Third: In order for the ITU to recover its processing costs for the communication satellite networks, each notifying Administration is required to pay "filing fee." This measure is a market mechanism in line with the "user-pay" principle so that ITU is in a position to recover administrative expenses from the users of the radio frequencies and orbital slots. The 1998 ITU plenipotentiary conference held in Minneapolis adopted a decision to charge for all satellite filings received by ITU after 7 November 1998. This decision was implemented by the Council in its Decision 482 (adopted at its 2002 session), which has been further modified several times. The ITU has started charging a nonrefundable filing fee per satellite network.

The revised Article 9.2B1 of the ITU Radio Regulations now provides that "If the payments are not received (by the ITU) in accordance with the provisions of Council Decision 482, as amended, on the implementation of cost recovery for satellite network filings, the (Radiocommunication) Bureau shall cancel the publication, after informing the Administration concerned. The Bureau shall inform all Administrations of such action, and that the network specified in the publication in question no longer has to be taken into consideration by the Bureau and other Administrations." Almost every year, a number of communication satellites filings are canceled as a result of nonpayment of ITU processing fee invoices.[59] It should be noted that

---

[58] ITU Radio Regulations, Resolution 49, Annex 1, para. 11.

[59] For example: OPTOS satellite system (see Director, Radiocommunication Bureau, Report to the 56th meeting of the Radio Regulations Board, ITU Document: RRB11-1/3-E, Annex 4, of 1 March 2011); LARKSAT-IORR, LARKSAT-AOR2R, LARKSAT-NAR, LARKSAT-PORR, INSAT-NAV-A-GS, SWANSAT-3A, THAICOM-LS2, and THAICOM-LS3 satellite systems (see Director, Radiocommunication Bureau, Report to the 55th meeting of the Radio Regulations Board, ITU Document : RRB10-3/4-E, Annex 4, of 29 October 2010); INSAT-EXC55E (see Director, Radiocommunication Bureau, Report to the 50th meeting of the Radio Regulations Board, ITU Document: RRB09-1/1-E, Annex 4, of 4 February 2009); and GOES-89.5W and GOES-105W satellite systems (see Director, Radiocommunication Bureau, Report to the 48th meeting of the Radio Regulations Board, ITU Document: RRB08-3/3-E, Annex 4, of 4 August 2008).

only those satellites that have been properly registered with ITU are entitled to protection against harmful interference.

Problem of Interference

Irrespective of the efforts made by Administrations and the ITU for the avoidance of harmful interference, problems of interference often arise. Article 1.166 of the ITU Radio Regulations defines "harmful interference" as "the effect of unwanted energy due to one or a combination of emissions, radiations, or inductions upon reception in a radiocommunication system, manifested by any performance degradation, misinterpretation, or loss of information which could be extracted in the absence of such unwanted energy."

There exists no compulsory international dispute settlement machinery within the ITU with respect to the resolution of interference problems. Under the ITU agreements, dispute resolution is dealt with under Article 56 of the ITU Constitution. According to this Article, Member States may settle their disputes on questions relating to the interpretation or application of the ITU Constitution, of the Convention or of the Administrative Regulations (including Radio Regulations governing space communications) by negotiation, through diplomatic channels, or according to procedures established by bilateral or multilateral treaties concluded between them for the settlement of international disputes, or by any other method mutually agreed upon. If none of these methods of settlement is adopted, any Member State party to a dispute may have recourse to arbitration in accordance with the arbitration procedure as specified in Article 41 of the ITU Convention. Member States have also concluded an Optional Protocol on the Compulsory Settlement of Disputes Relating to the ITU regulatory regime[60] which is applicable among Member States parties to that protocol. This protocol essentially makes the arbitration procedure as defined in Article 41 of the ITU Convention compulsory for the settlement of disputes among the States Parties to the protocol, numbering 64 at present. In practice, neither Article 56 of the ITU Constitution, nor Article 41 of the ITU Convention, nor the Optional Protocol has ever been used. Therefore, all the harmful interference problems have been, and are, resolved according to the provisions of Article 15 of the ITU Radio Regulations.

---

[60] As of February 2011, there are 64 States Parties to this protocol. They are Australia, Austria, Bahrain, Barbados, Belarus, Belgium, Belize, Benin, Bosnia and Herzegovina, Botswana, Canada, Chile, Colombia, Congo (Rep. of the), Cyprus, Denmark, Egypt, El Salvador, Estonia, Finland, Greece, Guinea, Iceland, Ireland, Italy, Japan, Jordan, Kenya, Kiribati, Korea (Rep. of), Kuwait, Lao P.D.R., Latvia, Libyan Arab Jamahiriya, Liechtenstein, Lithuania, Luxembourg, Madagascar, Malta, Mauritius, Mexico, Monaco, Netherlands, New Zealand, Oman, Panama, Peru, Philippines, Portugal, San Marino, Slovenia, South Africa, Sudan, Sweden, Switzerland, Togo, Tunisia, Turkey, United Arab Emirates, United Kingdom, Uruguay, Uzbekistan, Vietnam, and Zimbabwe. http://www.itu.int/cgibin/htsh/mm/scripts/mm.final-acts.list?_languageid=1&_agrmts_type=PROT-92.

Under Article 15 of Radio Regulations, cases of harmful interference are resolved exclusively through bilateral negotiations between the concerned Administrations who are obliged to exercise the utmost goodwill and mutual assistance in the application of the provisions of the ITU Constitution and Radio Regulations to the settlement of problems of harmful interference. In this regard, if considered appropriate, the concerned States may seek the administrative support of the RB. It is the RB which acts as an executive arm of the RRB conducting investigations into harmful interference allegations and registering frequency assignments. The RRB provides recommendations to the concerned Administrations in cases of harmful interference after a report has been received from the director of the RB.[61]

Frequency jamming amounts to intentionally caused interference which is illegal. Article 45 of ITU Constitution specifies that "All stations, whatever their purpose, must be established and operated in such a manner as not to cause harmful interference to the radio services or communications of other Members. . . which carry on a radio service, and which operate in accordance with the provisions of the Radio Regulations." In addition, according to Article 15 of ITU Radio Regulations, "(1) All stations are forbidden to carry out unnecessary transmissions, or the transmission of superfluous signals. . .. (2) Transmitting stations shall radiate only as much power as is necessary to ensure a satisfactory service." Jamming could also have serious consequences for national and international security because it is considered to be an unfriendly act or a sort of war action. The incidence of possible interference with, or jamming of, radio signals of satellites has been well-known in the operation of both civilian and military satellites, including global navigation satellite systems, like the Russian GLONASS and American Global Positioning System (GPS) (Butsch 2011). For instance, in 2009 and 2010, Iran intentionally jammed television and radio signals transmitted by the EUTELSAT communication satellite system to Europe and the Middle East (Brown 2010).

The ITU does not possess any mechanism or power of enforcement or imposition of sanctions against the violators of its rules, regulations, and processes. It is true that the voluntary compliance approach has worked well in the past and that States have largely been following the ITU rules mainly due to the fact that noncompliance would not augur well for their own individual and collective self-interest. It is doubtful whether this tradition will work well in the future as the number of State and non-State players has been increasing and the competition for scarce resources is becoming severe.

In order to diagnose the problem of harmful interference and its elimination or reduction, it is imperative to have an appropriate monitoring system. Currently, there exists no independent international monitoring system. Under Article 16 of the ITU

---

[61] Article 10.2 of the ITU Convention of the International Telecommunication Union (Geneva, 1992) as amended by the Plenipotentiary Conferences in 1994, 1998, 2002, 2006, and 2010.

Radio Regulations, Administrations have agreed to continue the development of their monitoring facilities. These stations may be operated by an Administration or, in accordance with an authorization granted by the appropriate Administration, by a public or private enterprise, by a common monitoring service established by two or more countries, or by an international organization. However, the so-called international monitoring system *currently* comprises *only* those national monitoring stations which have been so nominated by Administrations in the information sent to the ITU Secretary-General. Lack of independent international monitoring system(s) inhibits the availability of objective information which is critical in unbiased settlement of disputes related to harmful interference.

Conclusion

More and more telecommunication satellites are being, and will be, established worldwide. This means that there is, and will be, increased pressure on the already scarce and seriously congested radio frequency spectrum without which no radio telecommunications system can be operated. The competition for very limited radio frequencies and also orbital positions will only grow fierce as the demand for telecommunications increases. Costs for accessing and using these resources to operators and regulators are expected to amplify as access becomes difficult and cases of harmful interference increase.

Good faith has been the main basis for the implementation of the ITU procedures and regulations (there are no sanctions against violations). In recent years, due to rapidly increased demands and competition among applicants, ITU’s coordination and registration procedures have been abused. Thus, the ITU registration processing system is seriously clogged and it takes a few years for an application to get processed. The problem of orbital congestion is particularly significant in the geostationary orbit. The above-discussed three steps taken by the ITU in order to reduce the registration of “paper or virtual satellites” may be small but are important developments for enhancing the effectiveness of the ITU regulations and processes. It is in the interest of all communication satellite operators and States to comply fully with the ITU regulatory regime and processes.

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Economics and Financing of Communications Satellites
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends

- ▶ Satellite Communications and Space Telecommunication Frequencies
- ▶ Satellite Communications Antenna Concepts and Engineering
- ▶ Satellite Communications Overview
- ▶ Satellite Communications: Regulatory, Legal, and Trade Issues
- ▶ Satellite Orbits for Communications Satellites
- ▶ Satellite Radio Communications Fundamentals and Link Budgets
- ▶ Satellite Spectrum Allocations and New Radio Regulations from WRC-15: Defending the Present and Provisioning the Future
- ▶ Space Telecommunications Services and Applications
- ▶ Trends and Future of Satellite Communications

References

P.J. Brown, Iranian in a jam over satellite blocking, ATIMES (2010), http://www.atimes.com/atimes/Middle_East/LC25Ak04.html.

F. Butsch, GPS and GLONASS radio interference in Germany (2011), http://elib.uni-stuttgart.de/opus/volltexte/1999/278/pdf/278.pdf.

Satellite Spectrum Allocations and New Radio Regulations from WRC-15: Defending the Present and Provisioning the Future

Audrey L. Allison

Contents

Abstract

The International Telecommunication Union (ITU) convenes World Radiocommunication Conferences (WRC) for the purpose of concluding a treaty on emergent issues related to the operation of radio-based systems. This effort includes allocation of radiofrequency spectrum and procedures for accessing the orbit for satellites and results in amendments to the international Radio Regulations. The 2015 World Radiocommunication Conference (WRC-15) took place on November 2–27 in Geneva, Switzerland, with an agenda that included some forty topics. One of the key themes of WRC-15 was competition to access scarce radio spectrum resources, while also finding a way to enable introduction of innovative new services and technologies.

WRC-15 featured the latest campaign in the ongoing confrontation between the mobile telephony/broadband industry and the satellite industry over spectrum

---

A.L. Allison (✉)
International Space University, Strasbourg, France
e-mail: Audrey.allison@boeing.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_97

resources. The objective was to grant access to premium regional and global spectrum allocations to enable extension of desired terrestrial services without harming established satellite services providing lifeline connectivity and other important connections. However, this spectrum duel did not end up being the defining issue of WRC-15 as there were so many fractious issues that deeply divided the proceedings. But, in the end, the conference found a way forward on every issue and approved plans for its next proceedings in 2019 and 2023.

As described in this chapter, the satellite industry, led by major industry players and fueled by aspiring newcomers, not only defended its essential spectrum resources but accomplished key regulatory improvements to pave the way for future innovation – including new spectrum access, lifting constraints on mobile applications by satellite, and preparing the way for newly announced non-geostationary satellite systems. WRC-15 was thus a banner conference for the satellite industry.

Keywords

International Mobile Telecommunications (IMT) • International Telecommunication Union (ITU) • ITU Development Sector (ITU-D) • ITU Radiocommunication Sector (ITU-R) • Radiocommunication Bureau (BR) • Mobile Satellite Service (MSS) • 2015 World Radiocommunication Conference • World Radiocommunication Conference (WRC) • Radio Regulations • Spectrum allocations • Earth Stations on Mobile Platforms (ESOMPs) • Earth Stations in Motion (ESIM) • Fixed-Satellite Service (FSS) • Non-geostationary satellite orbit (non-GSO) • Unmanned Aerial Systems (UAS)

Introduction

Every 4 years or so, the International Telecommunication Union (ITU) convenes a World Radiocommunication Conference (WRC) for the purpose of concluding a treaty on emergent issues relating to the operation of radio-based systems, including satellites. WRCs amend the international Radio Regulations on the use of radiofrequency spectrum and satellite orbits. As the United Nations (UN) agency uniquely charged with the responsibility of harmonizing and coordinating the planet's use of the shared natural resources of radio waves and orbits, the ITU, together with its government and private sector members, is constantly at work creating international law and standards to accommodate new space systems and services within the congested radio environment and crowded orbits.

The most recent WRC took place on November 2–27, 2015, in Geneva, Switzerland, the ITU's headquarters just catty-corner from the European Headquarters of the UN, the original home of the League of Nations. WRC-15s significance is evident from its broad attendance – 3,300 delegates from 163 Member States and more than 80 companies (not counting those that attended as part of Member State delegations). Its agenda included more than forty topics covering a broad range of radio services, from maritime to aeronautical, amateur radio to broadcasting, and satellite to space sciences. The theme of WRC-15, like others before it, can be described as competition over access to scarce shared resources while finding a way both to enable introduction of new services and technologies for the benefit of mankind and protecting operation of established, incumbent radio operations.

As expected, WRC-15 featured the latest campaign in the sustained confrontation over spectrum resources between the voracious terrestrial mobile broadband industry and the more mature satellite industry. The rapidly growing spectrum needs of the mobile industry fueled its efforts to gain access to large swaths of radio frequencies that had been allocated on a global basis to satellite services by previous WRCs.[1] However, this issue that has defined recent World Radio Conferences did not end up being the focus of WRC-15, as several unexpectedly fractious issues arose that served to deeply divide the proceedings as proponents of competing innovative technologies clashed with countries preferring to maintain the status quo and to slow the rapid pace of change. In the end, after four long weeks including evening and weekend sessions, the conference found a path forward on every issue and approved plans for its next proceedings in 2019 and 2023.

As will be described in this chapter, the satellite industry, led by major industry players and fueled by newcomers with big ideas reminding us of the heyday of the 1990s, not only defended its current access to essential spectrum resources but built a pathway for future growth and innovation, including new spectrum allocations, liberalized regulations to allow new services, and measures to support newly proposed non-geostationary satellite orbit (non-GSO) systems. Thus, WRC-15 proved to be a banner conference for the satellite industry and for the governments and the people who rely on satellite services.

WRC-15 Results for the Satellite Industry

Upon the conclusion of WRC-15, the global satellite industry announced:

> World Radiocommunication Conference 2015 decides satellite spectrum is central to future vision for global connectivity; long-term delivery of innovative satellite services are assured a pivotal role alongside wireless and other complementary technologies. (Satellite Spectrum Initiative, 27 November 2015)

While the satellite industry, with the backing of like-minded governments, combined forces to successfully defeat the mobile industry's campaign to reallocate satellite spectrum to the terrestrial mobile service, it also made a stunning number of

---

[1] An allocation is an: "Entry in the Table of Frequency Allocations of a given frequency band for the purpose of its use by one or more terrestrial or space *radiocommunication services* or the *radio astronomy service* under specified conditions. This term shall also be applied to the frequency band concerned." International Telecommunication Union: Radio Regulations, No. 1.16, International Telecommunication Union, Geneva (2012)

Fig. 1 One of the satellite industry's promotional materials used in the campaign to defend C-band satellite spectrum allocations at WRC-15 (Intelsat, 2015)

advances, both in terms of obtaining additional spectrum resources and in liberalizing regulatory constraints on delivery of services to mobile platforms, including ships and aircraft (manned and unmanned). Moreover, the results included studies to support future spectrum resources and to facilitate the implementation of newly proposed non-GSO constellations. Indeed, there were so many actions taken to benefit the satellite and space communities than can be fully described in the space of a single chapter. Thus, this discussion will focus on key highlights and major impacts to the communications satellite community (Fig. 1).

Successful Defense of Satellite Spectrum Allocations

The prime agenda item of WRC-15, Agenda Item 1.1, considered the allocation of spectrum for the mobile service and identification of additional spectrum resources below 6 GHz for "International Mobile Telecommunications" (IMT) within that allocation.[2] IMT is the ITU's overarching and evolving term for a framework of standards for mobile telephony within the mobile service, such as 4G (fourth generation) and Long Term Evolution (LTE). At WRC-15, the world was unified on the need to address burgeoning requirements for additional spectrum resources to

---

[2] IMT is an application of the mobile service. Thus, to reserve a spectrum in a mobile service allocation for IMT usage, the frequency band is "identified" for IMT via footnote to the Table of Frequency Allocations in the Radio Regulations.

power the world's growing reliance on wireless broadband for smartphones, tablets, and the like. The core challenge was agreeing upon which spectrum resources to repurpose in light of the current and planned use of the subject frequency bands and the difficulty of sharing spectrum between the incumbent operations and the ubiquitous, nomadic, and relatively high-powered IMT services, without resulting in harmful interference. Earlier technical studies had demonstrated that co-frequency operation of IMT and satellite downlinks in the C-band (3.4–4.2 GHz) is infeasible due to saturation of Earth station receivers listening for distant satellite signals in the presence of stronger terrestrial signals. Frequency bands supporting a wide range of industries and government users were studied during the 5-year WRC-15 preparatory period for potential reallocation and identification.

A second major front of this spectrum battle at WRC-15 was Agenda Item 10, under which the conference was to recommend items for the agenda of the next WRC, expected to be in 2019, as well as to give views on the preliminary agenda for the subsequent conference in 2023. The priority item for most Member States in the preparation of the draft WRC-19 agenda was to study frequency bands above 6 GHz for the next-generation IMT, "5G," now on the drawing board. Again, satellite allocations were among those in the cross-hairs for these future fifth-generation terrestrial systems. Not until the final hours of the conference was this matter settled and the Ka-Band satellite allocation (27.5–30 GHz) removed from the list of bands to be studied for IMT for consideration at WRC-19. Both of these issues will be described below.

C-Band. The last two WRCs, in 2007 and 2012, addressed the issue of IMT use of globally harmonized C-Band satellite downlink spectrum. The physical properties of this band are unique among FSS allocations. C-band provides the capability of using broad intercontinental beams for worldwide coverage from just three spacecraft. C-band also features the best resistance to rain-fade, of particular importance in the tropics and other rainy regions of the world. C-band satellites are ideal for point-to-multipoint transmissions such as video distribution. They also are relied upon throughout the world to provide basic connectivity for remote and rural and especially tropical areas, such as the Pacific Island nations, who are among the most vocal defenders of this satellite allocation. C-band also supports a tremendous variety of other uses, including feeder link operations of MSS networks, backhaul for cellular telephones in developing countries, services to ships, tracking telemetry and control, and disaster recovery, among many others.

In 2007, the WRC was marked by a fiercely divisive battle over C-band. IMT proponents raised great excitement over the promise of delivery of affordable telecommunications services directly to consumers using the globally harmonized C-band allocation, with resulting benefits of economic and social development, not to mention revenues to governments from spectrum auctions and regulatory fees. C-band satellite services were portrayed as old-school, moribund, and no longer needed, in light of the satellite industry's exploitation of satellite spectrum allocations in higher frequency bands. After multiple all-night sessions during the final week of that 4-week conference, the IMT takeover of C-band was largely avoided. WRC-07 retained the satellite C-band allocation with "no change." However, 81 Member States broke from the decision and opted to place their name in country footnotes to the Table of Frequency Allocations declaring their intention to operate IMT within their territories, subject to coordination with their neighbors to avoid causing harmful interference to their lawful operations, including avoidance of disruptions to reception to Earth stations.

In 2012, the WRC's discussions of the proposed agenda for WRC-15 proved difficult. WRC-12 developed WRC-15 Agenda Item 1.1 which included further study of the C-band for potential IMT identification.

Following 3 years of additional studies and further entrenchment of positions by both sides, the C-band matter proved again to be stubbornly controversial. A satellite industry coalition, the "Satellite Spectrum Initiative," led by Intelsat, SES, Inmarsat, Eutelsat, and others, embarked on a global advocacy campaign to alert and educate regulators, customers, suppliers, and relief agencies as to the real threat of loss of the satellite C-band services. It succeeded in convincing most governments of the need to retain satellite access to the band to ensure continuity of these needed services (Fig. 2).

WRC-15 concluded with the decision for "no change" to the global satellite allocation in the upper C-Band (3600–4200 MHz) – a major win for the satellite industry and for the people who rely on satellite services. In Regions 1 and 2, however, the lower portion of the band (3400–3600 MHz) was allocated to the mobile service and identified for IMT. In some countries in these regions, the lower portion of the band had already been domestically repurposed for mobile services although still supporting limited international satellite services. Although "no change" was made to the FSS allocation in Region 3 across the entire C-band, a handful of countries added their names to a footnote to the International Table of Frequency Allocations expressing their interest in IMT use of this spectrum within their borders, but with cross-border power-flux density limits and an indication that this required

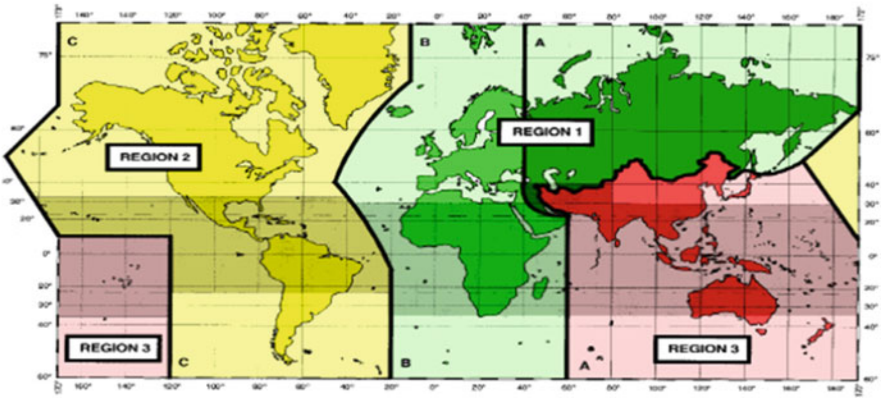


Fig. 2 The ITU's three radio regions, as defined in **No.** 5.2 of the Radio Regulations. Region 1 comprises Europe, Africa, and the Middle East; Region 2 contains the Americas; and Region 3 is Asia-Pacific. (Radio Regulations 2012)

coordination with neighboring countries. Four Region 2 countries (Canada, United States, Colombia, and Costa Rica) also entered into a new footnote containing an IMT identification in the sub-band 3600–3700 MHz, which matches domestic implementation of terrestrial services in their territories. The footnote also includes regulatory provisions for protection of receiving FSS Earth stations in neighboring countries.

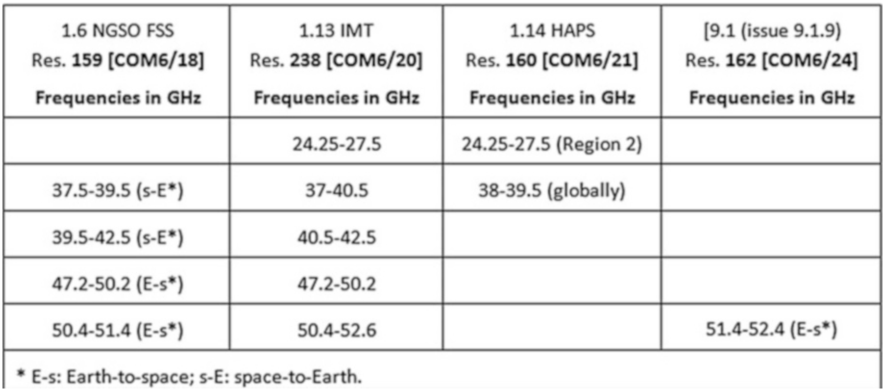
Over the course of three WRCs, the ITU managed to forge a consensus over the evolving use of the C-band. This judicious solution fortunately left much of the global satellite allocation intact while allowing the requirements of another desired service to be met. The C-band satellite uplink allocation 5925–6425 MHz was not touched by the WRC. GSMA, the mobile industry's trade association, observed:

> We welcome the decisions taken at WRC-15 to identify critical new spectrum to secure the future of the mobile internet. After weeks of intense treaty negotiations, governments agreed three new globally harmonised spectrum bands, representing a major step forward in meeting the growing demand from citizens worldwide for mobile broadband. Global harmonisation of spectrum bands through the WRC process is key to driving the economies of scale needed to deliver low-cost, ubiquitous mobile broadband to consumers around the globe. The GSMA applauds the strong support from governments in all regions for the global harmonisation of 200 MHz of the C-band (3.4–3.6GHz) to meet capacity requirements in urban areas. (GSMA Press Release, November 27, 2015)

Ka-Band. The complement to WRC-15 Agenda Item 1.1 was the consideration of a follow-on agenda item for the next WRC to identify spectrum to accommodate the next generation of mobile services – the so-called Fifth Generation or 5G – in higher frequency ranges above 6 GHz. Every region of the world submitted proposals to WRC-15 to adopt a 5G item on the WRC-19 Agenda. After the frustrations and difficulties experienced in completing sharing studies and proposals for WRC-15 on Agenda Item 1.1, which had not limited the specific frequency bands for study and consideration, Member States were determined to more carefully delineate the bounds of the next IMT conference item in order to structure a better and less costly preparatory effort for the next conference.

Region 2 (led by the United States) joined by some major voices from Europe (i.e., Sweden) and Asia (i.e., Korea and Japan) sought to include consideration of satellite Ka-band frequencies (27.5–30 GHz) in the studies for future 5G spectrum at the 2019 World Radio Conference. The inclusion of this frequency band in the 5G study proved to be exceedingly contentious and prolonged the consideration of the WRC-19 Agenda until the conference's very end. The satellite industry and concerned governments lobbied hard to keep this satellite spectrum off the table in light of the billions of dollars of recent investment in Ka-band satellites currently being built and launched. These include high-throughput satellites such as Intelsat's Epic, Inmarsat's Global Xpress, EchoStar's Jupiter, and ViaSat, all designed provide broadband services to the world.

WRC-15 finally adopted Resolution 238, which identifies more than 30 MHz of spectrum for study for IMT in bands between 24.25 and 86 GHz. Although the Ka-band satellite spectrum was not included in the final list of frequency bands for



1.6 NGSO FSS Res. **159 [COM6/18]** **Frequencies in GHz**	1.13 IMT Res. **238 [COM6/20]** **Frequencies in GHz**	1.14 HAPS Res. **160 [COM6/21]** **Frequencies in GHz**	[9.1 (issue 9.1.9) Res. **162 [COM6/24]** **Frequencies in GHz**
	24.25-27.5	24.25-27.5 (Region 2)	
37.5-39.5 (s-E*)	37-40.5	38-39.5 (globally)	
39.5-42.5 (s-E*)	40.5-42.5		
47.2-50.2 (E-s*)	47.2-50.2		
50.4-51.4 (E-s*)	50.4-52.6		51.4-52.4 (E-s*)

* E-s: Earth-to-space; s-E: space-to-Earth.

Fig. 3 Overlapping frequency bands planned for being simultaneously studied for satellite (geostationary and non-GSO FSS), IMT, and High Altitude Platform Station allocations or identification by WRC-19 (Radiocommunication Bureau, Results of the first session of the Conference Preparatory Meeting for WRC-19, 2015)

study, a few countries (Japan, Korea, United States, Sweden, Finland, and Colombia) nevertheless expressed plans to study the band for 5G implementation in other fora despite the ITU's failure to act in this regard. Indeed, the United States Federal Communication Commission has already launched a rulemaking proceeding to add a domestic mobile allocation in the 27.5–28.35 GHz band and other satellite bands to accommodate future 5G, a move which is being vigorously opposed by the satellite industry Federal Communications Commission (2015) (Fig. 3).

Notably, the new WRC-19 Agenda Item for IMT also includes study of satellite allocations in the higher V-band frequency range, namely, 37.5–40.5 GHz, 40.5–42.5 GHz, 47.2–50.2 GHz, and 50.4–51.4 GHz. As described further below, these bands are also being studied for additional satellite services. The 51.4–52.4 GHz band is further slated for study for possible use by High Altitude Platform Station (HAPS) Resolutions 159 and 162 (WRC-15). WRC-19 will assess the results of the studies that take place in the ITU-R Study Groups and Working Parties during the next study period and make the final decision as to the ultimate use of these frequency bands.

Regulatory Changes to Promote Innovation of Satellite Services

WRC-15 was also marked by the adoption of decisions to alleviate regulatory constraints discouraging use of satellites in the Fixed-Satellite Service (FSS) from providing services to Earth stations in motion. This particular topic has been at issue since at least 2000 when "Earth Stations on Vessels" were proposed to utilize C-band FSS satellites, rather than satellites operating in the mobile-satellite service (MSS), to provide service to ships in port (at fixed points) and at sea. Notably, there is no MSS allocation in C-band and one would be difficult to add as the band is widely shared with terrestrial fixed services. Use of FSS for mobility applications proved similarly controversial in 2003 when the conference considered and ultimately approved the use of Ku-Band FSS satellites for correspondence with Earth stations onboard aircraft operating in the Aeronautical Mobile Satellite Service. The challenge arose again in 2012, when the conference was unable to reach a conclusion on the use of FSS to support command and control links to unmanned aircraft.

At root, the difficulty lies in the long-standing definition of the Fixed-Satellite Service, which defines FSS as service to Earth stations at fixed points. Mobile Earth stations are intended to be used while in motion but operate in the Mobile Satellite Service. These regulatory definitions were based on the state-of-the art technology that existed when these definitions were adopted many years ago. These precise and now dated definitions have remained constant, while technology has continued to evolve, creating a growing disconnect between treaty-level regulations versus actual usage. However, there is a great reluctance to reopen these long-standing definitions that are foundational to the Radio Regulations and its Table of Frequency Allocations. Thus, these new services are accommodated on an exceptional, one-off basis (Fig. 4).

Earth Stations in Motion. High-throughput satellites (HTS) operating in Ka-Band FSS allocations have recently begun to deliver broadband services to airplanes and ships. This FSS capability is the product of a host of technology advancements – including developments in satellite manufacturing, solar panel efficiency, satellite antenna, and Earth station technology accompanied by changes

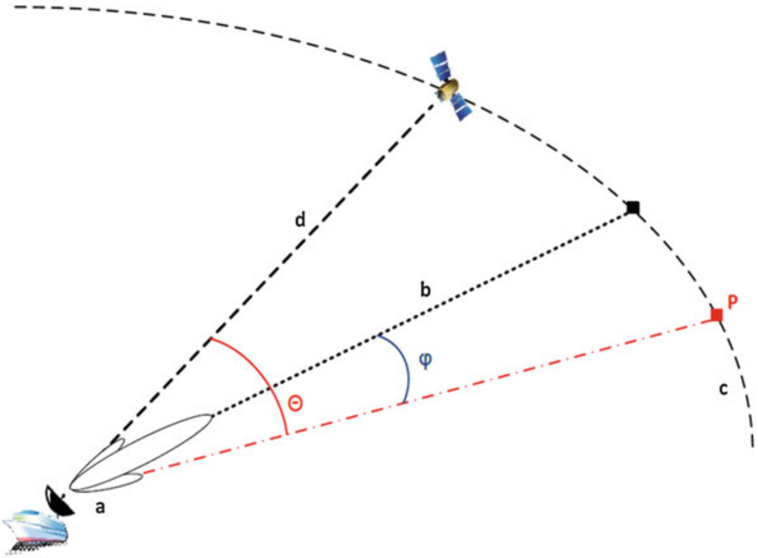


Fig. 4 WRC-15 approved an agenda item for 2019 addressing the use of FSS space stations in the Ka-band to provide service to Earth stations on airplanes, ships, and other moving platforms (Resolution 158 (WRC-15))

in the launch industry. These new technical advances allow these HTS networks to meet growing market demands for mobile broadband solutions. Moreover, the highly directional, multi-axis, stabilized Earth station antennas that are capable of maintaining a very high degree of pointing accuracy on rapidly moving platforms make it possible for FSS networks to provide mobility services within the FSS regulatory parameters governing their operations and the existing coordination agreements with their neighbors in the geostationary orbit. In the ITU, these new mobile applications are variously referred to as Earth Stations on Mobile Platforms (ESOMPs) or the more elegant Earth Stations in Motion (ESIM). The challenge is that these satellites operate in frequencies allocated to FSS and they are coordinated as FSS networks. Only some portions of these bands are also allocated to MSS. One particular range of the FSS allocations was further constrained by a footnote provision of the Radio Regulations which limited service to ESIMs to those bands also allocated to MSS (Fig. 5).

At WRC-15, the Americas and European regional groups submitted proposals to seek regulatory clarity for ESOMPs operations in the 29.5–30 GHz/19.7–20.2 GHz bands. While there are global FSS allocations in these bands, the allocation is limited by footnote, No. 5.526, which provides:

> **5.526** In the bands 19.7-20.2 GHz and 29.5-30 GHz in Region 2, and in the bands 20.1-20.2 GHz and 29.9-30 GHz in Regions 1 and 3, networks which are both in the fixed-satellite service and in the mobile-satellite service may include links between earth stations at specified or unspecified points or while in motion, through one or more satellites for point-to-point and point-to-multipoint communications.

As can be seen from this example from the table, the MSS allocation in these ranges is not uniform, casting doubt on use of the availability of the entire 500 MHz in each direction to serve ESIMs.

Although satellites with mobile service capability were already being built, launched, and coordinated, the ESIMs issue was not an item on the WRC-15 agenda. ITU-R studies on the technical and operational requirements for ESIMS Earth station operations were well underway, but had not been finally agreed before the conference. The Radiocommunication Bureau had helpfully responded to requests for coordination of these networks with the development of a new class of station and notice that its findings would be based on use of the existing criteria for FSS links in the relevant bands. The Director's report to WRC-15 observed these difficulties in applying the Radio Regulations and invited the conference to consider approaches for accommodating ESOMPs in the subject frequency bands. Thus, WRC-15 was able to address the issue under Agenda Item 9.1, under which each WRC is tasked to consider and approve the report of the Director of the Radiocommunication Bureau on activities of the Radiocommunication Sector since the previous WRC on any difficulties or inconsistencies encountered in the application of the Radio Regulations.

WRC-15 decided not to change No. 5.526 but, instead, it added a new footnote to the FSS allocation in the bands 19.7–20.2 GHz and 29.5–30.0 GHz to clarify that Earth stations in motion may operate in these bands subject to a number of

18.4-22 GHz

<table>
<tr><th colspan="3">Allocation to services</th></tr>
<tr><th>Region 1</th><th>Region 2</th><th>Region 3</th></tr>
<tr><td colspan="3">18.4-18.6
FIXED
FIXED-SATELLITE (space-to-Earth) 5.484A 5.516B
MOBILE</td></tr>
<tr><td>18.6-18.8
EARTH EXPLORATION-SATELLITE (passive)
FIXED
FIXED-SATELLITE (space-to-Earth) 5.522B
MOBILE except aeronautical mobile
Space research (passive)
5.522A 5.522C</td><td>18.6-18.8
EARTH EXPLORATION-SATELLITE (passive)
FIXED
FIXED-SATELLITE (space-to-Earth) 5.516B 5.522B
MOBILE except aeronautical mobile
SPACE RESEARCH (passive)
5.522A</td><td>18.6-18.8
EARTH EXPLORATION-SATELLITE (passive)
FIXED
FIXED-SATELLITE (space-to-Earth) 5.522B
MOBILE except aeronautical mobile
Space research (passive)
5.522A</td></tr>
<tr><td colspan="3">18.8-19.3
FIXED
FIXED-SATELLITE (space-to-Earth) 5.516B 5.523A
MOBILE</td></tr>
<tr><td colspan="3">19.3-19.7
FIXED
FIXED-SATELLITE (space-to-Earth) (Earth-to-space) 5.523B, 5.523C, 5.523D 5.523E
MOBILE</td></tr>
<tr><td>19.7-20.1
FIXED-SATELLITE (space-to-Earth) 5.484A 5.516B
Mobile-satellite (space-to-Earth)
5.524</td><td>19.7-20.1
FIXED-SATELLITE (space-to-Earth) 5.484A 5.516B
MOBILE-SATELLITE (space-to-Earth)
5.524 5.525 5.526 5.527 5.528 5.529</td><td>19.7-20.1
FIXED-SATELLITE (space-to-Earth) 5.484A 5.516B
Mobile-satellite (space-to-Earth)
5.524</td></tr>
<tr><td colspan="3">20.1-20.2
FIXED-SATELLITE (space-to-Earth) 5.484A 5.516B
MOBILE-SATELLITE (space-to-Earth)
5.524 5.525 5.526 5.527 5.528</td></tr>
<tr><td colspan="3">20.2-21.2
FIXED-SATELLITE (space-to-Earth)
MOBILE-SATELLITE (space-to-Earth)
Standard frequency and time signal-satellite (space-to-Earth)
5.524</td></tr>
</table>

Fig. 5 Table of Frequency Allocations for the bands 18.4–21.2 GHz (primary allocations are indicated in upper case letters; secondary allocations in lower case) (Radio Regulations 2012)

conditions provided in new Resolution 156 (WRC-15), "Use of the frequency bands 19.7–20.2 GHz and 29.5–30.0 GHz by Earth stations in motion communicating with geostationary space stations in the fixed-satellite service." These conditions include: operation within the envelope of coordination agreements, permanently monitoring and controlling Earth station operations by a Network Control and Monitoring Centre, having the capability of limiting operations over territories that have not authorized their operations, and maintaining a point of contact for the purpose of tracing any cases of possible harmful interference. Furthermore, these Earth stations may not be used for "safety-of-life" applications. As these conditions are similar to those that would typically be included in an operator's license from domestic regulatory authorities, this approach was completely acceptable and a notable step forward for the satellite industry.

In addition to obtaining needed regulatory relief for these particular segments of the Ka-band, ESIMs proponents obtained an agenda item for WRC-19, "to consider the use of the frequency bands 17.7–19.7 GHz (space-to-Earth) and 27.5–29.5 GHz (Earth-to-space) by Earth stations in motion communicating with geostationary space stations in the fixed-satellite service and take appropriate action, in accordance with Resolution 158 (WRC-15)." This resolution invites the ITU-R:

> 1. to study the technical and operational characteristics and user requirements of different types of Earth stations in motion that operate or plan to operate within geostationary FSS allocations in the frequency bands 17.7–19.7 GHz and 27.5–29.5 GHz, including the use of spectrum to provide the envisioned services to various types of Earth station in motion and the degree to which flexible access to spectrum can facilitate sharing with services identified [above];
> 2. to study sharing and compatibility between Earth stations in motion operating with geostationary FSS networks and current and planned stations of existing services allocated in the frequency bands 17.7–19.7 GHz and 27.5–29.5 GHz to ensure protection of, and not impose undue constraints on, services allocated in those frequency bands....
> 3. to develop, for different types of Earth stations in motion and different portions of the frequency bands studied, technical conditions and regulatory provisions for their operation, taking into account the results of the studies above, [and]
>
> *resolves to invite the 2019 World Radiocommunication Conference*
>
> to consider the results of the above studies and take necessary actions, as appropriate, provided that the results of the studies referred to in *resolves to invite ITU-R* are complete and agreed by study groups. (Resolution 158 (WRC-15))

These incremental regulatory improvements help to lift the cloud of regulatory uncertainty over the rollout of innovative new satellite services and promote their ability to obtain required domestic approvals to operate ESIMs services. These solutions, while neither elegant nor comprehensive, are politically expedient and are what is achievable in the highly charged, consensus-based regulatory crucible that is the WRC. But they leave in their wake complex layers of outdated regulatory texts (such as No. 5.526) for future cleanup (Fig. 6).

Unmanned Aerial Systems. WRC-15 continued the consideration from the previous conference of the issue of whether FSS could be used to support control and non-payload communication (CNPC) links of unmanned aircraft systems (UAS) in nonsegregated airspace (WRC-15 Agenda Item 1.5). This issue was triply controversial: it dealt with the thorny regulatory issue of using FSS to provide services to Earth stations in motion; it concerned "safety-of-life" communications, i.e., the command and control of unmanned aircraft via a non-safety-rated spectrum allocation; and it concerned a politically divisive issue in a UN body due to the dual-use nature of drones (i.e., civil and military). Of course, UAS are used for a growing number of civil applications, including resource monitoring and management, weather forecasting, geological surveying, and search and rescue, among many others. WRC-12 had been unable to resolve the issue and the 3 years of preparations for WRC-15 continued to be inordinately difficult and complex.

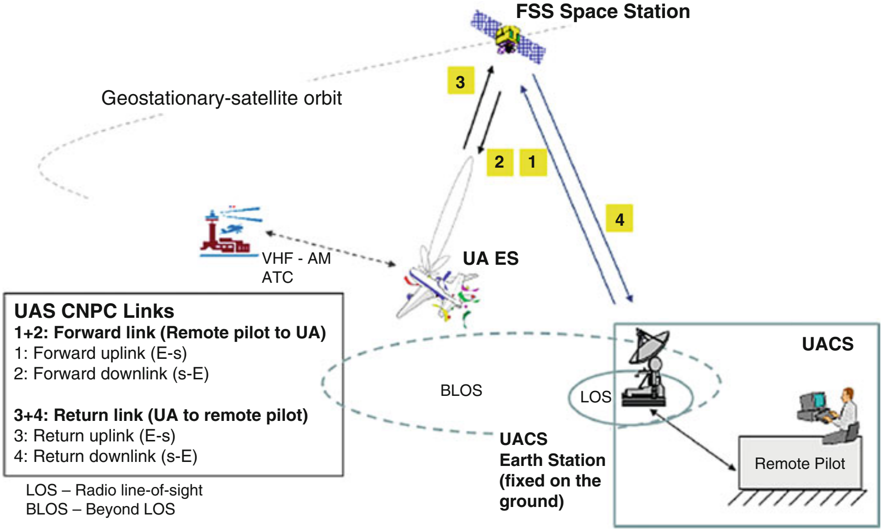
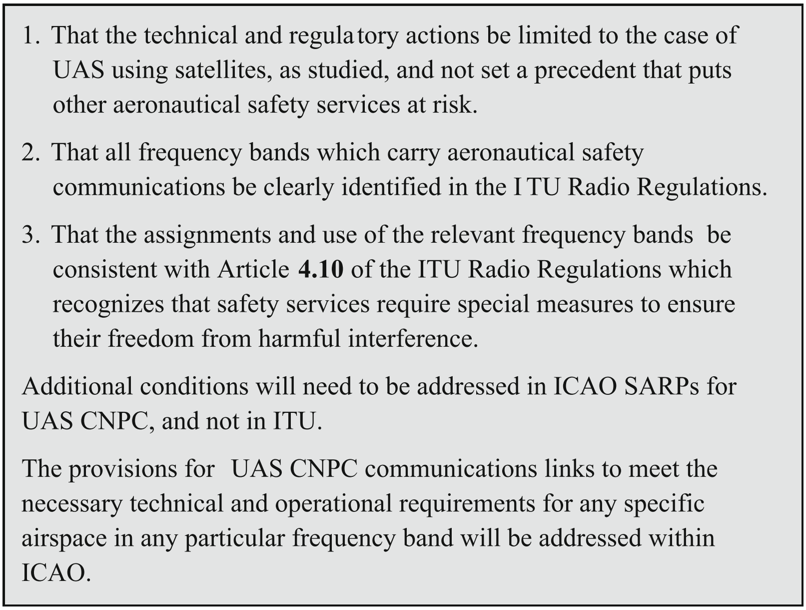


Fig. 6 Elements of UAS architecture using the FSS (Resolution 155 (WRC-15))

UAS operating beyond line of sight of terrestrial stations must necessarily rely upon satellite communications to control their operations. However, control of an aircraft is clearly a safety-of-life service that requires a higher level of guaranteed availability than is typically ensured by a regular non-safety radio service allocation (such as, in the case of UAS, payload communications). Traditionally, under existing regulations, only satellites operating in the Aeronautical Mobile Satellite (Route) Service (AMS(R)S) would be qualified to provide safety-of-life services to aircraft. FSS operators were not interested in operating in the AMS(R)S allocation because it would require new coordination agreements with higher levels of protection as well as acceptance of priority and preemption requirements which are not considered to be commercially reasonable. However, there is insufficient AMS(R)S capacity available or planned to meet the growing requirements for UAS CNPC links, while there is an abundance of FSS capacity available globally now. Again, technology and market needs were rapidly outstripping the bounds of the existing regulations and a more pragmatic, expedient approach was desired.

During the WRC-15 study cycle, the ITU-R Study Groups assessed the growing requirements for UAS CNPC spectrum: the technical characteristics and performance of CNPC links in varying operating conditions and the compatibility between this use and other services co-allocated in FSS bands. The views of the ITU's sister UN body, the International Civil Aviation Organization (ICAO), were crucial in finally resolving this agenda item after two conferences and two study periods. Although the ITU is the primary UN body for determining spectrum use, only ICAO has the authority to establish the safety requirements and standards for civil aircraft operation.

At WRC-12, ICAO had firmly stuck to the traditional view that CNPC links could only be provided by satellites in a safety-of-life allocation. However, in the final



1. That the technical and regulatory actions be limited to the case of UAS using satellites, as studied, and not set a precedent that puts other aeronautical safety services at risk.
2. That all frequency bands which carry aeronautical safety communications be clearly identified in the ITU Radio Regulations.
3. That the assignments and use of the relevant frequency bands be consistent with Article **4.10** of the ITU Radio Regulations which recognizes that safety services require special measures to ensure their freedom from harmful interference.

Additional conditions will need to be addressed in ICAO SARPs for UAS CNPC, and not in ITU.

The provisions for UAS CNPC communications links to meet the necessary technical and operational requirements for any specific airspace in any particular frequency band will be addressed within ICAO.

Fig. 7 ICAO Positions for the conference (Doc. CMR15/17 (2015))

lead-up to WRC-15, ICAO refined its views in light of the growing UAS requirements and, perhaps, the will of its own members. This proved to be the decisive element to finding a consensus solution. ICAO recognized that UAS “have great potential for innovative civil applications, provided that their operation does not introduce risks to safety of life.” ICAO announced that it was willing to support use of FSS systems for UAS CNPC links in nonsegregated airspace, but only if they could satisfy these conditions (Fig. 7):

In light of ICAO’s WRC-15 position, the conference was able to reach consensus to allow UAS CNPC operations in the following Ku and Ka-band FSS allocations that are not widely used to supporting terrestrial services:

10.95–11.2 GHz (space-to-Earth)
11.45–11.7 GHz (space-to-Earth)
11.7–12.2 GHz (space-to-Earth) in Region 2
12.2–12.5 GHz (space-to-Earth) in Region 3
12.5–12.75 GHz (space-to-Earth) in Regions 1 and 3
19.7–20.2 GHz (space-to-Earth)
14.0–14.47 GHz (Earth-to-space)
29.5–30 GHz (Earth-to-space)

Resolution 155 (WRC-15), “Regulatory provisions related to earth stations on board unmanned aircraft which operate with geostationary-satellite networks in the fixed-satellite service in certain frequency bands not subject to a Plan of Appendices

30, 30A and 30B for the control and non-payload communications of unmanned aircraft systems in non-segregated airspaces," provides a long list of prerequisites before CNPC links from FSS networks can be introduced, including:

- ICAO's development of international standards and recommended practices and procedures (SARPs) for implementation of CNPC links in these FSS allocations
- Definition by the ITU Radiocommunication Bureau of a new class of station for Earth stations providing UAS CNPC links
- That Earth stations of UAS CNPC links shall operate within the notified and recorded parameters of the associated satellite network, including specific or typical Earth stations of the geostationary FSS satellite network(s) published by the Bureau
- That Earth stations on board UA be designed and operated so as to be able to accept the interference caused by terrestrial services and from other satellite networks operating in conformity with the Radio Regulations
- That operators be able to ensure real-time interference monitoring, estimation and prediction of interference risks, and planning solutions for potential interference scenarios
- That power-flux density hard limits for UAS CNPC links be developed so that they do not cause harmful interference to terrestrial systems of other administrations and that the example of limits provided in the resolution be reviewed by the next WRC in 2019
- That ITU Radiocommunication Sector studies on technical, operational, and regulatory aspects of the implementation of this resolution be completed along with adoption of an ITU-R Recommendation defining the technical characteristics of CNPC links and conditions of sharing with other services

Resolution 155 (WRC-15) invites the 2023 World Radio Conference to consider the results of the studies referred to in the resolution and to take necessary actions. It further invites ICAO to provide information in time for WRC-19 and WRC-23 on its efforts on the implementation of UAS CNPC links, including development of SARPs. Thus, the ability to use FSS to support UAS CNPC links will be reviewed at the next two WRCs. Again, WRC-15 found a way chart a course forward to meet growing requirements for telecommunications services and marketplace realities, despite entrenched opposition and long-standing regulatory traditions and provisions.

Additional Spectrum Allocations for Satellite Services

Although much of the space industry's success at WRC-15 was based on defending existing spectrum access, the industry made notable headway on the offensive side as well – obtaining access to additional spectrum resources and improving existing satellite allocations. In this vein, this conference added spectrum allocations in the Ku-band and addressed future improvements to satellite allocations in the Ku and V-bands. In addition, WRC-15 created a new satellite allocation for provision of global flight tracking services.

Ku-Band. The satellite industry has long-sought additional satellite spectrum in the Ku-Band to correct the imbalance between uplinks and downlinks and among the world's three radio regions. The Ku-band FSS allocations support a broad range of satellite services, including service to very small aperture terminals (VSAT), enterprise and direct-to-consumer broadband services, satellite news gathering, and backhaul, to name a few. The desired frequency bands are currently used to support many incumbent services, including military operations, fixed services, broadcasting, and space sciences.

In the preparations for WRC-15, and at the conference itself, the solution proved to be elusive, despite genuine efforts by the parties to find an acceptable technical compromise. After long and detailed consideration, the industry finally succeeded in obtaining some relief. In Region 1, a downlink allocation was added at 13.4–13.65 GHz and an uplink allocation at 14.5–14.8 GHz was added in 39 countries via footnotes to the Table of Frequency Allocations incorporating resolutions containing restrictions on antenna size and power limits in order to protect operations of the incumbent services from harmful interference. The results are contained in Resolutions 163 and 164 (WRC-15).

V-Band. The Americas and European administrations proposed WRC-19 agenda items to study and consider additional FSS allocations in various frequency ranges above 37.5 GHz, including regulatory changes to accommodate proposed non-GSO systems in this spectrum range.

In approving three new FSS spectrum agenda item for future WRCs, the conference delegates agreed that:

- That satellite systems are increasingly being used to deliver broadband services and can help enable universal broadband access
- That next-generation fixed-satellite service technologies for broadband will increase speeds (45 Mbps is already available), with faster rates expected in the near future
- That technological developments such as advance in spot-beam technologies and frequency reuse are used by the fixed-satellite service (FSS) in spectrum above 30 GHz to increase the efficient use of spectrum

(Resolution 161, "Studies relating to spectrum needs and possible allocation of the frequency band 37.5–39.5 GHz to the fixed-satellite service," and Resolution 162, "Studies relating to spectrum needs and possible allocation of the frequency band 51.4–52.4 GHz to the fixed-satellite service (Earth-to-space)" (WRC-15))

The conference agreed that the ITU-R will conduct and complete studies in time for WRC-19 to consider additional spectrum needs for the development of FSS in light of other FSS spectrum allocations and whether those allocations are optimized for the most efficient spectrum use. Subject to this analysis, and the results of sharing and compatibility analysis with existing services, WRC-19 will determine the suitability of new primary allocation of the 51.4–52.4 GHz (Earth-to-space) to FSS, limited to FSS feeder links for geostationary orbit use. In addition to fixed and mobile services, the FSS will need to protect radioastronomy observations in the band and passive services in a neighboring frequency band.

The Americas Region also proposed that WRC-19 consider adding a "reverse-band" allocation to FSS in the band 37.5–39.5 GHz (Earth-to-space), limited to operation of FSS feeder links for GSO and non-GSO use. This allocation would complement the existing FSS allocation at 37.5–42.5 GHz (space-to-Earth). The 36–37 GHz band is also allocated to the Earth exploration-satellite service (passive) and the space research service. WRC-15 adopted Resolution 161 which conditions the proposed new allocation on demonstration through sharing and compatibility studies that the primary and secondary incumbent services would be protected from harmful interference. However, in the final moments of the negotiation, the new agenda item was changed from WRC-19 to WRC-23, due to the large number of items in the WRC-19 agenda (Fig. 8).

Global Flight Tracking (1090 MHz). Perhaps the most heralded action by WRC-15 concerned a matter that was not even on its original agenda. In October 2014, the Plenipotentiary Conference (PP-14), the governing authority of the ITU, took an unusual and extraordinary action of adding an agenda item to the WRC, which was then little more than one year away. In reaction to the loss of Malaysian Airlines Flight 370 and the worldwide discussions on global flight tracking, the governing body of the ITU adopted Resolution 185, "Global flight tracking for aviation" (Busan 2014), which served to place an item on the agenda for WRC-15, as a matter of urgency, to consider global flight tracking, taking into account ITU-R studies. PP-14 noted that ICAO had encouraged the ITU to take action to provide "necessary spectrum allocations for satellite when emerging aviation needs are identified," but also noted that "flight tracking for civil aviation is currently available across the globe, apart from some parts of polar regions."

The global flight tracking agenda item was not without controversy for multiple reasons. For one, there was insufficient time before WRC-15 for the rigorous

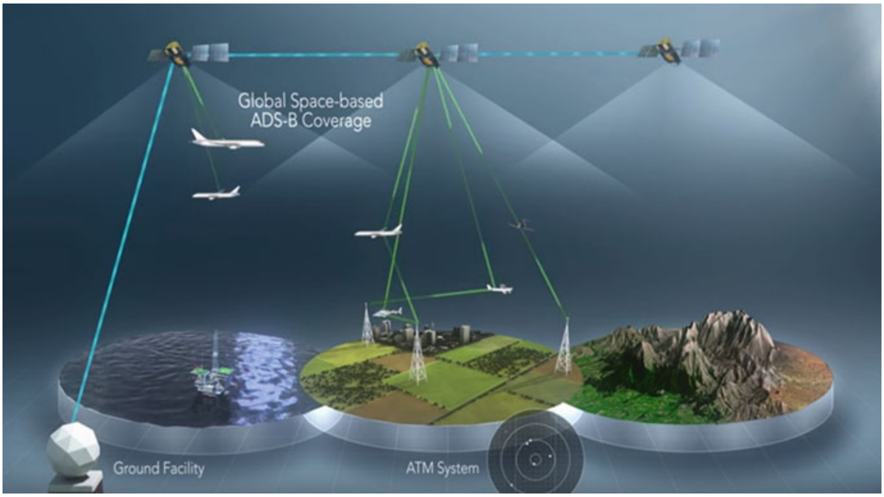


Fig. 8 "Radio allocated for global flight tracking," ITU Press Release (Geneva, 2015)

technical studies which are normally required before action is undertaken under any WRC agenda item. Incumbent providers of GFT services didn't view the item to be an urgent matter in light of the current availability of their services. Some thought that the ITU was "jumping the gun" and acting too far in advance of ICAO. Finally, a few Member States and Sector Members were concerned that the proposed new spectrum allocation could limit use by military aircraft systems (which use is not recognized by the Radio Regulations or ICAO). The issue continued to prove contentious in the remaining preparatory period before the WRC.

Thus, many were surprised when WRC-15 made quick work of this agenda item, resolving the matter early in its second week. WRC-15 added via footnote a global allocation of the aeronautical mobile-satellite (R) service (AMS(R)S) (Earth-to-space) to the frequency band 1087.7–1092.3 MHz, limited to the space station reception of Automatic Dependent Surveillance-Broadcast (ADS-B) emissions from aircraft transmissions that operate in accordance with recognized international aeronautical standards. The allocation is subject to the application of WRC-15 Resolution 425, "Use of the frequency band 1 087.7–1 092.3 MHz by the aeronautical mobile-satellite (R) service (Earth-to-space) to facilitate global flight tracking for civil aviation," which requires accommodation of other uses of this frequency bands, including non-ICAO systems. WRC-15's actions serve to extend the existing terrestrial aeronautical system for flight tracking to oceanic, remote, and polar regions – particularly if deployed on an non-GSO constellation, such as iridium.

In addition, in response to a request by ICAO and the proposals of several regional groups, WRC-15 adopted a related item for the WRC-19 agenda to consider and address spectrum and regulatory requirements for the future Global Aeronautical Distress and Safety System (GADSS) being developed by ICAO and the air transport industry. GADSS is being designed to "address the timely identification and location of an aircraft during all phases of flight as well as distress and emergency situations." It will also support search and rescue and flight data retrieval. It is envisioned to be a system of systems employing terrestrial and space components.

WRC-15 approved Resolution 426, (WRC-15), "Studies on spectrum needs and regulatory provisions for the introduction and use of the Global Aeronautical Distress and Safety System," which invites the ITU-R to conduct relevant studies on the GADSS radio requirements, analyze existing spectrum allocations and determine whether any additional spectrum is required, conduct studies on sharing and/or compatibility with existing services, and study existing regulatory methods and any new methods that may be required. WRC-19 can then take appropriate action based on the results of the studies or refer the matter to a later conference.

WRC-15 Consideration of Nanosatellites and Picosatellites

WRC-15 was tasked to consider the procedures for notifying space networks and to consider possible modifications to enable the deployment and operation of nanosatellites and picosatellites, taking into account their short development time, short mission time, and unique orbital requirements. These small satellites support a wide range of missions and applications and are often employed by new space entrants, such universities and research institutions, which may not be fully aware of the Radio Regulations requirements concerning coordination and operation of satellite systems and their responsibility for avoidance of harmful interference to other radio systems. In any case, the short mission life of these small satellites and their lack of specific orbital parameter information that would typically be required for an ITU filing would make it difficult to strictly apply the requirements of Articles 9 and 11 to these satellites.

In the preparations for WRC-15, the ITU-R had concluded that there was a need to educate and inform small satellite operators about their responsibilities under the Radio Regulations. At the 2015 Radiocommunication Assembly (RA-15) that immediately preceded the WRC, the topic was further considered. The RA approved ITU-R Resolution 68, "Improving the dissemination of knowledge concerning the applicable regulatory procedures for small satellites, including nanosatellites and picosatellites." It resolved "to develop material, such as Recommendations, Reports or a Handbook on small satellites (in particular, satellites whose mass is less than 100 kg), containing detailed information that would help to improve knowledge of the applicable procedures for submitting filings of satellite networks to ITU." It invited administrations to inform their national entities involved in the development, manufacturing, operation, and launch of small satellites about the ITU regulations and to encourage their application.

WRC-15 considered the RA's action and decided that no changes to the Radio Regulations were needed at this time, cautioning that any such changes could have broader applicability to all satellite systems. WRC-15 also considered proposals from three regional groups for future WRC agenda items concerning small satellites. WRC-15 decided that a dedicated small satellite agenda item was not needed to support consideration of modifications to the regulatory procedures by a future WRC because it could be considered under Agenda Item 7 which is automatically included on every conference's agenda to update satellite regulatory procedures. After much debate, WRC-15 did approve one new WRC-19 agenda item concerning small satellites that had been proposed by Europe. However, because the Radio Regulations do not address the size or mass of satellites, the action was instead directed to satellites with short duration missions, which is another way of categorizing nanosats and picosats. The conference adopted Resolution 659 (WRC-15), "Studies to accommodate requirements in the space operation service for non-geostationary satellites with short duration missions."

The resolution takes note of the growing number of satellites which have short duration missions and recognizes that they provide an affordable means to access orbital resources for new entrants in space. These satellites require reliable control and tracking for management of space debris and there is thus a demand for suitable allocations to the space operations service for their telemetry, tracking, and command (TT&C). The resolution invites the ITU-R to study spectrum requirements for TT&C for the growing number of non-GSO satellites with short duration missions and to assess the suitability of existing allocations to the space operations service below 1 GHz and whether additional allocations are needed. If so, the ITU-R is to conduct sharing and compatibility studies and studies on mitigation techniques to protect incumbent services or to consider possible new allocations or upgrades of existing allocations within the ranges 150.05–174 MHz and 400.15–420 MHz. WRC-19 is invited to consider the results of these studies and to take appropriate action.

Emerging Non-GSO Systems

The recent emergence of newly proposed large non-GSO satellite systems, popularly referred to as “Mega LEOs,” was evident at WRC-15. In addition to existing non-GSO commercial operators (O3B Networks, Iridium Satellite Communications, and Globalstar), new players (such as OneWeb and Space Exploration Technologies Corporation (SpaceX)) were active participants in the conference. WRC-15 received proposals to accommodate non-GSO interests, including making permanent the provision for Globalstar’s feederlinks and adding an allocation to support Iridium’s plans for its next constellation to provide Global Flight Tracking, and to include non-GSO systems in the studies for an additional FSS allocation at 51.4–52.4 GHz. Additional proposals addressed spectrum resources and regulatory modifications to accommodate the newly proposed non-GSO systems.

In his report to the conference, the Director of the Radiocommunication Bureau described the challenges faced by the Bureau in managing recent non-GSO filings:

> Since November 2014, the Bureau has received numerous requests for coordination for non-GSO systems operating in the FSS subject to equivalent power-flux density (epfd) limits in Article 22 and also to coordination under No. 9.7B of the Radio Regulations. A non-exhaustive list of such requests is provided below:
>
> (i) Satellite systems consisting of hundreds of satellites (about 800 satellites) on low Earth circular orbits with a single inclination value and with an indication that all frequency assignments of the system would be operated simultaneously
> (ii) Satellite systems consisting of tens of satellite (about 40 satellites) in different orbit planes, including one Tundra, one Molniya, and one TAP (three Apogee) orbits, with an indication that satellites at the proposed orbits would not be operated simultaneously and that only one of these orbit configuration would be implemented and notified for recording in the MIFR
> (iii) Satellite systems consisting of tens of thousands of satellites (from 70,000 to more than 230,000 satellites) in more than 1000 orbit planes, low Earth orbit for some systems, and medium Earth orbits for others, including different inclination values with the indication that the satellites in this system would be operated in different technically compatible subsystems corresponding to a unique altitude
> (iv) Satellite systems consisting of thousands of satellites (about 4000 satellites) on low Earth circular orbits with different inclination values and with an indication that all frequency assignments of the system would be operated simultaneously (Doc. CMR 15/4 (Add.1) (2015)).

The report also included updates on the processing and analysis tools and ideas for managing the coordination of these systems.

V-Band. Based on a request from O3B Networks, the Americas region submitted a proposal to WRC-15 to study and address the regulatory status of future non-GSO systems in the V-band. This proposal recognized that the V-band represents the next expansion band for satellite services following exploitation of the Ka-band but that prospective operators face great uncertainty in that "there are currently no mechanisms in the RR establishing coordination procedures applicable to NGSO systems operating in the frequency bands currently allocated to the FSS in the range from 37.5 to 51.4 GHz" (Inter-American Telecommunication Commission 2015). Except where otherwise provided, No. 22.2 of the Radio Regulations provides that non-GSO systems may not cause unacceptable interference to GSO FSS and broadcasting-satellite service (BSS) networks and shall not claim protection from GSO FSS and BSS satellite networks. This provision effectively makes non-GSO permanently secondary relative to GSO operators, even those who come later, a regulatory uncertainty that could serve to hinder investment by non-GSO systems in bands where No. 22.2 applies, currently including the entirety of the V-band FSS allocation.

WRC-15 agreed to establish a WRC-19 agenda item "to consider the development of a regulatory framework for non-GSO FSS satellite systems that may operate in the frequency bands 37.5–39.5 GHz (space-to-Earth), 39.5–42.5 GHz (space-to-Earth), 47.2–50.2 GHz (Earth-to-space), and 50.4–51.4 GHz (Earth-to-space)" in accordance with new Resolution 159 (WRC-15). In adopting the agenda item, the conference noted the promise of GSO and non-GSO satellite constellations to provide high-capacity, low-cost communications to the most isolated regions of the world. The conference invited the ITU-R to conduct studies on technical, operational, and regulatory issues concerning non-GSO FSS systems in these frequency bands while ensuring protection of GSO satellite networks. These studies are to focus on "the development of equivalent power flux-density limits produced at any point in the GSO by emissions from all the earth stations of a non-GSO system in the fixed-satellite service or into any geostationary FSS earth station, as appropriate". In addition, studies are to address protection of all incumbent services in the band, including fixed, mobile, passive services, and radioastronomy, as provided in Resolution 159 (WRC-15), "Studies of technical, operational issues and regulatory provisions for non-geostationary fixed-satellite services satellite systems in the frequency bands 37.5–39.5 GHz (space-to-Earth), 39.5–42.5 GHz (space-to-Earth), 47.2–50.2 GHz (Earth-to-space), and 50.4–51.4 GHz (Earth-to-space)."

C-Band. WRC-15 also considered a proposal from the United States, requested by The Boeing Company, on updating the technical provisions regarding non-GSO use of C-band frequencies. The United States requested that the WRC adopt a resolution calling for studies and review by WRC-19, in light of the technical advancements since the sharing criteria for GSO and non-GSO systems had been developed 12 years previously and the capability of non-GSO systems to provide low-cost, global broadband communications to meet growing requirements. The conference agreed and adopted Resolution 157 (WRC-15), "Study of technical and operational issues and regulatory provisions for new non-geostationary satellite orbit systems in the 3700–4200 MHz, 4500–4800 MHz, 5925–6425 MHz and

6725–7025 MHz frequency bands allocated to the fixed-satellite service". The resolution takes note of the fact that the Article 21 power-flux density limits and Article 22 equivalent power-flux density limits in portions of these bands were developed at the WRC in 2003 based on requirements for a highly elliptical orbit (HEO) configuration, which was the only type of system under consideration at that time. Moreover, there are no Article 22 limits for some of these bands. In his report to the conference, the Director had observed that the power limits may need to be reviewed or confirmed "taking into account the characteristics of systems recently submitted and the overall trend for a growing interest in operating non-GSO FSS systems, with a view to ensure that all existing services are adequately protected" (Doc. CMR15/4(Add.2) (Rev.1) (2015)).

WRC-15 resolved that the ITU-R would undertake studies throughout these bands with emphasis on the possible revision of Articles 21 and 22 while ensuring protection of existing services and without change to No. 22.2 or the existing protection criteria for GSO FSS networks. The Director was further instructed to report on the results of these studies in his report to WRC-19.

Coordination Considerations. The United States and United Kingdom also contributed proposals to WRC-15 regarding the issue of future coordination of large non-GSO FSS systems, an issue which was also addressed in the Director's report to the conference. The US proposal addressed coordination among multiple non-GSO systems, particularly the challenging case involving several hundreds or thousands of satellites in a constellation. At the request of SpaceX, the United States proposed a new resolution instructing the ITU-R to carry out studies on the effectiveness of the procedures for coordination between non-GSO FSS satellite systems in the bands 10.7–13.25 GHz, 13.75–14.5 GHz, 17.3–17.7 GHz, 17.7–20.2 GHz, and 27.5–30 GHz to identify possible mechanisms to facilitate coordination and co-frequency sharing among them and for WRC-19 to consider and take action on the results of the studies under Agenda Item 7, which recurs at every conference to consider modifications to the coordination process. Moreover, it sought to instruct the Director to convene, upon request of a notifying administration, voluntary multilateral meetings with the goal of facilitating the completion of coordination among non-GSO FSS systems (Doc. CMR15/6 (Add.23) (Add. 2) (Add.3) (2015)).

Although this proposal did not succeed at WRC-15, the issue could be raised again at WRC-19 under Agenda Item 7, and studies can be carried out within the ITU-R. In addition, the following point was recorded in the minutes of the Plenary: "WRC-15 recognizes that notifying administrations may mutually agree on the organization of multilateral coordination meetings for non-GSO FSS systems and may wish to seek the assistance of the Bureau under existing procedures" (Doc. CMR15/505 (2015)).

The United Kingdom proposed that WRC-15 amend the Radio Regulations to add new provisions regarding non-GSO systems. It observed that there is a provision in the Radio Regulations defining the bringing into use of frequency assignments to GSO stations, for example, No. 11.44B. This is important because such assignments are required to be brought into use within 7 years of the receipt by the Bureau of their advance publication information, else face cancellation pursuant to No. 11.44. However, the Radio Regulations lack definition of the bringing into use of frequency assignments to stations of non-GSO satellite systems. One is left to wonder whether bringing into use a single or a few satellites of an intended thousand or more satellite constellation would truly suffice for that purpose.

The United Kingdom asserted that:

> the absence of appropriate provisions for non-GSO satellite systems may leave open the possibility for spurious claims that assignments to non-GSO networks or systems have been brought into use. We are also of the view that sharing spectrum resources between GSO networks and non-GSO systems and between different non-GSO systems is already a complex task. Therefore, if frequency assignments to stations of non-GSO systems are brought into use on spurious grounds with the aim of warehousing spectrum, this would inevitably lead to an inefficient use of that limited resource. (Doc. CMR15/132(Add.23) (2015))

Thus, it proposed to add a new provision to Article 11 of the Radio Regulations:

> A frequency assignment in the fixed-satellite or mobile-satellite services to a space station in the non-geostationary-satellite orbit shall be considered as having been brought into use when at least the minimum number indicated in the coordination request information of non-geostationary-satellites with the capability of transmitting or receiving that frequency assignment has been deployed in at least one of the notified orbital planes. (Doc. CMR15/132(Add.23) (2015))

This proposal, too, did not succeed at WRC-15, due to lack of sufficient time to fully consider its potential implications. The delegates agreed that the issue raised a real concern that deserved further consideration at a future conference after further study. The Plenary concluded:

> WRC-15 invites ITU-R to examine, under the standing WRC agenda item 7, the possible development of regulatory provisions requiring additional milestones beyond those under RR Nos. 11.25 and 11.44 on the systems referred to in the paragraph above. This study may also consider the implications of the application of such milestones to non-GSO FSS/MSS systems brought into use after WRC-15. (Doc. CMR15/504 (2015))

Further WRC-15 Regulatory Considerations

Under Agenda Item 7, every WRC considers:

> Possible changes, and other options, in response to Resolution 86 (Rev. Marrakesh, 2002) of the Plenipotentiary Conference, an advance publication, coordination, notification and recording procedures for frequency assignments pertaining to satellite networks, in accordance with Resolution 86 (Rev.WRC-07) to facilitate rational, efficient, and economical use of radio frequencies and any associated orbits, including the geostationary-satellite orbit.

At WRC-15, 19 issues were considered under this item, which actually represents a decline over the previous conference. The issues addressed various detailed aspects of the finer points of satellite regulatory procedures, often further clarifying matters addressed at the previous conference. Some highlights from the Agenda Item 7 discussions follow.

Advance Publication. One notable decision of WRC-15 under Agenda Item 7 was the decision to effectively eliminate the advance publication information (API) stage in the satellite network coordination process. This had long been discussed in the ITU, as the required information for the API filing had been streamlined to such a degree over the years that it no longer conveyed much useful information, yet added unnecessary time (at least six months) to the coordination process. But instead of undertaking the extensive review and editing of the Radio Regulations that would be required to fully remove the API stage from its complex provisions, the conference quite pragmatically opted instead to have the Bureau automatically generate an API upon receipt of a coordination request and then to publish both. This decision minimized changes to the current Radio Regulations while allowing an incremental improvement. Thus, the API would still be used to start the 7-year regulatory period, but the six-month delay to start coordination (and secure one's place in the queue) would be removed.

Satellite Hopping. This issue concerns using one space station to bring frequency assignments into use at different orbital locations within a short period of time – such as drifting a satellite across multiple slots in order to meet ITU timing deadlines. Although satellite hopping is often described as a method of "gaming" the Radio Regulations, there could also be legitimate reasons, such as fleet management, for this practice. Thus, WRC-15 decided not to prohibit this activity, but to require the submission of information to the ITU when it is utilized. The conference adopted Resolution 40, "Use of one space station to bring frequency assignments to geostationary-satellite networks at difference orbital locations into use within a short period of time" clarifying the additional information that an administration should provide the Bureau when bringing into use a frequency assignment using a space station that has previously been brought into use.

Satellite Launch Failure. The conference considered whether the Radio Regulations should be amended to provide for an extension of the 7-year regulatory period for bringing into use in the event of a launch failure. The conference instead decided to retain the current ad hoc approach under which such extensions can be granted through request to the Radio Regulations Bureau: "The Board may address requests for a time-limit extension based on either a co-passenger issue or *force majeure* taking into account internationally applicable rules and practices in this regard so long as any extension is 'limited and qualified' " (Doc. CMR15/504 (2015)).

Excessive Satellite Filings. The conference also considered whether measures should be taken to mitigate "excessive satellite filings," a perennial issue for the ITU-R and its members, one which in previous decades had led to adoption of Satellite Network Cost Recovery, Administrative Due Diligence, and several revisions to the Radio Regulations, including the reduction of the time period for bringing into use of a satellite network from nine years to seven Allison (2014). The champion of this issue was Egypt, on behalf of the Arab countries.

> Issues were raised regarding coordination difficulties that arise for newcomer networks, as a result of multiple advance publication and multiple coordination requests submitted to the BR which may be in excess of what is actually required and practically implementable, in which many of these networks are usually suppressed after the expiry of the regulatory deadline time-limit of seven years as a result of not being brought into use or not being notified to the BR. However, during such regulatory time-limit, these networks need to be taken into account by subsequently filed networks and thus complicate the coordination process or even prevent subsequently filed networks to have timely access to the orbital/spectrum resources. This may result in misuse or irrational usage of frequency assignments and associated orbital resources. Taking into account the number of coordination requests that are suppressed after the seven-year regulatory lifetime, one may infer that such filings, in some cases, could be considered as excessive and could create barriers and difficulties for coordinating later filed satellite networks. However, uncertainties associated with procedures of effecting coordination properly, may be resolved by submitting multiple filings to provide flexibilities for notifying member states. (Doc. CMR15/3 (2015))

The conference did not agree to take any specific actions in relation to this issue. Instead, the matter was recorded in the minutes of the plenary to highlight the ongoing concerns with excessive filings, especially with coordination filings which may have a negative impact on later submitted networks. The proponent administrations noted that excessive filing may need to be addressed at a future WRC under Agenda Item 7 Doc. CMR15/505 (2015).

National Defense Article 48 of the ITU Constitution provides Member States with an exemption for purposes of their installations for national defense services:

> 1. Member States retain their entire freedom with regard to military radio installations.
> 2. Nevertheless, these installations must, so far as possible, observe statutory provisions relative to giving assistance in case of distress and to the measures to be taken to prevent harmful interference, and the provisions of the Administrative Regulations concerning the types of emission and the frequencies to be used, according to the nature of the service performed by such installations.
> 3. Moreover, when these installations take part in the service of public correspondence or other services governed by the Administrative Regulations, they must, in general, comply with the regulatory provisions for the conduct of such services. (ITU Constitution, (2015))

In its report to WRC-15, the Radio Regulations Board noted that a growing number of administrations had responded to the Bureau's inquiries under No. 13.6 of the Radio Regulations on whether a frequency assignment had been implemented stating that the subject assignment was used for defense, military, or other governmental purposes. The Board sought clarification from the conference as to whether Article 48 should be interpreted broadly to apply to such broad responses and to all categories of service, including public correspondence Doc. CMR15/14 (2015). The conference decided that the exemption would only be applied where the administration specifically invoked Article 48. However, the class of station or service would not be restricted Doc. CMR15/505 (2015).

Equitable Access. The subject of equitable access to the orbits appears on every WRC agenda and presents the opportunity to reopen and revisit the very foundation of the regulatory framework that has successfully supported today's satellite ecosystem. However, at WRC-15, the issues relating to equitable access were resolved in a routine manner indicating a continuation of the consensus over the current arrangements for use of the orbits.

The conference took action to confirm the Radio Regulations Board's extension of the regulatory deadline for bringing into use the frequency assignments of the satellite networks of two developing countries: Colombia (SATCOL 1B) and Laos (LAOSAT-128.5E). A Russian satellite network, CSDRN-M, also received a waiver of a regulatory deadline due to its importance in providing safety-of-life services for manned space flights and the international space station.

WRC-15 also suppressed Resolution 11 from the previous conference on "Use of satellite orbital positions and associated frequency spectrum to deliver international public telecommunication services in developing countries." This resolution, which had resulted from proposals by African nations to recent Plenipotentiary conferences and WRC-12, grew out of concern over maintaining the "Common Heritage" orbital positions that had initially been registered on behalf of Intelsat when it was an intergovernmental organization and in adopting regulatory measures to improve the access of developing countries to international public telecommunication services delivered via satellite. Although there were proposals to WRC-15 to modify the resolution to repeat the call for studies to support new regulatory measures, the conference instead referred to the action of the Radiocommunication Assembly on this issue during the week immediately prior to WRC-15.

RA-15 adopted ITU-R Resolution 69 "Development and deployment of international public telecommunications via satellite in developing countries" resolving to take a number of measures in the Radiocommunication Sector to work with the Development Sector improve access by developing countries, including:

- Collaborating with ITU-D on satellite technologies and applications as defined in ITU-R Recommendations and Reports and on satellite regulatory procedures in the Radio Regulations that will help developing countries with development and implementation of satellite networks and services
- Supporting ITU-D's development and deployment of international public telecommunication services via satellite in developing countries
- Undertaking studies to determine whether it might be necessary to apply additional regulatory measures to facilitate the development, deployment, and availability of international public telecommunications via satellite in developing countries
- Reporting the results of these studies to the 2019 World Radiocommunication Conference (WRC-19) (ITU-R Resolution 69 (2015)).

The Director of the Development Sector was also invited to organize workshops, seminars, and training on "sustainable and affordable access to satellite telecommunications, including broadband, and to continue activities between the relevant study groups of ITU-D and ITU-R that will assist developing countries in building capacities in the development and use of satellite telecommunications" and to take up the issue at the next World Telecommunications Development Conference in 2017.

Conclusion

Since the initial space radio conferences of the 1980s, the ITU has concluded seven treaties governing spectrum allocation and orbit use for space-based systems.[3] The Final Acts of each of these World Radiocommunication Conferences contain amendments to the international Radio Regulations, a basic instrument of the ITU and itself, a treaty. These outcomes, including globally and regionally harmonized satellite spectrum allocations, technical and operational parameters, and regulatory and procedural methods for coordination and registration, have made possible the implementation of today's vast number of commercial and government satellite networks and systems. The stable global legal and regulatory framework that has resulted from this process, and its clear pathway to international recognition and protection from harmful interference, has supported and sustained the investment and constant innovation necessary to foster introduction of new services and technologies, making ever more efficient and effective use of the available spectrum and orbital resources in space-based systems.

With the rise of the terrestrial mobile broadband industry and the mounting tsunami of broadband applications, where now even our household appliances are becoming wireless spectrum consumers (the Internet of things), the precious regionally and globally harmonized satellite spectrum allocations have understandably become a desirable target of other industries. The last three World Radio Conferences since 2007 have featured deeply bitter battles over access to these invaluable spectrum bands and WRC-19 is slated to continue on this theme. At the same time, the insatiable drive for unfettered broadband has also fueled new satellite solutions to meet that demand as well. Thus, there are now high-throughput satellites providing broadband service from geostationary FSS networks and increasingly high-performance moving platforms. Most recently there has been renewed interest in large-scale non-GSO satellite system in lower Earth orbit, which promise to provide low-latency broadband services to every corner of the Earth, no matter how remote. Satellite services also support terrestrial wireless operators with backhaul solutions, among other support.

The challenge for the satellite industry for the foreseeable future is clear: to continue to make strides in rolling out invaluable new services and vigorously

---

[3] These include the World Administrative Radio Conference (Málaga-Torremolinos, 1992) and the World Radiocommunication Conferences of 1995, 1997, 2000, 2003, 2007, 2012, and 2015, all in Geneva but for WRC-2000, which took place in Istanbul. This tally does not include the ITU's quadrennial Plenipotentiary Conferences which also address satellite issues, albeit not in as much detail. The 2014 Plenipotentiary, for example, adopted a resolution on spectrum for global flight tracking.

utilizing its spectrum allocations – while retaining government and public support for these services. This activity will demonstrate the value of maintaining satellite allocations in these regions. However, it is clear that the terrestrial industry's spectrum needs must also be accommodated and that its need will remain a priority of ITU members. It is thus in the satellite industry's interest to find a way to share some of its spectrum resources with appropriate terrestrial services or to help find solutions in alternative frequency bands.

After 150 years, the International Telecommunication Union continues to be a place for nations and companies to come together to forge technical and political solutions so that telecommunications services, including services delivered via satellite, can be efficiently and equitably provided to the world's inhabitants free from harmful interference. It is a great demonstration of global cooperation, public-private partnership, and the power of space law. WRC-15 was the most recent example of such a success, and it set a course for 2019 and beyond.

Cross-References

- ► Regulatory Process for Communications Satellite Frequency Allocations
- ► Satellite Communications and Space Telecommunication Frequencies

References

A. Allison, *The ITU and managing satellite orbital and spectrum resources in the 21st century* (Springer, Cham, 2014)

F. Daudu. *Minutes of the Eighth Plenary Meeting*, Doc. CMR15/505 (Geneva, 2015b)

F. Daudu. *Minutes of the Seventh Plenary Meeting*, Doc. CMR15/504 (Geneva, 2015c)

Director, Radiocommunication Bureau, *Report of the Director on the activities of the Radiocommunication Sector; Part 1: Activities of the Radiocommunication Sector between WRC-12 and WRC-15,* Doc. CMR15/4 (Add.1) (Geneva, 2015a) https://www.itu.int/md/dologin_md.asp?lang=en&id=R15-WRC15-C-0004!A1!MSW-E. Accessed 31 Jan 2016

Director, Radiocommunication Bureau, *Report of the Director on the activities of the Radiocommunication Sector; Part 2: Experience in the application of the radio regulatory procedures and other related matters,* Doc. CMR15/4 (Add.2)(Rev.1) (Geneva, 2015b) https://www.itu.int/md/dologin_md.asp?lang=en&id=R15-WRC15-C-0004!A2-R1!MSW-E. Accessed 31 Jan 2016

Federal Communications Commission, Use of spectrum bands above 24 GHz for mobile radio services (2015). http://apps.fcc.gov/ecfs/comment/view?id=60001304804. Accessed 31 Jan 2016

GSMA, *Press Release: GSMA commends allocation of additional spectrum for mobile broadband at WRC-15* (Geneva, 2015b). http://www.gsma.com/newsroom/press-release/gsma-commends-allocation-of-additional-spectrum-for-mobile-broadband-at-wrc-15/?utm_campaign=CEO_member%20announcement%20wrc15_30%20Nov%202015&utm_medium=email&utm_source=Eloqua. Accessed 31 Jan 2016

Inter-American Telecommunication Commission, *Proposals for the work of the conference*, Doc. CMR15/7(Add.24)(Add.8) (Geneva, 2015). https://www.itu.int/md/dologin_md.asp?lang=en&id=R15-WRC15-C-0007!A24-A8!MSW-E. Accessed 31 Jan 2016

International Telecommunication Union, *Final Acts of the 2015 World Radiocommunication Conference* (Geneva, 2016 in press). http://www.itu.int/pub/R-ACT-WRC.12-2015/en. June 25, 2016

International Telecommunication Union, *Press Release: Radio spectrum allocated for global flight tracking; ITU World Radiocommunication Conference enables Earth-to-space reception of ADS-B transmissions* (Geneva, 2015b). http://www.itu.int/net/pressoffice/press_releases/2015/51.aspx#. Accessed 31 Jan 2016

International Telecommunication Union, *Collection of Basic Texts of the International Telecommunication Union Adopted by the Plenipotentiary Conference* (International Telecommunication Union, Geneva, 2015c). http://www.itu.int/pub/S-CONF-PLEN/e. Accessed 31 Jan 2016

International Telecommunication Union, *Book of ITU-R Resolutions* (Geneva, 2015d). http://www.itu.int/pub/R-VADM-RES-2015. Accessed 31 Jan 2016

International Telecommunication Union, *Radio Regulations* (Geneva, 2012). http://www.itu.int/dms_pub/itu-s/oth/02/02/S02020000244501PDFE.PDF. Accessed 31 Jan 2016

Radiocommunication *Bureau, Administrative Circular: Results of the first session of the Conference Preparatory Meeting for WRC-19 (CPM19-1)* CA/226 (Geneva, 2015) https://www.itu.int/md/dologin_md.asp?lang=en&id=R00-CA-CIR-0226!!MSW-E. Accessed 31 Jan 2016

Satellite Spectrum Initiative, *Press Release: World Radiocommunication Conference 2015 decides satellite spectrum is central to future vision for global connectivity (London, 2015)*. http://www.icontact-archive.com/YaDOh0pPV3BYaZ1ULmguJw_Mh04NiVam. Accessed 31 Jan 2016

Secretary-General, *ICAO position for the conference*, Doc. CMR15/17 (Geneva, 2015a). https://www.itu.int/md/dologin_md.asp?lang=en&id=R15-WRC15-C-0017!!MSW-E. Accessed 31 Jan 2016

Secretary-General, *Report of the Radio Regulations Board to WRC-15 Resolution 80 (Rev. WRC-07)*, Doc. CMR15/14 (Geneva, 2015b). https://www.itu.int/md/dologin_md.asp?lang=en&id=R15-WRC15-C-0014!!MSW-E. Accessed 31 Jan 2016

Secretary-General, *Report of the Conference Preparatory Meeting on operational and regulatory/procedural matters to WRC-15*, Doc. CMR15/3 (Geneva, 2015c). https://www.itu.int/md/dologin_md.asp?lang=en&id=R15-WRC15-C-0003!!MSW-E. Accessed 31 Jan 2016

Secretary-General, *Agenda of the Conference,* Doc. CMR15/1 (Geneva, 2015d). http://www.itu.int/md/R15-WRC15-C-0001/en. Accessed 31 Jan 2016

United Kingdom of Great Britain and Northern Ireland, Proposals for the work of the Conference, Doc. CMR15/132(Add.23)(Geneva, 2015a). https://www.itu.int/md/dologin_md.asp?lang=en&id=R15-WRC15-C-0132!A23!MSW-E. Accessed 31 Jan 2016

United States of America, Proposals for the work of the Conference, Doc. CMR15/6 (Add.23) (Add.2) (Add.3)(Geneva, 2015b). https://www.itu.int/md/dologin_md.asp?lang=en&id=R15-WRC15-C-0006!A23-A2-A3!MSW-E. Accessed 31 Jan 2016

New Millimeter, Terahertz, and Light-Wave Frequencies for Satellite Communications

Joseph N. Pelton

Contents

Abstract

The last 50 years of satellite communications has followed a consistent trend to produce networks that can provide higher and higher rates of throughputs at lesser cost. Closely linked to this trend has been a parallel effort to seek more efficient use of the allocated frequencies. The first satellite systems were power limited, but as satellite engineers designed more powerful spacecraft, the challenge has been more and more to find ways to use frequencies more efficiently. In short, in a digital world, the objective has become to send more bits of information per available Hz of radio frequency. This has been primarily accomplished by using greater complexity in the coding and multiplexing systems. This has also been achieved by polarization isolation and higher-gain antennas (and thus narrower

---

J.N. Pelton (✉)
International Space University Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_98

spot beams) that enable geographic isolation of the transmitted beams. If these narrow beams are spread sufficiently apart, this reduces interference and the radio frequencies can be reused over and over again. This process also minimizes the effective path loss of irradiated power by concentrating the beam to a tighter area.

This 50 years of satellite progress can be summarized by the ever-increasing power levels, frequency allocations, and system complexity to increases throughput efficiency. This "complexity" has allowed more throughput of information via the spectrum that is available. This is now typically measured in the metric of "digital bits" per hertz.

The fixed-satellite services (FSS), the mobile satellite services (MSS), and the broadcast satellite services (BSS) each in their own ways have applied this process to exploit the available frequency bands progressively over time. The lower-frequency bands have been used up first to meet initial demand in the earliest years. This is simply because these bands are easier to use. This is primarily because there is less rain attenuation in the lower frequencies and the radio transmission equipment and antennas are easier to design, manufacture, and use. In the case of the fixed-satellite services, the C band (at 6 and 4 GHz) was used first. Then the Ku bands (at 14 and 12 GHz) were utilized next and then they became largely saturated. Currently the greatest amount of expansion is in the so-called Ka band (this is the 30 and 20 GHz bands) that requires high power and encoding complexity to overcome rain attenuation issues. Despite the efficiency gains that come with the use of higher power, high-gain antennas, and coding complexity the current commercial satellite frequencies (in C, Ku, and Ka band) wiil eventually saturate. This is because of ever increasing demand for broadband video and data services and expanding access to users around the world. Further the satellite allocation for C-band was reduced as a fully protected service.

The next frontier thus seems to be the so-called Q/V bands of 47.2–50.2 GHz and 37.5–40.5 GHz, and beyond that, the expansion will be to even higher frequencies such as the W band, the terahertz (THz) band, and even the lightwave frequencies. The use of such high frequencies with ever-shrinking wavelengths is a challenge for satellite service, because atmospheric conditions make the use of such spectrum very difficult indeed. The one area where satellites have an advantage would be for transmissions that occur above the Earth's atmosphere. The use of light waves or laser communications for intersatellite links (ISLs) or cross-links to connect satellites in orbit is not only possible but is starting to be used for this purpose. Laser cross-links for low Earth orbit constellations is easiest, but this is also possible for medium Earth orbit constellations or even GEO satellites.

There are many challenges represented by the higher radio frequency (RF) bands. These challenges include building radio equipment that can operate effectively and efficiently at these exceeding challenging frequencies. The great challenge is to utilize these microscopic wavelengths and to cope with the atmospheric interference that tends to block the signals at the Q/V and W bands and higher. Here it is a matter of not only rain scatter of the signal but also the

oxygen absorption, scintillation, and other problems that weaken, distort, or otherwise interfere with satellite signals in these millimeter wave and even the THz frequency ranges. Light-wave transmissions from ground to space and back constitute an even greater difficulty.

Nevertheless, satellite communications systems of the future can be expected to operate in these challenging frequencies. In order to do so, however, new modulation and multiplexing equipment, signal regeneration, coding systems, antennas, and power systems will likely all be needed to deliver secure and reliable service in the future. In the meantime, better coding processes, higher power transmission, and improved and higher-gain antenna can extend and expand the efficiency of usage of the lower-frequency bands. It is possible that instead of such a heavy reliance on satellites in the GEO orbits, lower orbit satellite constellations and high-altitude platforms (HAPS) can also be deployed in the Ku band and Ka band to provide additional throughput capabilities. A third factor to consider in assessing future demand is the additional build-out of high-capacity fiber-optic networks. These extremely broadband systems could also serve to reduce the demand for future satellite services. Nevertheless, the demand for mobile, rural, and remote services plus broadcasting and multi-casting services should still sustain satellite growth for some time to come.

This chapter, in particular, focuses on the technical, operational, and practical issues associated with the development and future deployment of future satellites in the Q/V and W bands. It also briefly discusses the even higher terahertz frequencies and light-wave or laser transmission.

Keywords

Aldo Paraboni Q/V band hosted payload • Allocation of frequencies • Alphasat • Antenna design • C band • Coder/decoder (codec) • European Space Agency • High-altitude platform systems (HAPS) • Hosted payloads • Intersatellite links (ISLs) • Ka band • Ku band • Oxygen absorption • NASA • Q/V band • Rain attenuation • Scintillation • Terahertz (THz) frequencies • US Air Force Department • W band

Introduction

The demand for new satellite frequencies has mirrored the worldwide demand for additional frequencies as more and more applications for radio frequencies have evolved over the years. The desire for additional terrestrial applications in the broadband cell phone services – and especially the most recent fourth-generation long-term evolution (LTE) services – has been enormous with over seven billion cell phones of various types now in operation. Long-distance microwave, military applications, instructional television, and satellite communications have ended up losing radio frequency allocations in the lower very high-frequency (VHF) and ultrahigh-frequency (UHF) bands to this ever-expanding demand for terrestrial cellular service. The current rapid build-out of new satellite systems to claim more and more of the available assignments of frequencies in the GEO orbit now accentuates the need to prove the viability of systems that can operate in the remaining allocated bands for satellite communications.

Technically the overall bands are 35–75 GHz for Q/V band and 75–110 GHz for W band. The specific bands that have been allocated to geosynchronous satellite communication for the Q/V band are 37.5–40.5 GHz for the downlink and 47.2–50.2 GHz or a total of 3000 MHz of spectrum. A lesser amount of only 1000 MHz at the lower end of these bands are available for non-GEO constellations.

The available spectrum for the W band is even wider. Here the International Telecommunication Union (ITU) allocation is for 71–76 GHz frequencies for downlinking and 81–86 GHz for uplinking. This means a remarkable 5000 MHz is potentially available for commercial services in this very high-frequency band.

These new bands are different in terms of their readiness for commercial use. In the Q/V bands there have been experimental satellite packages flown, actual applications for satellite systems processed through the US Federal Communications Commission (FCC) and the start by ground system manufacturers to build equipment for these frequencies. In the case of W band, only military experiments are now planned, and no immediate commercial systems are anticipated around the world. The technical challenges for Q/V band systems are indeed great, but the difficulties of W band, THz frequencies, and laser communications are far greater still in terms of both space and ground systems. Therefore, this chapter of new frequency bands for commercial satellite communications will be addressing these new and ever higher-frequency bands separately.

Q/V Band for Satellite Communications

There is only one in-orbit experimental package currently testing the feasibility of using the Q/V-band frequency. This is the Q/V experiment that flew on the Inmarsat Alphasat. On July 23, 2013, the large-scale commercial satellite Alphasat was launched carrying several hosted payload experimental packages aboard. Of the experiments, one of the most significant was the Alphasat's "Aldo Paraboni Q/V band" hosted payload. In late January 2014, tests on this experimental package began. A series of signals were transmitted across Europe using the package's three spot beams with the uplink signal being at 48 GHz and the downlink signal at 38 GHz. Dr. Aldo Paraboni is the scientist who first conceived of these experiments. He felt that it was critical for such tests to be carried out before seeking full-scale commercial satellite system deployment in these challenging frequencies.

The other experimental payloads that included laser and other experiments by ESA and other experiments were carried as hosted payloads on Inmarsat's Alphasat, the largest European telecom satellite ever built at 6.6 metric tons. This huge satellite for maritime communications is now in its final orbital position at 25°E and is providing mobile satellite communications in the Atlantic Ocean region, while the Q/V-band tests are also being conducted. Alphasat and its hosted payloads are also

Fig. 1 The prelaunch Inmarsat Alphasat satellite displaying the hosted payload experimental packages (Graphic Courtesy of ESA)

the result of one of European Space Agency's largest public–private partnerships to date for commercial satellite experiments. This particular project included participation by ESA, Inmarsat, and a dozen institutional and industrial partners from the European space community (Alphasat) (Fig. 1).

The experimental objectives for the Q/V-band tests included the following:

- To demonstrate the effectiveness of PIMT (propagation impairment mitigation techniques) in improving the achievable data throughput for a Q/V-band satellite link
- To test ACM (adaptive coding and modulation), based on the European-developed digital video broadcast standard (DVB-S2) to improve effective throughput
- To test the effectiveness of on-demand uplink power control (ULPC) to increase throughput in rain attenuation conditions
- To test these systems in a dynamic trade-off between overall service availability and efficiency of transmission
- To experiment with adaptive transmission schemes for the PIMT system to find the most effective and efficient system over links using the DVB-S2 standard

using the three spot beams on the experimental package to carry out dynamic tests in different rain and atmospheric conditions

These tests, carried out over some 2 years, have shown that adaptive coding and modulation based on the DVB-S2 standards work with a good degree of effectiveness (Giuseppe Codispoti).

NASA has also been conducting research in the area of millimeter wave transmissions to support future satellite communications. Although rain attenuation is the area of most concentrated concern, there are a number of other key areas that propagation measurements are taken to assess performance at these very high frequencies. In the proposed NASA experimental measurements program that would take RF propagation measurements at 27.5 and 76 GHz, the objective would be to take readings and consider the impact of all of the following factors that are quite challenging in Q/V and especially W-band frequencies:

- **Scintillation** – Rapid fluctuations in the refractive index of the atmosphere can cause rapid variations in the attenuation of the propagating signal. These scintillation effects occur with a half of a second or less and will limit the performance of high data rate transmissions, especially those that are using complex modulation and coding systems.
- **Depolarization** – For maximum spectral efficiency (frequency reuse systems), it is important to characterize signal leakage between polarizations of the same signal via simultaneous co-polarization and cross-polarization measurements. This can especially occur during localized rain events.
- **Group delay** – The advantage of exploiting millimeter bands is the tremendous amount of bandwidth available to operate high data rate systems. However, group delay, or dispersion, across the bandwidth can limit the exploitation of the entire available spectrum.
- **Atmospheric noise** – The atmosphere through which millimeter wave transmission must transit has an equivalent black-body temperature that is about 10–40 K closer to the ambient temperature. This is a larger problem for transmissions in the Q/V bands and W bands.
- **Precipitation-covered ground antenna** – Dew, snow, and condensation on the antenna that is particularly prone to distortion because of the small nature of the wavelengths can cause signal losses. These losses can be as large as several dB in the W band (Acosta et al. 2015).

NASA has not developed experimental packages to test transmission in the Q/V band, but the experiments with the Advanced Communications Technology Satellite (ACTS) did allow tests of quite similar techniques to those being tested on the Alphasat Q/V experimental package. These tests included power, dwell time, and other "on-demand" strategies to cope with specific geographic areas experiencing significant rain attenuation. The results from the ACTS tests in Ka band are thought to be generally transferable to the higher-millimeter wave frequencies.

Compensation Techniques to Improve Satellite Transmission

The compensation techniques tested on ACTS and the current Alphasat Q/V-band package include the following types of approaches:

- Enhanced power on demand in antenna beams experiencing significant rain attenuation and other transmission efficiency effects as noted above
- Increased dwell times on demand for TDMA and CDMA multiplexing systems in beams experiencing rain attenuation, etc.
- Advanced coding techniques such as turbo-coding (and in the current Q/V-band experiments tests of the DVB-S2 standard) to test means to compensate for rain attenuation, etc.
- Active onboard signal processing and regeneration to break down uplink signals back to baseband and thus regenerate the signal to restore its full integrity before it is downlinked. This type of regeneration of signals onboard the satellite can, for instance, during the heaviest rainstorm provides a 10–14 db advantage.

Proposed Q/V-Band Commercial Satellites

Remarkably a total of 16 satellite systems that would have operated within the Q/V bands were proposed to the US Federal Communications Commission by US companies almost 20 years ago in 1998. Of these, all 14 systems as presented in Table 1 were intended to provide global, or nearly global, service. One other was intended for US domestic service, and one was to be a package to provide additional store-and-forward capability on an earlier proposed "Little LEO" system. Despite these extensive and formal proposals to the FCC, none of these systems were ever initiated and thus never built and deployed. As can be seen in the table, some of these systems proposed to use a variety of orbits that include the geosynchronous orbit, constellations in medium Earth orbit, low Earth orbit, and Molniya orbit or some combinations of two of these orbits. Most of these new systems proposed to employ new technologies such as multiple narrow spot beam antennas, onboard demodulation, processing and routing of traffic between beams, intersatellite links, and in some cases scanning beams to continuously illuminate the service area as the satellite passes overhead. In short, these imaginative and forward-looking filings with the USA seemed to foresee a massive new movement to use the broadband allocations available in the Q/V bands. But other advances in the Ku band and Ka band have postponed a movement up to the Q/V bands.

The 14 filings for global or near-global systems as summarized in Table 1 are derived from the article prepared by John Evans and A. Dissanayake in an IEEE article written in 1998. Almost all of these systems, as filed with the FCC, used several of the abovementioned techniques designed to improve transmission effectiveness during precipitation attenuation (Evans and Dissanayake).

This table is remarkable in that it reflects $45 billion of proposed satellite system investment, all within official filings to the US Federal Communications for a total of

Table 1 Listing of proposed satellites in millimeter wave band in 1998

Proposed global or new global Q/V satellite systems						
Company	System	Orbit	No. of sats.	Satellite throughput (Gb/s)	Intersat link	Capital invest. (US $)
Denali Telecom LLC	Pentriad	Molniya	9	36 Gb/s	No	$1.9 billion
GE Americom	GE*Star Plus	GEO	11	About 70 Gb/s	27 optical	$3.4 billion
Globalstar L.P.	GS-40	LEO	80	About 1 Gb/s	No	$7 billion
Hughes Comm. Inc.	Expressway	GEO	14	About 65 Gb/s	Optical 3 Gb/s	$3.9 billion
Hughes Comm. Inc.	Space Cast	GEO	6	Abourt 64 Gb/s	Optical 3 Gb/s	$1.7 billion
Hughes Comm. Inc.	Star Lynx	GEO and MEO	4 20	About 5.9 Gb/s About 6.3 Gb/s		$2.9 billion
Lockheed Martin	Q/V band	GEO	9	About 45 Gb/s	3 optical/ 2 radio	$4.75 billion
Loral Space & Comm. Ltd.	Cyber Path	GEO	10	17.9 Gb/s	2 radio	$1.17 billion for four sats w/ possibility of six more
Motorola	M-Star	MEO	72	About 3.6 Gb/s	27 radio	$6.4 billion
Orbital Sciences Corp.	Orblink	MEO	7	About 75 Gb/s	2 radio	$0.9 billion
PanAmSat.	VStream	GEO	12	About 3.2 Gb/s	27 radio	$3.5 billion
Spectrum Astro. Inc.	Aster	GEO	25	About 10 Gb/s	2 optical	$2.4 billion
Teledesic	VBS	LEO	72	4 Gb/s	4 optical	$1.9 billion
TRW	GESN	GEO and MEO	14 and 15	About 50 Gb/s About 70 Gb/s	10 optical 4 optical	$3.4 billion
Totals			378 satellites			$45.2 billion

some 378 new satellites. And this was in US government filings nearly 20 years ago in 1998. If the Loral CyberPath system had been completely built out to complete a global network of ten satellites (as opposed to just the first four GEO satellite as costed in their proposal), then this would have actually entailed 384 satellites and a total investment of over $47 billion of space segment facilities (Evan et al.). Yet none of the proposed satellites were ever built and deployed. It turns out that these filings were only an attempt to stake a claim to the very large allocations in Q/V-band spectrum by the would-be Q/V-band satellite operators. Instead, system network operators have found that intensive use of the Ku bands and Ka bands was a more cost-effective solution. A significant part of the reason is not only the space segment cost, but also the ground systems cost would have been very substantial. Until Q/V-band ground antennas were designed, developed, and manufactured in high-volume production, this will remain a major impediment to use of this new satellite band.

Commercial satellite systems intended to serve a consumer or business market must be designed to operate with relatively low cost terminals if they are to be commercially successful. It has taken several years for Ku-band ground terminals and then Ka-band ground terminals to be reduced from relatively expensive to moderate costs. It will undoubtedly take some time for the development of Q/V-band ground antennas and even longer to see these produced in volume and available at reasonable cost.

While the indoor electronics portion of the terminal can largely be constructed of application-specific integrated circuits (ASICs) at low cost, this is not true of the outdoor antenna and related electronics of the feed system. For this part of the ground system, solid-state pseudomorphic high-electron mobility transistors (P-HEMTs) will likely be used, but these are relatively low-power systems and it is unclear whether these can provide sufficient amplification of the satellite signal for the user terminals. For receiver applications, P-HEMTs used in the low-noise amplifiers need high-volume production to get costs down. Thus there is a vicious circle of the need for high-volume production to get costs down, but high-volume usage depends on reasonably low costs. The parallel trend analysis that has seen the cost of Ku-band ground antennas drop over time and that the same process is now working for Ka-band ground antennas is nevertheless encouraging. The success of the tests being conducted on the Q/V onboard package on the Alphasat and the very wide 3000 MHz broadband spectrum available will likely lead to deployment of satellite networks in the coming decade.

Factors that could serve to delay the deployment of Q/V-band satellite systems, however, include the following:

- The huge capacity of the current high-throughput satellites (i.e., Ka-band and Ku-band systems by Via Satellite, Intelsat, Hughes Network Systems, etc.) and related concerns about the oversupply of satellite system capacity. (The launch of these systems all within a few years of each other provide for a fourfold increase in global satellite capacity.)
- The various new constellations of low and medium Earth orbit satellites that are being planned to provide Internet-optimized services. (On one hand, there is concern about the oversupply of service capacity that such megaLEO satellite systems like the OneWeb and the SpaceX networks might entail. On the other, if these systems should prove financially unviable and end in bankruptcy, it might

serve to dry up capital investment to support the new and presumably more risky Q/V satellite networks.)

- Finally, the development of high-capacity fiber-optic networks in urban areas in combination with broadband mobile cellular systems and other options such as space-based optical rings (see Laser Light Communications below), high-altitude platform systems, and projects like high-flying balloons with wireless Internet platforms are seen as limiting future satellite demand.

W-Band Satellite Networks

Although NASA has not initiated frequency propagation and transmission experiments in the Q/V band, it is undertaking a joint development and research program in the W band. Currently, the Air Force Research Laboratory, Space Vehicles Directorate (AFRL/RV), in collaboration with the Space and Missile Systems Center, Military Satellite Communications Directorate (SMC/MC) and NASA Glenn Research Center, is conducting new fundamental research to study the atmospheric effects on radio frequency signal propagation in the W band. Specifically, the objective is to statistically characterize channel propagation effects in the 71–76 GHz and 81–86 GHz bands. Their joint study is exploring signal attenuation, phase dispersion, and depolarization. This is quite similar to the experimental objectives of the ESA Q/V experiments on Alphasat, but this program is being undertaken by ground-based measurements.

The measured data from this project, however, will be used to develop new modeling and design tools that can be used to design and assess future military satellite communication architectures, rather than in the pursuit of commercial satellite network design (W/V band).

Despite the attractiveness of the 5000 MHz that is available in the W band, the difficulties of these frequencies between 70 and 86 GHz are sufficiently daunting that it does not seem likely that commercial communications satellite networks will be designed and deployed in any reasonably near-term framework and even military satellite communications systems seem a good ways away.

Terahertz (THz) Frequencies for Satellite Communications

The radio frequencies about 100 GHz are largely unassigned for specific uses by the International Telecommunication Union. This means that from 0.1 to 30 THtz, there is a huge amount of spectrum that might be used if suitable and cost-efficient applications were to be found. In light of the extremely difficult atmospheric interference problem that prevents use of these frequencies for satellite communications, there are obvious problems with this application for satcom links. The use of the THz RF bands for intersatellite links (ISLs), which operate above the Earth's atmosphere, could be an attractive possibility. Researchers in this area suggests that very broadband intersatellite links (ISLs) that would be capable of data relays in excess of 10 Gb/s could be designed and built at lesser cost than optical ISLs. These researchers claim that THz wave transmitters and receivers could be designed to combine the advantages of both optical (i.e., broader bandwidth) and microwave (i.e., greater simplicity and lower cost) in the context of designing future intersatellite links. THz transmission systems are different from optical links in that the THz wave is not visible. There are also problems with free-space coherent optical communications that involve the need for high-cost intensity modulation/direct detection (IM/DD) devices. Such optical links entail costly and complex equipment including higher powered devices and narrower beam tracking than is the case for RF-based systems.

There are some researchers that are proposing the development of broadband (i.e., greater than 10 Gb/s) terahertz ISLs for relays between satellites. These THz relays would then be converted to Ka band for Earth to space transmission or space to Earth links. Initial research results suggest that such ISL applications could be technically possible for GEO to GEO satellite links, MEO to MEO, LEO to LEO, or LEO to GEO for data relay satellites that should thus continue. Ultimately such applications might also be developed for a satellite to Moon (i.e., cislunar link) or even links to Mars using THz links might prove technically feasible and cost-effective (Han et al. 2015).

Optical Links for Satellite Communications

There has been over 20 years of development of optical communications links using satellite communications. The Japanese communications development programs to develop optical communications systems date back to at least the Experimental Test Satellite VI (known as Kiku) that included a laser communications experiment.

On August 28, 1994, Japan launched the ETS-VI that included an optical payload on board known as the laser communications experiment (LCE). LCE was one of the world's first space demonstrations of optical ISL technology and was conducted by the Communications Research Laboratory (CRL) of the Ministry of Posts and Telecommunications (MPT) of the Japanese government that is now known as the National Institute for Information and Communications Technology (NICT). The LCE laser downlink used a GaAlAs semiconductor laser that operated at 0.83 μ wavelength and the uplink used an argon 0.51 μ wavelength laser. The tests were conducted using a ground observatory near Tokyo (See Fig. 2).

The small laser communications package weighed only about 22 kg and consumed only about 80 W when operational. The 7.5 cm telescope on the spacecraft had a fixed optical beam pointed by a steering flat mirror. This design feature of the steering mirror reduced the difficulty of the acquisition, tracking, and pointing system. But this greatly reduced the data rate handling capability. The ground telescope diameter was 1.5 m. The space package was able to acquire the ground beacon spatially by using a charge-coupled device (CCD) with a field of view of 8 mrad. Fine tracking for the LCE was accomplished via a four-quadrant avalanche photo detector (APD) and two single-axis fine-steering mirrors. The flight package

Fig. 2 The ETS-VI Japanese satellite that conducted some of the first laser communications experiments (Graphic Courtesy of CRL of Japan)

was able to demonstrate laser communications in space during periods of visibility in clear sky conditions. This set the stage for the Japanese Optical Inter-orbit Communications Engineering Test (OICET) satellite experiments that took place over a decade later (Brandon et al. 1998) (Fig. 2).

OICET (Kirari): The Optical Interorbit Communications Engineering Test Satellite (OICETS) was named “Kirari” in Japanese. This low Earth orbit satellite was developed by NICT and the Japan Aerospace Exploration Agency (JAXA) to perform optical interorbit communication experiments. The OICET satellite was able to use a laser beam to communicate with another satellite tens of thousands of kilometers away. Kirari was designed to be able to reposition its pointing attitude, and its optical antenna could point not only toward a geostationary satellite but also toward a ground station. Laser communication experiments between the ground and this low Earth orbit (LEO) satellite were carried out from the roof of NICT’s optical ground station (OGS) in Koganei, Tokyo. This flexible pointing capability allowed the experiments with ground optical telescopes prior to the launch of the SILEX experiment on the European Artemis. The Artemis-OICET laser connections were ultimately completed from March to May 2006 (“Satellite Laser Technology”) (Fig. 3).

Artemis: The Artemis satellite was launched in December 2005. This European Space Agency (ESA) satellite conducted the world’s first bidirectional intersatellite laser communication experiments by successfully connecting between the Japanese OICET Kirari and the Advanced Relay and Technology Mission Satellite (ARTEMIS). Perhaps the key element on the Artemis satellite was the semiconductor laser intersatellite link experiment (SILEX) system. The SILEX was able to support around 50 Mb/s of throughput using a 25 cm telescope. At the other end of the transmission was the OICETS. This was designed to support a LEO to GEO optical ISL connection with the SILEX package on the ARTEMIS.

The experiments by the Japanese, Europeans, and Americans, as discussed below, have all focused thinking on how to use cross-links in satellite constellations and even more advanced communications systems. Dr. Takashi Iida, former director of

Fig. 3 The OICETS spacecraft that conducted laser communications experiments with the SILEX optical package on the European Artemis satellite (Graphics courtesy of JAXA)

the Communications Research Laboratories of Japan, who helped to engineer the ETS-VI and OICETs, began thinking about such advanced concepts while working with the author of this chapter at the University of Colorado, Boulder, in the late 1980s. He first wrote about the idea of a global ring of optical satellites at that time. In a book about the future of satellite communications, he published a chapter outlining the technology and the implementation strategy concerning how a ring of satellites in medium Earth orbit with laser communications cross-links could create a very broadband global telecommunications infrastructure. Thus, Dr. Iida was one of the first to explain how such an idea could be accomplished. His ideas and those of other researchers in optical satellite communications now seems to be coming forward toward reality. The actual work in this area follows in the next section (Iida et al. 2003).

US Optical Intersatellite Link (ISL) Experiments and the Canceled TSAT Program

NASA and the Jet Propulsion Labs (JPL) have spent a good deal of effort to develop optical space communications systems. In the case of the JPL development, the prime objective was to develop a capability for high data rate interplanetary communications. The complication in this area is that the US development was moved to a classified program and the US defense research activity assumed control of this technological research. This led to the plans for the so-called Transformational Satellite Communications (TSAT) program that was announced in 2004.

This program was to feature 10 Gb/s laser cross-links between satellites and between satellites and high-altitude manned and unmanned aircraft. This program was to include an eight-satellite TSAT global ring constellation that was envisioned to be fully operational in 2016, when announced by the director of the US Department of Defense (DOD) Military Satellite Communications (MILSATCOM) Joint

Program Office at Los Angeles Air Force Base in El Segundo, Calif. TSAT's data links to ground stations were designed be extremely high-frequency (EHF) RF links able to move data at rates up to 2 Gb/s (Keller 2004).

This program was subsequently canceled by the US Congress during the budgetary process to cut costs as part of the 2010 US Federal Fiscal Budget and the termination of the Lockheed Martin $2 billion contract. The decision was to instead purchase more of the high-throughput Advanced Extremely High Frequency (AEHF) Milstar III satellites to support the US net-centric warfare capabilities. Although this was much less of a capability than the planned TSAT program, these Milstar III AEHF satellite were still able to provide over ten times the capacity and six times better data rate transfer than the current Milstar II satellites (TSAT).

Currently there is a commercial project called "Laser Light Communications" that has indicated plans to design and deploy a laser-based global satellite ring that is an updated version of the TSAT program.

Laser Light Communications

The planned Laser Light Communications constellation is seeking to establish a global IP backbone in space. It is still too early to confirm whether this project will be able to raise the capital funding and necessary launch arrangements to implement this system. Even if this project is unsuccessful, it seems that others may in time be able to proceed to create such a laser-based global IP network in space.

This all-optical network is to be comprised of eight medium Earth orbit satellites plus four spares. It is seeking to create a massive backbone capability with an operating system that is estimated to have a total capacity of 7.2 terabits/sec (Tbps).

Laser Light's All Optical Hybrid Global Network, known as "HALO," will integrate with the existing communications infrastructure to help Optus and other regional fixed-satellite service (FSS) providers extend data throughput and range. This network is being planned in partnership with Australia's Optus and perhaps other partners (The Speed of Light).

This will include satellite to satellite optical cross-links and sat-ground optical up-/downlinks that will be able to sustain digital speeds of over 200 Gb/s, without reliance on radio frequency transmissions. Laser Light™ intends to interconnect its proposed Optical Satellite System (OSS) with the global fiber network – both terrestrial and undersea optical cables. Thus when deployed there will be both terrestrial- and space-based global IP backbone service capabilities. This network will provide worldwide coverage at service levels and connectivity options previously unattainable by other satellite platforms. Such an optical network would be much broader band than even high-throughput satellite networks such as those deployed by Via Satellite, Jupiter by Hughes Network Systems, Intelsat, or others.

Laser Light™'s planned key differentiator is its all optical-wave transmission interface platform, fully integrated via its proprietary ground access nodes, enabling seamless, real-time handoffs to terrestrial fiber carriers at their points of presence.

This compatibility is economically beneficial as the same equipment and optical protocols used in today's fiber-optic transport industry forms the Laser Light™ platform, ensuring carrier-to-carrier interconnection at global co-location points. Severe weather conditions will be bypassed by Laser Light™'s dynamic rerouting capabilities. Laser Light™ intends to design a redundant network topology – based on real-time atmospheric analysis and network redundancy – which affords alternative routing to the customer notwithstanding weather conditions at the point of delivery (Laser Light Communications).

Conclusion

The developmental history of the commercial satellite communications industry can in a way be charted by the way increasingly higher frequencies and smaller wavelengths have been employed over time to meet new and expanded demand for information and telecommunications services.

Satellite systems have worked their way through C-band systems and Ku-band systems and are now implementing Ka-band systems at a rapid pace.

The current European experiments with the Q/V package on Alphasat may well clear the way for the future deployment of new commercial or governmental communications satellites in the 3000 MHz of spectrum that is available in this band. There remain challenges not only in terms of the spacecraft but especially in designing and manufacturing cost-effective ground systems in this new millimeter wave band. Since this spectrum is difficult to use because of rain attenuation and other issues, and because of alternative technological solutions to obtaining additional space-based capacity, the implementation is still perhaps a decade or more away.

This means that the even more difficult to use spectrum in the terahertz spectrum (i.e., from 0.1 to 30 THz) is still further away, even though some experiments by military research agencies are moving forward to explore future implementation. Another option that has been identified is to use this spectrum for satellite cross-links as an option to laser-based intersatellite links (ISLs).

The area where there has been continuing research and development for over 20 years that seems to be on the brink of major new developments is in the area of optical (or laser-based) intersatellite links (ISLs). Here the number of applications can be of many different types. The uses can be for cross-links in constellations such as in LEO, MEO, or even GEO to GEO connections at very high speeds. There can be applications such as to connect LEO to GEO data relay satellites to support surveillance or remote sensing satellite operations. It is also possible that this technology could be used effectively for interplanetary or cislunar links in a new generation of deep space telecommunications systems.

One of the most interesting concepts would be to create a ring of optical-based satellite connections that could create a very high speed broadband IP-based network in medium Earth orbit that could be connected to the ground by either millimeter wave units or optical links that could be flexibly routed around the world to always maintain clear sky connections. Such a system was contemplated some time ago as a global communications utility. Then it was seriously undertaken by the US military TSAT program as a multibillion-dollar project to be implemented by Lockheed Martin. This project, however, was canceled in 2010 due to budgetary constraints. Most recently, a new project to create a global laser-based ring of optical satellites has been initiated by Laser Light Communications. This project is still seeking capital funding to complete its network.

The bottom line is that the desire for expanded spectrum to meet consumer, business, and governmental/military communications needs will drive demand to use higher and higher spectrum that will ultimately likely include the millimeter, terahertz, and light-wave bands as technological advances makes use of this spectrum both possible and economically viable.

Cross-References

▶ Future of Military Satellite Systems
▶ Trends and Future of Satellite Communications

References

R. Acosta, J. Nessel, R. Simons, M. Zemba, J. Morse, J. Budinger, W/V-band RF propagation experiment design (2015), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20120016067.pdf. Last accessed 3 Dec 2015

Alphasat's pioneering high-frequency hosted payload set for experiments (2015), http://www.esa.int/Our_Activities/Telecommunications_Integrated_Applications/Alphasat/Alphasat_s_pioneering_high-frequency_hosted_payload_set_for_experiments. Last accessed 3 Dec 2015

W. Brandon, V. Chan, R. Kwan, Dec. 1998 Japanese research satellite projects (1998), http://www.wtec.org/loyola/satcom/c6_s1.htm. Last accessed 3 Dec 2015

J.V. Evans, A. Dissanayake, IEEE, 1998 Prospects for commercial satellite services at Q- and V-bands (IEEE, 1998), http://www.argreenhouse.com/society/TaCom/papers98/01_01i.pdf. Last accessed 3 Dec 2015

Giuseppe Codispoti, Italian space agency Feb. 13, 2014 "The Q/V band experiments". Presentation to the United Nations Committee on the peaceful uses of outer space (2014), http://www.unoosa.org/pdf/pres/stsc2014/tech-24E.pdf. Last accessed 3 Dec 2015

H. Han, J. Yuan, J. Tong, Design of the THz space application system. J. Comput. Commun. **3**, 61–65 (2015). doi:10.4236/jcc.2015.33011. Published Online March 2015 in SciRes, http://www.scirp.org/journal/jcc. Last accessed 3 Dec 2015

T. Iida, J.N. Pelton, E. Ashford, Satellite communications in the 21st century: trends and technologies, vol. 202 (AIAA, Reston, 2003)

J. Keller, Optical links are key to next-generation military communications satellite, 1 Apr 2004. http://www.militaryaerospace.com/articles/print/volume-15/issue-4/departments/electro-optics-watch/optical-links-are-key-to-next-generation-military-communications-satellite.html. Last accessed 3 Dec 2015

Laser Light Communications (2015), http://laserlightcomms.com/faqs.php. Last accessed 3 Dec 2015

“Satellite Laser Technology”, National Institute of Information and Communications Technology (NICT), http://www2.nict.go.jp/wireless/spacelab/lasersatellitetech/en/03past/past4.html. Last accessed 3 Dec 2015

The Speed of Light, Laser light and optus explain optical communications partnership to via satellite magazine, via satellite, 4 May 2015, http://www.laserlightcomms.com/newsroom.php. Last accessed 3 Dec 2015

TSAT, Timeline and recent developments, http://www.defenseindustrydaily.com/special-report-the-usas-transformational-communications-satellite-system-tsat-0866/. Last accessed 3 Dec 2015

W/V-band Satellite Communications Experiment (WSCE) program, U.S. Department of the Air Force, Solicitation Number: BAA-RVKV-2014- January 14, 2014 0002, https://www.fbo.gov/index?s=opportunity&mode=form&id=f54c60dd11d6f8473297918b46b3e5f6&tab=core&_cview=1. Last accessed 3 Dec 2015

Satellite Radio Communications Fundamentals and Link Budgets

Daniel R. Glover

Contents

Abstract

Satellite communications makes use of radiofrequency links. Particular frequencies are allocated for satellite communications through international regulatory registration and coordination processes which prevents interference between systems. In typical operation, a satellite's transponder receives an uplinked signal from Earth, changes its frequency slightly to avoid self-interference, and retransmits it on a downlink to Earth. Antennas provide gain by focusing the transmitted energy. Path loss describes a natural spreading out of the transmitted wave front as it travels through space. A link budget is an accounting of gains and losses throughout a system that is used as a design tool to provide sufficient power (or gain) to allow a satellite connection to be established. The link margin is the

---

D.R. Glover (✉)
International Space University, Rocky River, OH, USA
e-mail: danglover@gmail.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_17

excess amount of received signal power above what is required. Shannon's law implies that there are trade-offs possible in a communications system design between power, bandwidth, and complexity.

Keywords

Bandwidth efficiency • Bit error rate • Coding decibel • Effective isotropic radiated power (EIRP) • Forward error correction • Free space loss • Gain line-of-sight • Link budget • Link margin • Modulation path loss • Satellite slot Shannon's law • Signal-to-noise ratio • Spectrum • Trade-off • Transponder • Wavelength

Introduction

This chapter provides an overview of radio communications as they apply to commercial satellite communications systems. This chapter is not intended to be a manual for designing satellite communications systems but rather provides the basic concepts needed to understand a system design and link budget. The general concepts are presented briefly and simply and may only hint at the actual complexity of a satellite communications system. As in any system design, satellite communications must trade-off several parameters where a choice for a given parameter results in constraining the choices for the other parameters. The major trade-offs are between power, bandwidth, and system complexity.

Basic System Concepts

One of two images may spring to mind when the term "satellite communications" is encountered. The image may be that of a "satellite dish" antenna on the ground. Or one might picture a spacecraft in space with its antennas and solar arrays deployed. These two images represent two major subdivisions of a satellite communications system: the flight segment (or satellite) and the ground segment (or the ground station with its associated antenna being the predominant feature). The flight segment is the part of the system that makes it a satellite communications system, but it is the ground segment that interfaces with users and with terrestrial communications systems. The ground segment is becoming a more familiar sight, with large dishes at "satellite farms" or gateways, medium-sized dishes atop retail stores, and small dishes for satellite TV becoming a ubiquitous sight. To provide useful communications capabilities, a system must have at least two ground stations (one providing a transmission and another receiving the transmission) and one satellite that relays the information between the two ground stations. In addition, there is a control segment (which is usually transparent to the user) that controls the spacecraft operations. Of course, a particular system may have many satellites and many ground stations (or user terminals), but for our purposes it is best to start with the simplest picture.

A satellite is an object (in our case, a spacecraft) that orbits around another object (in our case, the Earth). The communications spacecraft has several design constraints placed upon it (and thus, on the overall communications system) because it must be placed in orbit. Spacecraft designs are limited in their mass and volume in order to fit on the launch vehicle that places them into orbit. The mass and volume limits affect the size of the power system on the spacecraft, and so the power available is also constrained. In addition, the space environment (thermal, radiation, atomic oxygen, space debris, micrometeoroids, etc.) imposes constraints on the design (such as parts and material selection).

A spacecraft may be considered as consisting of two parts: the spacecraft "bus" and the payload. The spacecraft bus provides support services to the payload, while the payload provides the useful, or moneymaking, part of the satellite. Examples of payloads are scientific instruments, remote sensing instruments, navigation service transmitters, or (in our case) communications equipment. A satellite may have one type of payload or a combination of payload types.

The spacecraft bus provides services to the communications payload including power, structural support, attitude control and pointing, propulsion and station keeping, thermal control, commands, and telemetry. These services are provided by spacecraft bus subsystems which may vary slightly in name and content among different organizations or spacecraft manufacturers. The spacecraft bus typically has its own communications system separate from the communications payload which is used to control the spacecraft bus and the payload operations. The bus communications system is spacecraft part of the control segment. The spacecraft bus communications system receives commands from the control segment ground station and also transmits data concerning the state and status of the spacecraft bus and payload to the control segment ground station.

A typical communications payload consists of antennas and electronics designed for the reception and transmission of radio signals. Some processing of these signals is done on board which may be simple or complex depending on the system. In effect, the communications payload is a radio relay station in space with the satellite bus analogous to a tower.

Let us take a look at generic radio communications, then see where satellites fit in. Two radio stations are able to communicate when they are in "line-of-sight" or have no obstructions or barriers between them. Line-of-sight is a term that is used to describe the path between two stations along which an electromagnetic wave may travel without obstruction. As long as electromagnetic waves can propagate from one station's antenna to the other and can be distinguished from background noise or other interfering signals, communications is possible. Due to the curvature and topographic features of the Earth, as the separation distance on the ground between two locations increases, eventually the two stations will not be within line-of-sight and communications will not be possible. For two ground stations that are not within line-of-sight of each other, a relay station may be used to enable communications if the relay station is within line-of-sight of each of the two ground stations. The relay station receives a signal from one ground station and retransmits it to the other ground station. Additional relay stations can be used to extend the distance between the two ground stations.

A relay station might be placed on a tower to allow the line-of-sight to reach farther over the horizon. The higher the tower, the farther the horizon is extended (or the line-of-sight can reach) and so the farther apart the two ground stations can be and still communicate. For two ground stations on either side of the Atlantic Ocean, a relay tower would have to be over 600 km high (above sea level) in the middle of the ocean to have a line-of-sight to each shore. Since such a large tower greatly exceeds the height of the largest structure ever built, it makes sense to use a satellite as a platform for a radio relay station. Even a series of small relay towers across an ocean would be prohibitively expensive, although undersea cables (originally telegraph cables but now using optical fibers and repeaters) have been used since the nineteenth century and are still an effective means of bridging the oceans.

An example of a passive relay is a phenomenon that occurs at certain frequencies at night. Radio waves are affected by the ionosphere, an electrically conductive layer of the atmosphere that contains charged particles from ionization by solar radiation. The ionosphere is dynamic, as illustrated by the aurora, and changes from day to night. Certain frequencies of radio waves have their paths bent by the ionosphere to the point of being reflected when conditions are right at night and so can propagate much farther around the world than normal. Another example of a reflective relay was the experimental Echo satellite (1960) which was a metalized balloon 30 m in diameter placed in low Earth orbit. Other frequency bands can penetrate the ionosphere with minimal effect and can be used for communication with spacecraft.

One parameter that describes electromagnetic waves (such as radio waves) is the frequency. An alternate parameter is the inverse of the frequency, wavelength. The frequency (f) of a wave in a vacuum is related to its wavelength (λ) by the relationship $f\lambda = c$ where c is the speed of light. Communications engineers tend to use frequency except when a physical representation is useful, as in antenna design, where wavelength is typically employed.

So, a satellite in orbit can act as a relay by receiving signals from one ground station and transmitting them back to another ground station that may not be in the line-of-sight of the first ground station. A satellite in geosynchronous Earth orbit (GEO) can "see" a little less than half the Earth's surface, so one satellite is not enough to provide global coverage. The path from one ground station to a satellite then on to another ground station is known as a "hop." It may take more than one hop through more than one satellite to communicate from a ground station to another on the other side of the world.

Three satellites spaced 120° around the equator at GEO can provide almost total coverage of the globe (with the poles being just out of sight). This idea was first put forward by Arthur C. Clarke in 1945 in an article in the magazine *Wireless World* entitled "Extra Terrestrial Relays." An example implementation of this concept is the NASA Tracking and Data Relay Satellite System (TDRSS) which, among other uses, provides global coverage for International Space Station and Shuttle operations.

A ground station or terminal may be a building full of electronics with a large parabolic dish antenna or as simple as a mobile telephone handset or a satellite television receiver in a home. A ground terminal might transmit, receive, or both transmit and receive. A transmitting ground station requires attention to detail in the ground terminal setup and operation as improper transmissions can interfere with other systems. Terminals may be set up as nodes for point-to-point communications between two stations, broadcast, point-to-multipoint, or in a mesh. Small ground stations that are used in satellite communications networks were named VSATs (very small aperture terminals) when they were introduced. The name comes from their relatively small antennas, although more recent ground terminals (such as satellite television receivers) have even smaller antennas.

Two communications systems that try to use the same frequency at the same time in the same location may interfere with one another such that reliable communications is not possible. Sharing frequencies requires careful coordination to prevent interference. Purposeful interference is known as jamming.

The radiofrequency (RF) spectrum is a portion of the electromagnetic spectrum that is in great demand for uses including communications. Many different uses (applications) require some amount of spectrum in order to operate. To prevent interference and to allow efficient use of spectrum, national and international regulations have been put in place to coordinate the assignment of frequencies to applications. The international body facilitating this coordination is the International Telecommunications Union (ITU), a specialized agency of the United Nations (UN). National regulations control the use of spectrum within the borders of a country and supersede international regulations as long as other countries' spectrum use is not affected.

Figure 1 shows the frequency allocations for spectrum in the USA between 10 and 30 GHz. The purpose of this illustration is to indicate the incredible intricacy of the allocations within just a single country. Applications that use a portion of the RF spectrum include television broadcast, radio broadcast, mobile telephones, wireless telephones, consumer goods, microwave ovens, wireless networks, aeronautical and maritime communications, radio navigation, meteorological radar, radio astronomy, space research, microwave relays, multipoint distribution systems, fixed satellite, mobile satellite, direct broadcast television by satellite, and satellite radio broadcast. Frequency allocations are negotiated among the administrations of the world via a process established by the International Telecommunication Union (ITU) that is described separately in the chapter on this subject by Dr. Ram Jakhu. Satellite communications operators, in short, are in competition for spectrum with many other applications.

There are other types of communications besides commercial. The military may use commercial systems or their own satellite systems. There are search and rescue locating systems that use communications technology. Global navigation satellite systems are essentially highly specialized communications systems. There are satellite communications systems for collecting scientific data on Earth and deep space communications systems for scientific spacecraft exploring the solar system.

On a deep space mission, the communications system is somewhat different from that of a commercial spacecraft. There is typically a "low-gain" and a "high-gain" system named after the characteristics of their respective antennas. The low-gain

Fig. 1 US frequency allocations from 10 to 30 GHz (From US Department of Commerce chart “United States Frequency Allocations, The Radio Spectrum” available at http://www.ntia.doc.gov/osmhome/allochrt.pdf)

Table 1 Radar band designations for some frequencies used for satellite communications

Letter designation	Frequency band (approx.)
L	0.4–1.55 GHz
C	6/4 GHz
Ku	14/12 GHz
Ka	30/20 GHz

system performs the functions that the spacecraft bus communications subsystem does on a commercial satellite, namely, command and data handling, tracking functions for navigation, and housekeeping telemetry. The high-gain system provides a wide bandwidth "pipe" for sending back large amounts of scientific data (e.g., image data, video, radar data, etc.). A deep space probe is like a television station in space that needs to transmit a large amount of data back to Earth. The radio system is also sometimes used as a science instrument itself as is planned with the European Space Agency's Mars Radio Science Experiment (MaRS). MaRS will use the radio signals that convey data and instructions between the spacecraft and Earth to probe the planet's ionosphere, atmosphere, surface, and even the interior. The low-gain antenna also can serve as a backup to the high-gain antenna as with NASA's Galileo spacecraft when a failure in the high-gain antenna resulted in the science data being sent back by the low-gain system (allowing much less data to be returned than would have been possible with the high-gain system but much better than nothing).

Satellite communications engineers frequently make use of old radar band designations when discussing frequency bands. Table 1 lists a few common letter designations for bands useful for commercial communications. Another convention is to list the uplink frequency first, then the downlink frequency when a band is used for both uplink and downlink. Two different frequencies are used to keep the uplink and downlink from interfering with each other. Thus, in C band, a frequency band (of, say, 500 MHz) around 6 GHz is typically used for the uplink and 4 GHz for the downlink (Gagliardi 1984).

Transponders

A *transponder* is the electronic portion of the communications relay station that receives a signal, changes the frequency, and retransmits it. The word transponder is a contraction of the words transmitter and responder. Sometimes one might hear the word "repeater" used, but that is something of a misnomer in this case because a transponder intentionally changes the frequency of the received signal, while a repeater usually refers to electronics that receive and retransmit a signal without any intended changes to it other than amplification.

The transponder is the basic unit of the communications payload. Another way of thinking of a transponder is as a communications channel of the satellite. A communications payload may have from 1 to over 100 transponders, with a typical number being a couple of dozen. Each transponder operates in a portion of the total frequency band available for use by the satellite. The bandwidth of an individual transponder might be designed to be 20 MHz or as much as 500 MHz. For many years, the typical bandwidth of a transponder was set at 36 MHz; because of this history, one may still speak of 36 MHz equivalent transponders as a unit of comparison when discussing satellite capacity. The bandwidth required to transmit an analog, studio quality video signal was 36 MHz which was a major driver in satellite communications design in the early days of the business when analog television distribution was an important application. Today, transponders can be found with bandwidths of 26, 27, 36, 54, 72 MHz, or even larger bandwidths.

A transponder that receives and retransmits a signal without affecting it (other than amplifying it and changing its carrier frequency) is said to be operating in “bent-pipe mode.” This term is an analogy that compares information flowing in a communications channel to fluid flowing in a pipe. A “bent pipe” merely changes the direction of the flow without performing any processes on the “fluid.” This is not a perfect analogy (since we know the signal is amplified and the carrier frequency changed) but is good enough to be widely used. A satellite in “bent-pipe mode” does not process the signals in the channels but is only a conduit that provides a little boost (perhaps a better analogy would include a pump at the bend, but that might be making the analogy a little too clumsy to use and still does not incorporate an analogy for the frequency translation).

Originally, satellite transponders were analog devices. With the advent of digital signals, the use of error correction coding was introduced. A regenerative transponder demodulates the signal and provides error correction on board the satellite before modulating and transmitting back to the ground. This has the advantage of correcting errors introduced on the uplink, thus regenerating the original signals before exposing them to potential errors on the downlink.

In contrast to a bent-pipe satellite is the satellite that includes onboard processing of signals. An early example of the use of onboard processing in a commercial communications satellite is the Iridium system. Iridium uses a set of satellites (known as a “constellation”) in low Earth orbit (LEO) to enable personal communications through satellite telephones. Telephone calls are switched on board the satellites and relayed through the constellation and the ground system. Iridium uses inter-satellite links between the satellites in the constellation to route an individual call to its destination.

A signal is a representation of information; that information could be a voice signal, video, data, or a combination. A channel (in our case, a particular transponder’s bandwidth) is spaced around a carrier frequency and has a capacity to contain a signal or collection of signals with a certain amount of information within that channel’s frequency bandwidth.

Antennas

An antenna radiates and/or captures radio frequency waves to and from free space. It is the interface that couples our electronics to free space and enables telecommunications without wires. Antenna design is a specialized topic within telecommunications engineering.

Different antennas have different patterns of radiation. An isotropic antenna radiates equally in all directions (a pattern described as “omnidirectional”). Other antennas concentrate their radiation in patterns known as “lobes” or “beams.” A design may attempt to concentrate all of the radiated energy into one main beam, but there are typically many additional smaller beams, known as side lobes, radiating in unwanted directions. The main lobe of the pattern has a beamwidth that is usually measured as the angle between points on that pattern that are at half the power of the peak of the main lobe. To simplify matters, we will ignore the side lobes and concentrate on the main lobe of the pattern when we talk about the antenna beam.

The antenna pattern characterizes the radiated energy (transmission) of a given antenna as well as its ability to collect energy (receive signals). If you are familiar with optics, it may help to think of an antenna as the radio equivalent of a telescope at optical wavelengths. There are similarities such as diffraction limits, diameter versus field of view, etc., since light and radio waves are both electromagnetic waves (just at different wavelengths). Telecommunications engineers tend to use frequency except in a few instances such as antenna design. With antenna engineers, wavelength is a more pertinent view since they are dealing with physical dimensions of shapes and sizes.

There are many different types of antennas, but the one most commonly associated with satellite communications is the parabolic dish antenna. These dish antennas have a narrow beamwidth, concentrating the energy of the radiated main beam into a smaller solid angle. This means more of the radiated energy reaches, or “illuminates,” the satellite when using a dish antenna as compared to an omnidirectional (or “omni” for short) antenna. A useful analogy is: a dish antenna is to an omni as a searchlight is to a bare light bulb. This advantage of the dish relative to the omni (that is due to focusing the energy) is known as “antenna gain.” The antenna gain of an omni is 1 (or 0 dB), while the antenna gain of a dish antenna is greater than one and related to the diameter (or aperture) of the dish in wavelengths.

Antenna beams allow us to reuse spectrum without interfering with other satellite systems. By using a narrow enough beam, we can communicate with one satellite and not illuminate its neighbors with unwanted electromagnetic energy. The beamwidth of ground antenna systems is the key parameter in defining the separation of spacecraft placed in the geosynchronous Earth orbit (GEO) arc 36,000 km above the equator at sea level. The spacing supporting a single communicating spacecraft at a particular frequency is commonly referred to as a *satellite slot*.

A crude analogy may help to illustrate the concept: Imagine communicating by Aldis lamp (a flashing signal lamp used for naval communications in the twentieth century) using Morse code. In flashing the lamp, the sender illuminates an area (at a given distance) that is a cross section of the beamwidth of the lamp. That illuminated area hopefully includes the intended receiver but may include other unintended receivers. Now imagine being surrounded by a ring of signalers, only one of which is the intended communications partner (or node). Now there are two problems with communicating: first is illuminating only the desired node in the ring (corresponding to a satellite in a GEO slot) and the other problem is filtering out the signal flashes coming from non-desired nodes in the ring and only receiving flashes from the desired node. The first problem is easily solved by using a lamp with a narrow enough beam such that only the desired node is illuminated (perhaps using a flashing laser if necessary). The second, reception, problem can be solved by using a mask with an aperture that emphasizes the desired node (say by cutting a hole in a piece of cardboard that masks the neighbors of the desired node).

If you were being indiscriminant in your illumination and inadvertently included neighbor nodes that you were not trying to communicate with, you would be a bad neighbor. By illuminating non-desired nodes, you are introducing distracting flashes (or interfering noise) into your neighbor's systems. To be a good neighbor and avoid jamming the communications of others, it is required to use a narrow enough beam to illuminate the desired node and not its neighbors.

In a satellite system, the antenna beamwidth solves both problems by illuminating only the desired satellite in the main lobe and masking the neighbors (since they are not in the main lobe of the antenna pattern) in the uplink. In the downlink, the ground station receives strong signals from the desired satellite and only small amounts of energy (which count as noise for our system) from the neighbors because the desired satellite is located within the main lobe (beam) and the neighbors are not. The separation between adjacent satellites needs to be at least the same amount that corresponds to the beamwidth of the antennas that are practical at that frequency (say 3° for C band).

Polarization can be used to cram a few more channels into a given amount of bandwidth. There are two types of polarization: linear and circular. Linear polarization is the same effect that is used in sunglasses to reduce glare. A filter can be used to allow a horizontally polarized wave through while blocking a vertically polarized one. In communications, polarized antennas can be used to capture or reject polarized radiofrequency carriers.

There are two bands of frequencies in the electromagnetic spectrum that are said to be windows through the atmosphere because electromagnetic waves can pass through relatively unchanged. One window is the radio window used for satellite communications (the other is the visible window). At low frequencies, the ionosphere bends the paths of electromagnetic waves to the point of reflection, while at higher radio frequencies, the atmosphere absorbs the waves energy. Water vapor absorbs radio waves, and at high frequencies, such as at Ka band, clouds and rain can produce significant signal loss (see Fig. 2) known as rain fade.

An antenna located at the equator pointing straight up at a satellite in the GEO arc will have a shorter path and look through less atmosphere (optical depth) than an antenna located at a higher geographic latitude. This distance from the ground station to the satellite is the slant range. Another parameter of ground station antenna pointing is the look angle (azimuth and elevation) by which the ground station antenna must be pointed to view a particular satellite. An antenna pointed at a GEO satellite will also be pointed at the sun for a short period of time around the equinox twice a year. The sun is a significant noise source at all frequencies and will cause interference (known as a sun outage) at those two short periods (typically 15 min) each year.

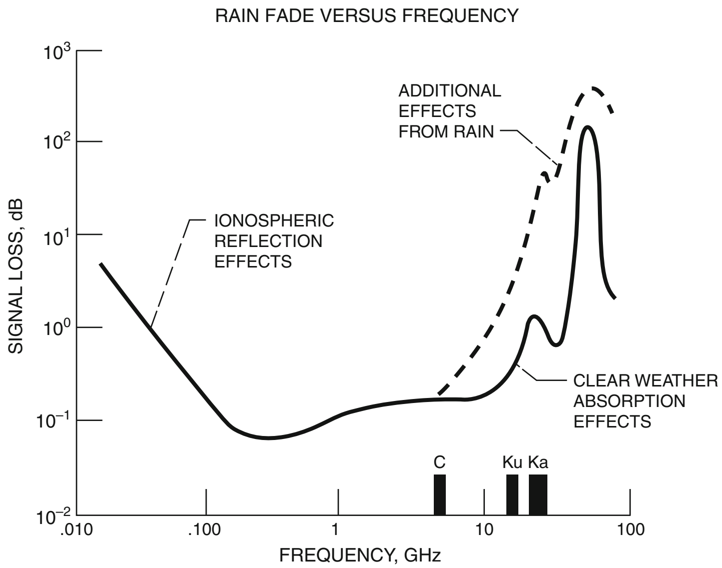


Fig. 2 Rain fade versus frequency (Courtesy NASA)

Digital Communications

Initially, satellite systems started out as analog systems and over the years have increasingly become digital systems. Digital communications allows flexibility in design through the use of compression and error correction coding. Digital signals can be perfectly reconstructed when a system is designed properly, while analog signals tend to become more corrupted in each stage of processing. Converting an analog signal to a digital signal does introduce quantization noise, but the amount of noise added can be selected (by finer quantization and sampling) and is only added once (at the time of conversion).

A signal can be converted from analog to digital through a process called pulse code modulation (PCM) that includes sampling (Jayant and Noll 1984). The analog signal is measured periodically at some sample rate. The samples are quantized to numerical values. The analog signal is reconstructed by essentially connecting the dots of the values at the sampling times. The reconstruction will be poor unless enough samples have been taken and the samples were quantized to a suitable range of values. The sample values can be represented by binary digital values (bits) and strung together into a bitstream. The bitstream (with values of 1 or 0) represents the signal.

The analog signal at its original frequency location and bandwidth, typically from 0 Hz up to some frequency that includes most of the signal energy, is the baseband signal. When digitized, the signal will typically take up much more bandwidth than the original analog signal, but that can be reduced through digital data compression. For example, a typical voice signal for telephony would have a baseband bandwidth of

4 kHz and would require 64 kbps in digital form after PCM at 8 bits/sample and 8 k samples/s. Using compression, the 64 kbps PCM data rate could be reduced to 10 kbps or even 4 kbps (with a quality vs. data rate trade-off). In the digital world, data rate is sometimes referred to as bandwidth and they are closely related. Approximate data rates for some types of signals range from 32 to 320 kbps for MP3 (Motion Picture Experts Group 1 Layer III) compressed audio, 1,411 kbps for CD audio, and 2–40 Mbps for MPEG-2 compressed video (typically ~5 Mbps for standard video resolutions, ~19 Mbps for high-definition TV, ~36 Mbps (1x) for Blu-ray video).

Modulation and Coding

As presented in the previous chapter, modulation is the technique of imposing a signal (say, a baseband signal) onto a radiofrequency carrier (or carriers) for transmission by varying the carrier's amplitude, frequency, or phase in accordance with the signal. A few modulation schemes are demonstrated in the figures below. These highly simplified representations are not to scale; the carrier frequency is typically much higher than the baseband signal.

Figure 3 depicts analog frequency modulation (FM). The carrier wave's frequency is varied in accordance with the change in the information signal's amplitude. Figure 4 depicts digital amplitude-shift keying (ASK or on-off keying) modulation. In this case, a digital string of data is represented by a digital waveform. The carrier amplitude is varied according to the digital information bit stream, on for a 1 and off for a 0. The envelope for this modulated carrier is obviously not constant which would present a problem for most transponders. Figure 5 depicts the digital version of frequency modulation, frequency-shift keying (FSK). Figure 6 depicts

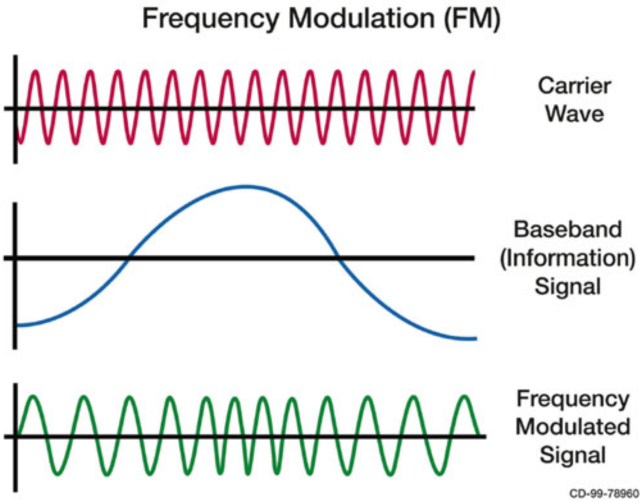


Fig. 3 A cartoon representation of analog frequency modulation, not to scale (Courtesy NASA)

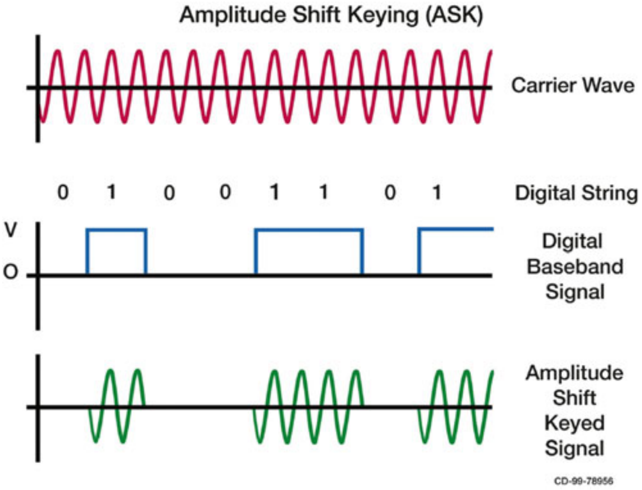


Fig. 4 A cartoon representation of digital amplitude-shift keying modulation, not to scale (Courtesy NASA)

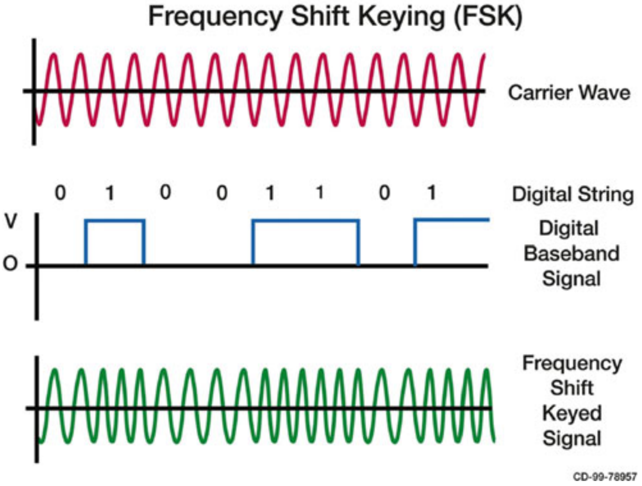


Fig. 5 A cartoon representation of digital frequency-shift keying modulation, not to scale (Courtesy NASA)

digital phase-shift keying (PSK). The digital waveform used is typically a non-return-to-zero (NRZ) representation where a bit value of 1 is represented by a positive value and a bit value of 0 is represented by a negative value, but here a bit value of 0 is represented as 0 for illustrative purposes only. There are many other methods of representing a bit stream as a waveform such as Manchester (or biphase) encoding.

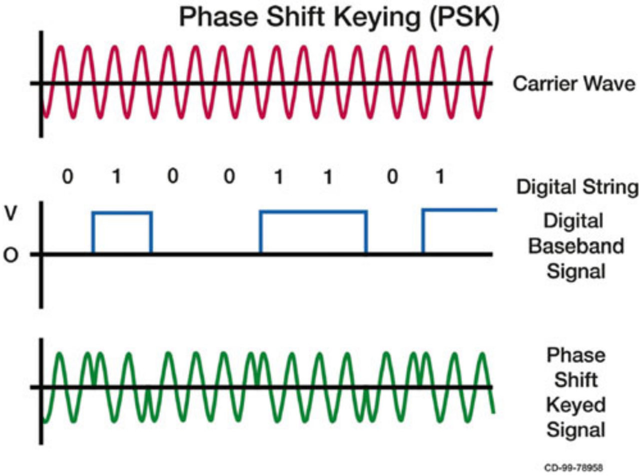


Fig. 6 A cartoon representation of digital phase-shift keying modulation, not to scale. V represents a positive voltage and O is usually a negative voltage but shown here as 0 for illustrative purposes (Courtesy NASA)

Table 2 Bandwidth efficiency of various digital modulation schemes

Modulation scheme	Theoretical bandwidth efficiency
BPSK	1 bit per second (bps)/Hz
QPSK	2 bps/Hz
8-PSK	3 bps/Hz
16-QAM	4 bps/Hz

A version of PSK, quadrature phase-shift keying (QPSK), is widely found in satellite systems. Biphase PSK (BPSK) is sometimes encountered. QPSK is better from the standpoint of bandwidth efficiency as it is equivalent to two independent (orthogonal) BPSK schemes combined. Although BPSK is a simpler method, QPSK chipsets have become widely available making it a cost-effective choice. Bandwidth efficiency refers to the amount of bandwidth required to accommodate a given amount of data. The more bandwidth efficient a scheme is, the more data can be accommodated in a given channel bandwidth, but generally this results in a more complicated scheme. In other words, there is usually a complexity versus bandwidth trade-off to be made in selecting a modulation scheme, among other considerations.

Advanced modulation schemes include 8-PSK, quadrature amplitude modulation (QAM) including 16-QAM, 64-QAM, and 256-QAM where the numbers refer to the number of symbol states in the scheme. Table 2 shows the theoretical bandwidth efficiency for various modulation schemes. A system's bandwidth efficiency is affected by more than just the modulation scheme and will be less than that shown in the table.

System bandwidth efficiency will be less than that shown in Table 7 due to things like pulse shaping (typically raised cosine filtering (Hayken 1983)) and coding. A

QPSK system might have a bandwidth efficiency of 1.5 bps/Hz or less. A 16-phase coding system using the latest technology, however, might use this level of complexity to achieve an efficiency of 5 bps/Hz or better.

An important measure of quality in a digital system is the bit error rate (BER). The BER is the number of bits in error divided by the total number of bits sent. Bit errors can occur randomly spaced or in blocks. The BER is usually assumed to be an average over some period of time. For example, if 1 bit is in error on average for every million bits sent, the BER would be $1/10^6$ or 10^{-6}. Error correction coding, also known as channel coding, can be used to correct some errors. Coding requires that additional bits are inserted which adds information redundancy to the bitstream at the cost of increasing the number of bits that must be sent.

In a simple example of error correction coding, each original data bit could be sent three times. The original data bit is combined with two parity bits that provide a 3-bit representation in the transmission data stream. One bit error can be detected and corrected since it will conflict with two unchanged bits. This triples the number of bits that must be sent compared with the original data, however. If the original bit was only sent once, then a bit error could be detected but not corrected since the receiver will not know which bit is the correct value. It turns out that there are more complicated codes that allow various levels of error detection and correction that use fewer than 3-bit symbols. The amount of redundancy that a particular code adds is described by the code rate where the rate is the number of original data bits divided by the total number of bits (data and parity) that are sent. For example, the rate of our inefficient tripling code above is 1/3.

There are several types of error correction coding including block codes, convolutional codes, concatenated codes, turbo codes, low-density parity-check (LDPC) codes, etc. (Peterson and Weldon 1991). They add varying amounts of complexity and provide varying levels of protection. Source coding refers to data compression coding. Channel coding adds redundancy to the bitstream while source coding removes redundancy. A bit error in a compressed data stream may cause damage to more than one data bit when uncompressed.

Coding gain is a measure of the power savings that a coding scheme achieves compared to an uncoded signal in the same system. The power savings come at the cost of increased bandwidth and complexity as coding adds redundancy to the original bitstream.

Multiple access (Bhargava et al. 1981) refers to the sharing of a communications channel by more than one signal. It is similar to multiplexing but is a bit more complicated because of the interactions between the signals. There are three main types of multiple access: time division multiple access (TDMA), frequency division multiple access (FDMA), and code division multiple access (CDMA). In TDMA, each signal (which may itself be a collection of signals) uses the channel for a specified period of time. This requires complex time synchronization. In FDMA, each signal gets its own frequency band subchannel. In CDMA, each signal is encoded with a pseudorandom code which spreads the signal over a larger bandwidth (Dixon 1976) (CDMA is a spread spectrum technique) and is extracted using the same code. Multiple access usually takes place at an intermediate radio frequency, and there are complex interactions between signals that can affect them.

Intermodulation distortion can occur when several carriers share a nonlinear device, such as some power amplifiers. An amplifier may have a linear transfer function in a portion of its operating curve but may be nonlinear at the high power output portion of its operating curve. The amplifier may be operated at a back off point compared to its maximum power output to ensure that it is operating as a linear device to minimize intermodulation distortion effects if several carriers are sharing it.

Shannon's Law

Claude Shannon was one of the seminal thinkers of the twentieth Century. He worked as a researcher at Bell Labs in its heyday. By applying Boolean algebra to telephone switching circuits, he set the stage for digital electronic design. He developed a mathematical theory of communication that included the foundation of information theory (Hamming 1986). One result of the theory is a description of the limit of the amount of information that can be transmitted on a given channel (that is sometimes known as "Shannon's law" or the "Shannon–Hartley theorem") commonly called the channel capacity. The "Shannon limit" refers the theoretical best rate of transmission for a given channel. The channel capacity equation is

$$R \leq W\ \mathrm{log2}\ (P_R/P_N+)$$

where

R = theoretical maximum transmission rate, bits per second
W = usable bandwidth, Hertz
P_R = power of received signal
P_N = power of noise

The Shannon limit implies that a system can be designed (i.e., with sufficient complexity, e.g., using error-correcting codes) that can allow transmission on a channel that approaches the rate of the Shannon limit. For error-free communications, E_b/N_0 has to be at least -1.6 dB, even with infinite bandwidth. Error correction coding can reduce the power required but uses more bandwidth. By adding coding to the system, complexity and bandwidth are increased, but power required is decreased (i.e., complexity and bandwidth are traded for power).

Link Budgets

Communications is achieved by introducing into free space a modulated radio wave (transmitting) which propagates over a distance to a point where a receiving station captures a portion of the transmitted energy and extracts information from it. As the radio wave propagates through space, it spreads out and the energy transmitted is spread across an increasingly larger cross-sectional area; less energy is available to be captured by a given antenna as the distance increases. Although we have been talking about beamwidth as a two-dimensional measurement characterizing an antenna, the transmitted wave actually propagates in a three-dimensional solid angle. A solid angle can be visualized as a cone whose apex is at the transmitting antenna having a base, or cross section of the beam, at any given distance from the antenna (this approximation is an illustrative aid, since waves are usually assumed to propagate spherically which would mean that the base bulges out as a spherical section). The area of the cross section of the beam gets larger as the distance from the antenna increases and the energy of the transmitted wave is spread over that increasingly larger cross section. The received power flux density drops off inversely proportional to the square of the distance.

As the distance between the transmitter and receiver increases, the signal power decreases (all else being equal). At some point the received signal power will be less than the noise that is received/generated by the receiver, and reliable communications will not be possible. The greatest contribution to the decrease in signal power received is the loss due to the propagation distance known as the path loss or free space loss. The free space loss, L_s, is given by

$$L_s = (4\pi d)^2/\lambda^2$$

where

d is the distance in meters (or whatever unit is used for λ)
λ is the wavelength in meters

The wavelength term is not part of the physical path loss but is related to the receiving antenna's aperture (and related gain) and is included with the physical path loss.

The propagation distance from the ground to a geostationary satellite is around 36,000 km (the distance varies depending on the latitude and longitude of the ground station and the location of the satellite in the geostationary arc). At a frequency of 6 GHz, the free space loss is approximately 1×10^{20}. Because this is a loss, this means that the signal will be 10^{-20} times weaker when the path loss is accounted for.

Although it is usually the largest factor by far, the path loss is not the only contribution to the received power; there are many factors that can boost or decrease the power of the transmitted wave or that can introduce noise to the system. The ability to communicate depends on being able to receive the desired signal at a level above the received and generated noise. Factors that decrease signal (loss) or increase noise are detrimental to reliable communications, whereas factors that increase signal (gain) or decrease noise are beneficial.

The system *link budget* is used to account for the various gains and losses across a communications link between two nodes. Part of the link budget is calculated according to a conceptually simple equation accounting for gains and losses:

$$T_p * G_{link}/L_{link} = R_p$$

where

T_p is the transmitted power
G_{link} is the product of all the gains in the link
L_{link} is the product of all the losses in the link
R_p is the received power

Once we have the received power, we can calculate the signal-to-noise ratio (SNR):

$$\mathrm{SNR} = \mathrm{S}/\mathrm{N}$$

where, in the simplest case, the received signal power is $S = R_P$ and N is the received noise (including noise generated by the receiver). The signal-to-noise ratio is the key parameter in determining whether or not a communications link will be successful (or in the vernacular, whether it is possible to "close the link"). The higher the SNR, the better the link. The signal is modulated onto a radiofrequency (RF) wave known as a *carrier*. Sometimes the carrier-to-noise ratio, C/N or CNR, is used in satellite system design to characterize a link at the RF level before demodulation, whereas the SNR is a measure of the signal strength after demodulation. An RF carrier will usually carry many baseband signals. In a digital system, the signal-to-noise quantity we are looking for is the ratio of the energy per bit to noise density, E_b/N_0.

In a satellite relay system, there are two links that need to be analyzed: (1) the uplink, or the link from the ground to the satellite, and (2) the downlink or the link from the satellite to a second ground station. The uplink and the downlink are analyzed separately and the results combined to get the overall link budget.

The key result from a link budget is a quantity known as the link margin. This is the difference between the actual received power and the required received power in decibels (see below for a discussion of decibels). Hopefully the link margin is a positive quantity; a negative quantity indicates that the link will not be closed, and communications will not be successful. Generally, the higher the link budget, the better it is from an engineering standpoint (too high and a cost accountant somewhere should be complaining).

Understanding Decibels

The range in values of the various factors involved in calculating the link budget is very large. Engineers use a dimensionless unit, the *decibel* or dB (named after Alexander Graham Bell), to simplify the calculations. A result in decibels, R, is defined as a ratio of two powers, P_1 and P_2, such that

$$\mathrm{R\ dB} = 10\log_{10}(\mathrm{P}_1/\mathrm{P}_2)$$

or ten times the logarithm (to the base 10) of the ratio of the two power quantities. (The logarithm of a number is the exponent to which the base is raised to get that number.) Some examples of the conversion to decibels:

Let the ratio $P_1/P_2 = 10$ (i.e., $P_1 = 10 \times P_2$). Then

$$R = 10\log_{10}(10) = 10 \text{ dB}$$

since $\log_{10}(10) = 1$. Admittedly, this example is not a very exciting transformation. Let us try some more.

Let the ratio $P_1/P_2 = 1$ (i.e., $P_1 = P_2$). Then

$$R = 10\log_{10}(10) = 10 \text{ dB}$$

$$\text{since } \log_{10}(1) = 0.$$

Let the ratio $P_1/P_2 = 1/10$ (that is $10 \times P_1 = P_2$). Then

$$R = 10\log_{10}(1/10) = -10 \text{ dB}$$

since $\log_{10}\left(10^{-1}\right) = -1$. That is still not very impressive. Let us look at something a little bigger.

Let the ratio $P_1/P_2 = 1{,}000{,}000$ (that is $P_1 = 1{,}000{,}000\, P_2$). Then

$$R = 10\log_{10}(1{,}000{,}000) = 60 \text{ dB}$$

since $\log_{10}\left(10^{6}\right) = 6$. Ah, now we are getting somewhere. Looking at the result (60 dB), we can see that there should be six zeros in the original ratio because of the logarithmic nature of the dB. Table 3 shows that working with decibels makes it easier to work with power ratios having large dynamic ranges.

Dealing with large ranges of ratios cannot be the only reason for using decibels since scientific notation also makes it easier to work with large ranges of numbers

Table 3 Decibels provide compact representations of large dynamic ranges

Ratio	Exponent	Decibels
1	10E0 or 10^0	0 dB
10	10E1	10 dB
100	10E2	20 dB
1,000	10E3	30 dB
1,000,000	10E6	60 dB
1,000,000,000	10E9	90 dB
1,000,000,000,000	10E12	120 dB
100,000,000,000,000,000,000	10E20	200 dB

and does not require calculations. Another reason for using decibels is that the equivalent function to multiplying ratios is adding decibels and the equivalent function to dividing ratios is subtracting decibels. This is because of the logarithm in the definition. That means that instead of multiplying gains and dividing losses to get a link budget, we can convert to decibels and add gains and subtract losses. For example:

$$1/10 = 0\ \text{dB} - 10\ \text{dB} = -10\ \text{dB}$$

as we saw before. We will come back to link budgets, but first a few more examples.

Let the ratio $P_1/P_2 = 2$ (i.e., $P_1 = 2P_2$). Then

$$\text{R} = 10\log_{10}(2) = 3.0\ 1\ \text{dB}$$

since $\log_{10}(2) = 0.30102999\ldots$. Most engineers round off and use 3 dB to mean a ratio of 2. The power ratio equivalent of 3 dB is calculated:

$$\begin{gathered} 10\text{E}\ (\text{R}/10) = \text{P}_1/\text{P}_2 \\ 10^{(3/10)} = 1.99526\ldots \end{gathered}$$

which is approximately 2 as we would expect. Some other useful values to remember are given in Table 4.

Knowing the conversions for 2, 3, 7, and 10, we can get the rest from easy manipulation of the decibel values (the examples below use various levels of rounding off):

$$\begin{aligned} 4 &= 2 \times 2 = 3\ \text{dB} + 3\ \text{dB} = 6\ \text{dB} \\ 5 &= 10/2 = 10\ \text{dB} - 3\ \text{dB} = 7\ \text{dB} \\ 6 &= 2 \times 3 = 3.0\ \text{dB} + 4.8\ \text{dB} = 7.8\ \text{dB} \\ 8 &= 2 \times 4 = 3\ \text{dB} + 6\ \text{dB} = 9\ \text{dB} \\ 9 &= 3 \times 3 = 4.77\ \text{dB} + 4.77\ \text{dB} = 9.54\ \text{dB} \end{aligned}$$

Table 4 Converting numbers from 1 to 10 to decibels (rounded to two decimal places)

Ratio	dB
1	0
2	3.01
3	4.77
4	6.02
5	6.99
6	7.78
7	8.45
8	9.03
9	9.54
10	10.00

Table 5 Decibel values from 0 to 10 and their corresponding ratio values to two decimal places

Decibels	Ratio
0	1
1	1.26
2	1.58
3	1.99
4	2.51
5	3.16
6	3.98
7	5.01
8	6.31
9	7.94
10	10

With these values, it is easy to put together whatever number we need in decibels. For example:

$$2,000,000 = 2 \times 1,000,000 = 3 \text{ dB} + 60 \text{ dB} = 63 \text{ dB}$$
$$400 = 4 \times 100 = 6 \text{ dB} + 20 \text{ dB} = 26 \text{ dB}$$

Table 5 shows the case for the inverse conversion from decibels to power ratios rounded off to two places. As we have already seen, 3 dB is typically rounded off to a value of 2, which would have happened if we rounded these values off to one decimal place.

To convert, say, 26 dB to a ratio, one could use the definition and calculate:

$$10\text{E}(26/10) = 398$$

or use some simple arithmetic and some memorized conversions (i.e., that 6 dB is approximately a ratio of 4):

$$26 \text{ dB} = 20 \text{ dB} + 6 \text{ dB} = 100 \times 4 = 400$$

which is hopefully close enough for the desired application.

Up to this point, we have been talking about converting power ratios to decibels because that is the way decibels are defined. Of course, if one needs to convert a number, the special case where $P_2 = 1$ can be used, so that the ratio P_1/P_2 becomes P_1. At the risk of confusing the reader, there are cases where dB are used for amplitude ratios where the definition is $20 \log_{10}(A_1/A_2)$. This definition is not used in the link budget discussion below but is provided for completeness in the decibel discussion. The amplitude definition of decibels, dBV (V for volts in this case), can be encountered by replacing the power quantities in the power ratio by voltages:

$$P = V^2/Z$$

where P = power, V = voltage, and Z = impedance, so that

$$\begin{aligned} R &= 10\log 10(P_1/P_2) \\ &= 10\log_{10}\left(\left(V_1{}^2/Z\right)/\left(V_2{}^2/Z\right)\right) \\ &= 10\log_{10}(V_1/V_2)^2 \\ &= 20\log_{10}(V_1/V_2) \end{aligned}$$

because $\log_{10}X^2 = 2\log_{10}X$.

The decibel is not an official unit of the International System of Units (SI); however, it is approved by the International Committee for Weights and Measures (CIPM, *Comité International des Poids et Mesures*) for use with the SI. A common practice, which is not in accordance with official SI rules, is to attach a reference unit indicator to the dB symbol, such as dBW, dBm, dBHz, etc., as a shorthand way of specifying a reference level (Ambler and Taylor 2008). These units indicate a reference value in order to use the decibel definition in a specialized way as an absolute unit rather than a ratio (relative) unit. For example, dBW indicates a measurement relative to a value of 1 W (or $P_2 = 1\text{ W}$). Thus

$$R\text{ dBW} = 10\log_{10}(P_1/1\text{W})$$

so that, if $P_1 = 1\text{ W}$, then $R = 0\text{ dBW}$. In SI units, the reference value of 1 W would have to be specified explicitly each time, i.e., R (re 1 W) = 0 dB or $R_{(1\text{ W})} = 0\text{ dB}$. Common reference values that may be encountered in non-SI literature are shown in Table 6.

Link Budget Calculation

The simplified power balance equation (see above)

$$T_p * G_{link}/L_{link} = R_p$$

using decibels becomes

$$T_p\text{dB} + G_{link}\text{dB} - L_{link}\text{dB} = R_p\text{dB}$$

Table 6 Informal decibel reference indications

Informal reference unit	Reference value
dBW	1 W
dBm	0.001 W or 1 mW
dBi	Isotropic antenna gain
dBHz	1 Hz

where

T_p dB is the transmitted power in decibels (referenced to W or mW)
G_{link}dB is the sum of all the gains in the link in decibels
L_{link}dB is the sum of all the losses in the link in decibels
R_pdB is the received power in decibels (referenced to W or mW as in T_p)

Although this equation is very simple, the challenge is in finding all of the significant gains and losses and keeping their signs correct. The definition of a loss, for example, could be as a positive quantity to be subtracted or as a negative quantity to be added. For example, the free space loss to GEO at 6 GHz (which we saw above is 10^{20}) is 200 dB which is subtracted in the link budget equation.

The transmitted power and transmitter antenna gain are usually combined in a quantity called the effective isotropic radiated power (EIRP). As the name implies, EIRP is the power that would have to be radiated from an isotropic antenna (an antenna that radiates equally in all directions) to provide the same power flux density at the receiver:

$$\mathrm{EIRP} = P_T - L_T + G_T \mathrm{dBW}$$

where

P_T = transmitter power
L_T = transmitter losses
G_T = transmit antenna gain

The important parameter to the receiver is the received power. Different combinations of transmitter hardware can provide the same received power, and EIRP is a way of representing the transmitted power in a standard way. For example, if we ignore the losses, an EIRP of 50 dBW could be obtained by any number of combinations of transmit power and antenna gain (e.g., $P_T = 15$ dBW and $G_T = 35$ dB, $P_T = 20$ dBW and $G_T = 30$ dB, etc.). Higher EIRP results in higher receiver power, all else being equal, but there are regulatory limits on the amount of power that can be radiated onto the Earth. These regulatory limits are for the purpose of limiting interference to other communications systems.

Another concept, G/T, is used to characterize the receiver. The receiver figure of merit is the ratio of the receiving antenna gain, G, to the receiver system noise temperature, T (in Kelvins). The higher the G/T, the better the receiver. The noise temperature is a way of lumping together various noise contributions, many of which are thermal in origin and the remainder of which are modeled as thermal. The received noise power is related to the noise temperature by Boltzmann's constant and the bandwidth of interest.

The simplified link budget equation (Miller et al. 1993) can be written using EIRP and G/T:

$$C/N = \mathrm{EIRP} - L_s + 10\log_{10}(G/T) - 10\log_{10}(kB)$$

where k is Boltzmann's constant and B is the equivalent noise bandwidth of the receiver.

For the purposes of this chapter, it is sufficient to realize that there are many quantities that need to be determined for entry into a link budget calculator, some of which are combinations of lower-level details of the system design. The level of detail in the link budget depends on its purpose. As a design tool, the system engineer will have a link budget that may have dozens of separate items that allow trade-offs to be made. Once a system is built and losses become more determined and tightly range bound, some quantities can be combined into a simplified link budget for use in, say, determining a particular ground station design.

A detailed design link budget may include gains and losses due to items such as circuits (may include cables, connectors, waveguides, switches, etc.), antenna size, antenna polarization, antenna alignment, weather, atmospheric gasses, thermal noise, active device noise, interference, back off or attenuation, coding, etc.

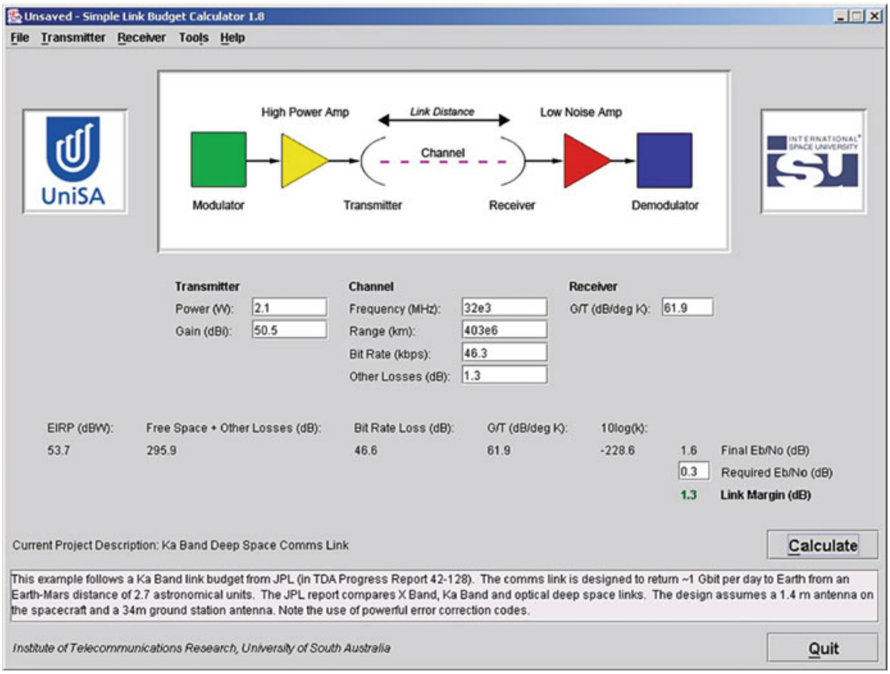
There are many free and commercial automated tools for calculating radiofrequency link budgets and especially link budgets for satellite systems. Many link budget tools for educational purposes require the user to enter all parameters but may provide a limited number of variable parameters by lumping detailed parameters together. This provides a nice platform for getting the feel of a link budget without getting bogged down in the details of a design (Fig. 7).

A system engineer doing a detailed design will probably use a spreadsheet to build their own tool so that they can introduce parameters as needed. Some of the parameters in the link budget are related to the satellite design (e.g., transponder bandwidth, transmitted power, antenna gain), and some are related to the ground station design (e.g., terminal location, antenna gain, receiver sensitivity). To add services with an existing satellite, the satellite design parameters are already determined since it is already deployed in space and can be obtained from the satellite service provider or from a database.

Commercial tools allow potential users to investigate system options by including a database of existing satellite and ground systems with parameters needed to conduct link budget analyses in addition to a tool for doing the calculations.

Some software packages provide a useful representation of a satellite transponder's coverage called a footprint map (see Fig. 8). A footprint is the intersection of a satellite's beam with the surface of the Earth. The contours of the map show the areas of signal strength due to the antenna pattern. Note in Fig. 4 that the designer of the spacecraft antenna shaped the antenna coverage contour to evenly spread the maximum EIRP output over the US northern continental area and the populous areas of Canada. Presumably, this is the area of the primary intended customer base.

Table 7 shows sample output from a commercial software package called SatFinder. This table indicates the considerable complexity found in an actual link



Transmitter

Power (W): 2.1

Gain (dBi): 50.5

Channel

Frequency (MHz): 32e3

Range (km): 403e6

Bit Rate (kbps): 46.3

Other Losses (dB): 1.3

Receiver

G/T (dB/deg K): 61.9

EIRP (dBW):	Free Space + Other Losses (dB):	Bit Rate Loss (dB):	G/T (dB/deg K):	10log(k):		
53.7	295.9	46.6	61.9	-228.6	1.6	Final Eb/No (dB)
					0.3	Required Eb/No (dB)
					1.3	**Link Margin (dB)**

Current Project Description: Ka Band Deep Space Comms Link

This example follows a Ka Band link budget from JPL (in TDA Progress Report 42-128). The comms link is designed to return ~1 Gbit per day to Earth from an Earth-Mars distance of 2.7 astronomical units. The JPL report compares X Band, Ka Band and optical deep space links. The design assumes a 1.4 m antenna on the spacecraft and a 34m ground station antenna. Note the use of powerful error correction codes.

Institute of Telecommunications Research, University of South Australia

Fig. 7 Example of a free tool for demonstrating link budget concepts (Courtesy UniSA, from http://www.itr.unisa.edu.au/itrusers/bill/public_html/software/)

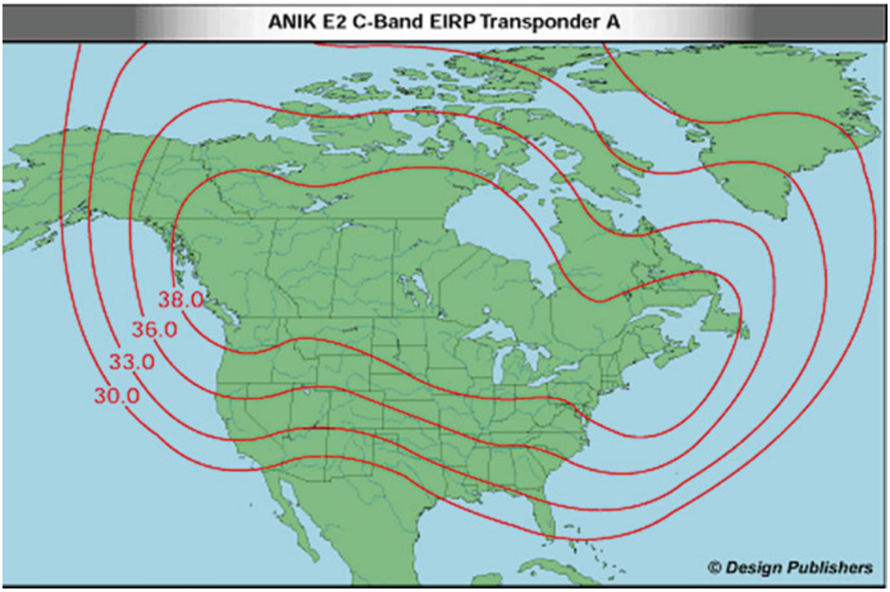


Fig. 8 Example footprint map from SatFinder (Courtesy SatNews.com)

Table 7 Example output from SatFinder link budget software

Digital link budget produced using SatFinder (http://www.satnews.com/linksample.htm)				
Service name	Unspecified			
Coverage	Unspecified			
Uplink earth station	Denver			
Downlink earth station	Los Angeles			
Satellite name	Telstar 402R			
Link input parameters	*Up*	*Down*	*Units*	
Site latitude	39.73 N	34.05 N	degrees	
Site longitude	105.00 W	118.25 W	degrees	
Site altitude	1	0.1	km	
Frequency	14.472	12.172	GHz	
Polarization	Vertical	Vertical	–	
Rain model	ITU (30.3)	ITU (19.7)	(mm/h or zone)	
Availability (average year)	99.9	99.9	%	
Water vapor density	3	10	gm/m^3	
Surface temperature	10	20	°C	
Antenna aperture	10	0.9	meters	
Antenna efficiency/gain	65	70	% (+ prefix dBi)	
Coupling loss	0.2	0	dB	
Antenna tracking/mispoint error	0	0.3	dB	
LNB noise figure/temp	–	0.75	dB (+ prefix K)	
Antenna noise	–	27	K	
Adjacent carrier interference	30	30	dB	
Adjacent satellite interference	30	12	dB	
Cross polarization interference	200	200	dB	
Uplink station HPA output back off	3	–	dB	
Number of carriers/HPA	1	–	–	
HPA C/IM (up)	200	–	dB	
Uplink power control	0	–	dB	
Uplink filter truncation loss	0	–	dB	
Required HPA power capability	MAX	–	W	
Satellite input parameters	*Value*	*Units*		
Satellite longitude	89.00 W	degrees		
Transponder type	TWTA	–		
Receive G/T	1	dB/K		
Saturation flux density	−86.05	dBW/m^2		
Satellite attenuator pad	0	dB		
Satellite ALC	0	dB		
EIRP (saturation)	47	dBW		
Transponder bandwidth	27	MHz		

(*continued*)

Table 7 (continued)

Digital link budget produced using SatFinder (http://www.satnews.com/linksample.htm)				
Input back off total	1	dB		
Output back off total	0.3	dB		
Intermodulation interference	30	dB		
Number of transponder carriers	AUTO	–		
Carrier/link input parameters	*Value*	*Units*		
Modulation	4-PSK	–		
Required bit error rate performance	10^{-9}	–		
Required Eb/No without FEC coding	20	dB		
Required Eb/No with FEC coding	5.1	dB		
Information rate	23.6	Mbps		
Overhead	0	%		
FEC code rate	0.59	–		
Spreading gain	0	dB		
Reed–Solomon code	1	–		
(1 + Roll-off factor)	1.2	–		
Carrier spacing factor	1.3	–		
Bandwidth allocation step size	0.1	MHz		
System margin	0.9	dB		
Calculations at saturation	*Value*	*Units*		
Gain 1 m^2	44.67	dB/m^2		
Uplink C/No	98.88	dBHz		
Downlink C/No	89.72	dBHz		
Total C/No	79.59	dBHz		
Uplink EIRP for saturation	76.65	dBW		
General calculations	*Up*	*Down*	*Units*	
Elevation	41.14	39.61	degrees	
True azimuth	155.84	134.99	degrees	
Compass bearing	145.77	121.46	degrees	
Path distance to satellite	37692.47	37809.06	km	
Propagation time delay	0.125728	0.126117	seconds	
Antenna efficiency	65	70	%	
Antenna gain	61.75	39.65	dBi	
Availability (average year)	99.9	99.9	%	
Link downtime (average year)	8.766	8.766	hours	
Availability (worst month)	99.615	99.615	%	
Link downtime (worst month)	2.809	2.809	hours	
Spectral power density	−59.11	−26.31	dBW/Hz	
Uplink calculation	*Clear*	*Rain up*	*Rain dn*	*Units*
Uplink transmit EIRP	75.65	75.65	75.65	dBW
Transponder input back off (total)	1	1	1	dB
Input back off per carrier	1	4.32	1	dB
Mispoint loss	0	0	0	dB
Free space loss	207.18	207.18	207.18	dB

(*continued*)

Table 7 (continued)

Digital link budget produced using SatFinder (http://www.satnews.com/linksample.htm)				
Atmospheric absorption	0.1	0.1	0.1	dB
Tropospheric scintillation fading	0.08	0.08	0.08	dB
Atmospheric losses total	0.18	0.18	0.18	dB
Total path loss (excluding rain)	207.36	207.36	207.36	dB
Rain attenuation	0	3.32	0	dB
Uplink power control	0	0	0	dB
Uncompensated rain fade	0	3.32	0	dB
C/No (thermal)	97.88	94.56	97.88	dB. Hz
C/N (thermal)	24.08	20.76	24.08	dB
C/ACI	30	26.68	30	dB
C/ASI	30	26.68	30	dB
C/XPI	200	196.68	200	dB
C/IM	200	200	200	dB
Eb/(No + Io)	13.26	13.26	13.26	dB
Downlink calculation	*Clear*	*Rain up*	*Rain dn*	*Units*
Satellite EIRP total	47	47	47	dBW
Transponder output back off (total)	0.3	0.3	0.3	dB
Output back off per carrier	0.3	3.62	0.3	dB
Satellite EIRP per carrier	46.7	43.38	46.7	dBW
Mispoint loss	0.3	0.3	0.3	dB
Free space loss	205.71	205.71	205.71	dB
Atmospheric absorption	0.1	0.1	0.1	dB
Tropospheric scintillation fading	0.3	0.3	0.3	dB
Atmospheric losses total	0.4	0.4	0.4	dB
Total path loss (excluding rain)	206.11	206.11	206.11	dB
Rain attenuation	0	0	2.74	dB
Noise increase due to precipitation	0	0	4.14	dB
Downlink degradation (DND)	0	0	6.89	dB
Total system noise	81.67	81.67	212.02	K
Figure of merit (G/T)	20.23	20.23	16.09	dB/K
C/No (thermal)	89.42	86.1	82.53	dB. Hz
C/N (thermal)	15.61	12.29	8.73	dB
C/ACI	30	26.68	30	dB
C/ASI	12	8.68	12	dB
C/XPI	200	196.68	200	dB
C/IM	30	26.68	30	dB
Eb/(No + Io)	−4.15	−4.15	−4.15	dB
Totals per carrier (end-to-end)	*Clear*	*Rain up*	*Rain dn*	*Units*
C/No (thermal)	79.59	79.59	79.59	dB. Hz
C/N (thermal)	5.78	5.78	5.78	dB

(*continued*)

Table 7 (continued)

Digital link budget produced using SatFinder (http://www.satnews.com/linksample.htm)				
C/ACI	23.19	23.19	23.19	dB
C/ASI	6.15	6.15	6.15	dB
C/XPI	193.19	193.19	193.19	dB
C/IM	26.2	26.2	26.2	dB
C/(No + Io)	59.96	59.96	59.96	dB.Hz
C/(N + I)	−13.85	−13.85	−13.85	dB
Eb/(No + Io)	−13.77	−13.77	−13.77	dB
System margin	0.9	0.9	0.9	dB
Net Eb/(No + Io)	−14.67	−14.67	−14.67	dB
Required Eb/(No + Io)	5.1	5.1	5.1	dB
Excess margin	−19.77	−19.77	−19.77	dB
Earth station power requirements	*Value*	*Units*		
EIRP per carrier	75.65	dBW		
Antenna gain	61.75	dBi		
Antenna feed flange power per carrier	13.9	dBW		
Uplink power control	0	dB		
HPA output back off	3	dB		
Waveguide loss	0.2	dB		
Filter truncation loss	0	dB		
Number of HPA carriers	1	–		
Total HPA power required	17.1016	dBW		
Required HPA power capability	10	W		
Spectral power density	−59.11	dBW/Hz		
Space segment utilization	*Value*	*Units*		
Overall link availability	99.8	%		
Information rate (inc overhead)	23.6	Mbps		
Transmit rate	40	Mbps		
Symbol rate	20	Mbaud		
Occupied bandwidth	24	MHz		
Noise bandwidth	73.8	dB.Hz		
Minimum allocated bandwidth required	26	MHz		
Allocated transponder bandwidth	26.1	MHz		
Percentage transponder bandwidth used	96.67	%		
Used transponder power	46.7	dBW		
Percentage transponder power used	100	%		
Max carriers by transponder bandwidth	1.03	–		
Max carriers by transponder power	1	–		
Max transponder carriers limited by	Power	[1.00]		
Power equivalent bandwidth usage	27	MHz		

budget calculation for a particular satellite. SatFinder has a database with information on over 500 satellites (including orbital location, frequencies, EIRP, G/T, bandwidth, etc.). Similar satellite link budget tools include products that can be found online such as Satmaster or Customizable Link Budget Tool.

Some commercial network modeling tools, such as OpNet, have satellite link modeling capability and can integrate with other tools, such as Analytical Graphics, Inc.'s Satellite Tool Kit (STK), to provide analysis of satellite systems. STK has several nice communications functions and tutorials that can help in understanding link budgets and system trade-offs. There are tools from other companies that are specialized for regulatory coordination to avoid interference between systems. Examples include Visualyse (from Transfinite) and Sat-Coord (from RPC Telecommunications).

The main result from a link budget analysis is the link margin. This is the amount of received power over and above the minimum amount required to close the link and meet the system requirements for quality. A good value for a link margin in a fixed satellite system (where the ground stations are well characterized and design parameters do not change radically) is around 3 dB. For high-frequency systems (Ka band), rain fade is an important design consideration. From Fig. 2 we can see that rain fade may require 10 dB or more of margin, but there are techniques for dealing with rain including site diversity and adaptive power control.

For a mobile satellite system, at least one ground terminal is a mobile device, perhaps handheld. In that case, the design parameters may be more variable depending on how and where the handset is being used. Handheld devices generally need to be low power (to minimize radiation of the user) and have low-gain antennas (to avoid having to point the antenna at the satellite). In an urban environment, there is shadowing from buildings as well as reflections that cause multipath distortions. Mobile systems may make use of LEO constellations or a GEO satellite. In a LEO system, there may be a need for some margin for handoffs between satellites. In general, a link margin of 10–15 dB might be appropriate for mobile systems. However, one must examine the link budget to see where various effects are accounted for. Assumptions and confidence in models used to arrive at parameters used in the link budget calculation (such as rain fade) will affect the amount of link margin required.

Adaptive power control can be used to vary the transmitted power to overcome short-term increases of the link path loss (known as fading). When conditions are good, the transmitted power can be reduced, and when conditions deteriorate, transmitted power levels can be increased. Some link margin is still required to accommodate dynamic changes in losses that occur faster than the adaptive feedback can respond.

Availability and signal quality are two requirements that will determine an acceptable link margin. Availability refers to the percentage of time that a satellite link can be closed during a certain time period (a year is typically used). An availability of 99.95 % implies that a link is down for a total of around 4 h per year. Signal quality depends on the type of signal being transmitted. For digital signals, the signal quality parameter of interest is the bit error rate. For other signals it might be the SNR or C/N.

Conclusion

Link budgets are an accounting of the gains and losses in a communications system. The link margin is a measure of the robustness of the link. The channel capacity equation implies that a communications engineer can trade off power, bandwidth, and complexity to achieve various system requirements in a design. Establishing link budget and link margins for different types of satellite systems represents a variety of challenges. As one moves to higher frequencies and one moves from fixed to mobile satellite systems, the more difficult the challenge becomes and the higher the link margin must be set to provide reliable service.

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Global Communications Satellite Systems
- ▶ US Domestic Communications Satellite Systems
- ▶ Common Elements Versus Unique Requirements in Various Types of Satellite Application Systems
- ▶ Economics and Financing of Communications Satellites
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ Geographic Information Systems and Geomatics
- ▶ History of Satellite Communications
- ▶ Introduction to Satellite Navigation Systems
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Overview of the Spacecraft Bus
- ▶ Regulatory Process for Communications Satellite Frequency Allocations
- ▶ Satellite Antenna Systems Design and Implementation Around the World
- ▶ Satellite Communications and Space Telecommunication Frequencies
- ▶ Satellite Communications Antenna Concepts and Engineering
- ▶ Satellite Communications Modulation and Multiplexing
- ▶ Satellite Communications Overview
- ▶ Satellite Communications: Regulatory, Legal, and Trade Issues
- ▶ Satellite Communications Video Markets: Dynamics and Trends
- ▶ Satellite Earth Station Antenna Systems and System Design
- ▶ Satellite Orbits for Communications Satellites
- ▶ Satellite Transmission, Reception, and Onboard Processing, Signaling, and Switching
- ▶ Space Telecommunications Services and Applications
- ▶ Telemetry, Tracking, and Command (TT&C)
- ▶ Trends and Future of Satellite Communications

References

T. Ambler, B.N. Taylor, *Guide for the Use of the International System of Units (SI)*. NIST Special Publication 811, 2008th edn. (U.S. Department of Commerce, Technology Administration, National Institute of Standards and Technology, Gaithersburg, 2008), p. 30

V.K. Bhargava, D. Haccoun, R. Matyas, P.P. Nuspl, *Digital Communications by Satellite* (Wiley, New York, 1981), p. 24

R.C. Dixon, *Spread Spectrum Systems* (Wiley, New York, 1976), p. 13

R.M. Gagliardi, *Satellite Communications* (Van Nostrand Reinhold, New York, 1984), pp. 17–20

R.W. Hamming, *Coding and Information Theory*, 2nd edn. (Prentice-Hall, Englewood Cliffs, 1986), p. 2

S. Hayken, *Communications Systems*, 2nd edn. (Wiley, Chichester, 1983), p. 469

N.S. Jayant, P. Noll, *Digital Coding of Waveforms* (Prentice-Hall, Englewood Cliffs, 1984), p. 221

M.J. Miller, B. Vucetic, L. Berry (eds.), *Satellite Communications Mobile and Fixed Services* (Kluwer, Boston, 1993), p. 51

W.W. Peterson, E.J. Weldon Jr., *Error-Correcting Codes*, 2nd edn. (MIT Press, Cambridge, 1991), p. 6

Satellite Communications Modulation and Multiplexing

Paul T. Thompson

Contents

P.T. Thompson (✉)
University of Surrey, Guildford, Surrey, UK
e-mail: p.thompson@surrey.ac.uk

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_86

Abstract

This chapter addresses the principles involved in three key areas in satellite communications, namely, modulation (and demodulation), forward error coding, and multiple access approaches or techniques. It focuses on features and technical approaches utilized in modern-day digital satellite communications systems. Since analog approaches or techniques are today seldom used in operating satellite networks, they are not addressed in detail. Nevertheless, several references are given concerning such analog techniques should there be a historical interest in these subjects. The materials provided in this chapter are aimed at giving an appreciation of the issues involved and the performance achievable in practice. For more detailed mathematical treatments, an extensive bibliography is also provided.

Keywords

ALOHA random access • Bit error rate (BER) • Bit rate • Block codes • Carrier to noise ratio • CDMA • Coding • Coding gain • Convolutional coding • Decoding • Demand assigned multiple access (DAMA) • Demand assignment • Demodulator • Digital video broadcast (DVB) • Encoding • Forward error correction (FEC) • Low-density parity-check (LDPC) coding • Medium access control (MAC) • Modem • Modulation • Multiple access techniques • Nyquist filtering • Random access • Reed–Solomon coding • Claude Shannon • Signal constellation • Spectral efficiency • Symbol rate • TDMA • Turbo coding • Viterbi algorithm or Viterbi coding • VSAT networks

Introduction

When information is to be conveyed over a satellite link, it is first processed to make it suitable for impressing this information on a radio frequency (RF) carrier. This process is called modulation. After transmission over the satellite link, the modulated RF carrier is demodulated to extract a replica of the original information. Various modulation formats are adopted in satellite communications. The choice of the modulation technique employed depends upon a range of considerations such as:

- Nature of the information to be conveyed (long streams, short bursts, etc.)
- Performance requirements
- Interference generation and tolerance
- Capability of the satellite link in terms of power and bandwidth available

To operate satellite communications systems to obtain the maximum value from the satellite resources available (power and bandwidth), it is important to exploit these resources in an efficient manner within the constraints of the system parameters. To this end, if bandwidth permits, forward error correction (FEC) or coding can be employed to reduce the power requirements and therefore achieve higher capacity overall. A range of FEC approaches exist that match the specific needs of the traffic and satellite channel. These have evolved with time as the hardware developed to allow the implementation of more complex schemes. A range of these will be presented along with performance metrics and associated performance outcomes.

The third key feature addressed in this chapter is that of multiple access, that is, how terminals in a network communicate to each other in terms of the number of carriers that are available for utilization, their contents, and whom they are intended to reach. Again, a range of approaches exist depending upon the nature of the traffic, the size of the network, and the ground terminals involved.

Modulation

Modulation is the process of impressing the wanted data on a radio frequency (RF) carrier which is then conveyed over the satellite link and demodulated at the receiving terminal. Thus, modulation translates a baseband spectrum (in a lower frequency range) to a carrier spectrum (at a much higher frequency range).

Demodulation is the process of recovering the data at the receiver end of the link. Figure 1 depicts a typical transmit and receive chain with the modulation and demodulation clearly identified.

Key Aspects of Modulation

Modulation is the name given to the process of impressing the signal to be transported onto a carrier. Demodulation is the name given to extracting the wanted signal from the carrier. Thus, the process requires a modulator and a demodulator, collectively known as a modem. The input to the modulator may require some initial processing such as filtering and amplitude limiting. Digital input signals may also be processed to provide a balance signal around zero and to spread the bandwidth of the signal (with a scrambler).

Modulation of a signal onto a carrier results in dynamic changes to one or more of the following:

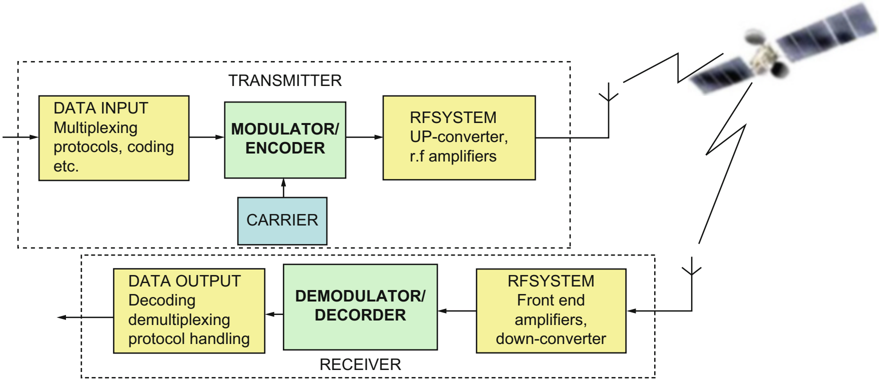


Fig. 1 Typical transmit and receive chain

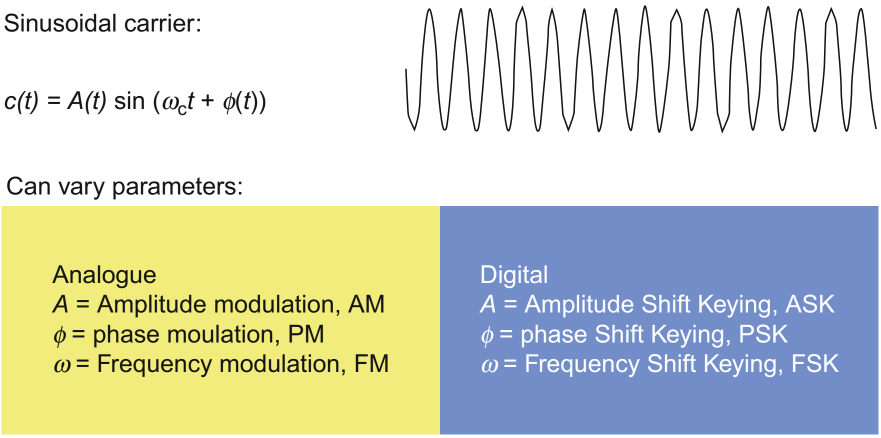


Fig. 2 Modulation of a sinusoidal carrier

- The amplitude of the carrier
- The phase of the carrier
- The frequency of the carrier

Thus, a sinusoidal carrier can be modulated in several ways as shown in Fig. 2.

The initial carrier is usually a CW signal with no perceptible bandwidth. The modulated carrier will have an associated bandwidth that is a factor of the bandwidth of the data signal. Two bandwidths are often considered in practice, namely, the occupied bandwidth and the allocated bandwidth. The occupied bandwidth is associated with the noise bandwidth in which the signal is being detected and demodulated. It is often expressed as the −10 dB bandwidth. The allocated bandwidth is that bandwidth which is allocated by the satellite operator and includes some extra bandwidth to avoid excessive adjacent channel interference. This bandwidth is often set at 1.4 times the symbol rate but can vary based on a number of transmission efficiency factors.

In this chapter, we address the digital modulation cases.

Digital Modulation

The modulation of a carrier with a digital signal may take one of many forms as indicated below.

- Amplitude-shift keying (ASK)
- Frequency-shift keying (FSK)
- Phase-shift keying (PSK)
- Bipolar phase-shift keying (BPSK)
- Quadrature phase-shift keying (QPSK)
- M-ary phase-shift keying (MPSK)
- Asymmetric phase-shift keying (APSK)
- Minimum-shift keying (MSK)
- Quadrature amplitude modulation (QAM)

By far the most popular is QPSK which has been employed in satellite systems for many decades, while BPSK has been adopted for critical cases where interference issues are highly important. There are systems, for instance, that can drop from high-efficiency 8-phase-shift keying to QPSK or even BPSK modulation depending on the degree of interference or precipitation attenuation that might be experienced at the ground antenna system due to heavy rainfall – particularly when much higher microwave or millimeter-wave frequencies are being utilized.

A sinusoid of any phase can be regarded as a combination of a cosine wave in a in-phase component and a sine wave in a quadrature component which can be plotted on a phasor diagram as shown in Fig. 3. This so-called I/Q quadrature approach is very popular in digital modem implementations.

BPSK has just two phase states, one representing a digital zero and the other representing a digital one. This can be depicted on a phasor diagram and is known as a constellation diagram. The receiver has to determine a digital zero or one by examining the phase of the carrier. The decision boundary can be plotted on the constellation diagram (for BPSK, this is the “Q axis” as shown in Fig. 4).

With QPSK, four-phase states represent the data. These are established with symbols where a symbol in this case represents 2 bits, for example, “01.” The decision boundaries in this case are both the I and the Q axes as shown in Fig. 5. It is clear from this figure that there is more scope to misinterpret the symbols and hence QPSK will need more immunity from noise than BPSK. This is normally achieved by operating at a higher carrier to noise level (i.e., typically 3 dB higher).

For 8PSK, a symbol represents 3 bits as depicted in Fig. 6.

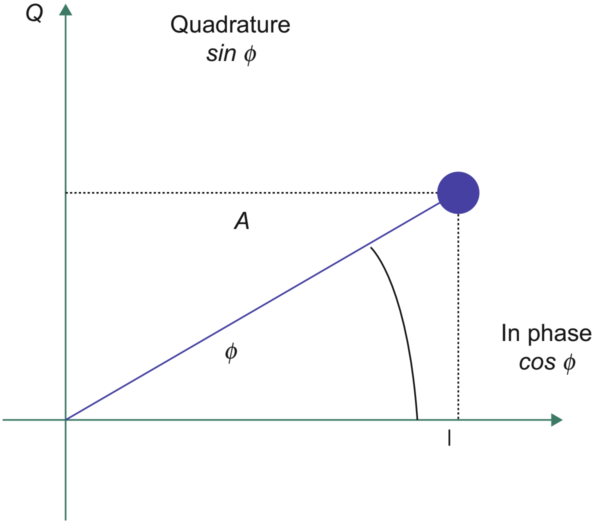


Fig. 3 Phasor representation of I and Q components

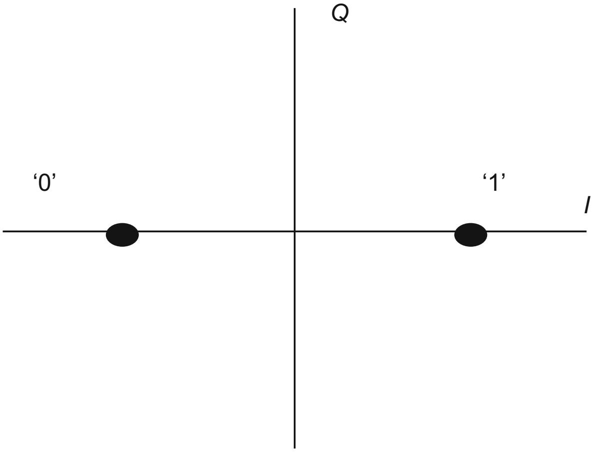


Fig. 4 BPSK constellation diagram

The DVB-S2 standard [1] has two special higher-order modulation formats that have been developed to operate in a more optimum manner in the presence of a nonlinear channel (caused by operating high-power amplifiers close to saturation). These are 16APSK and 32APSK. Their respective constellation diagrams are given in Fig. 7.

---

[1] Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications, ETSI EN 302 307 V1.2.1.

Fig. 5 QPSK constellation diagram

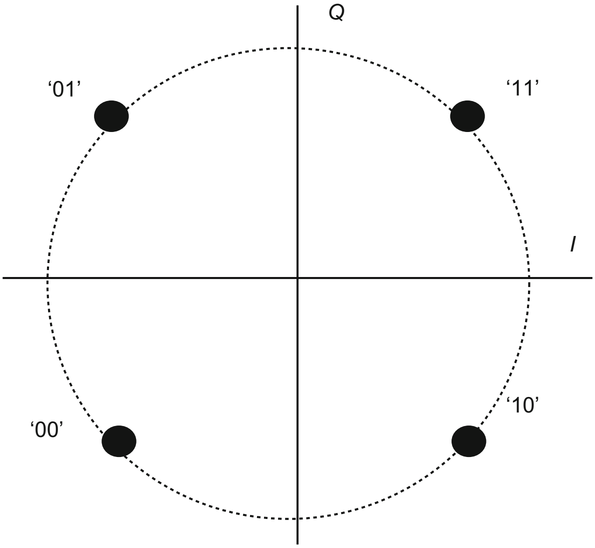
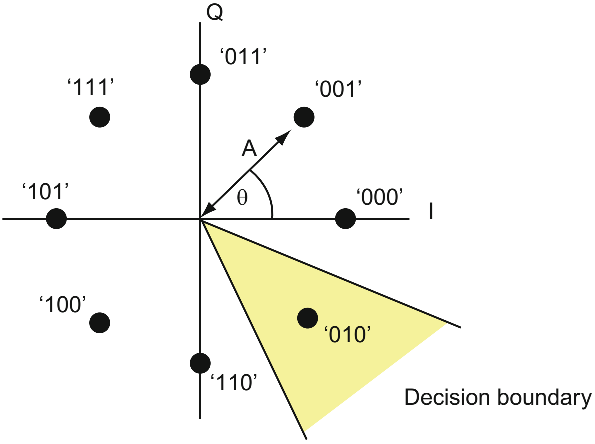


Fig. 6 8PSK constellation diagram



An alternative to APSK is quadrature amplitude modulation (QAM). Typically, QAM has a rectangular constellation diagram structure as depicted in Fig. 8 where 16-QAM is shown.

Consideration is now given to the issue of how to describe the performance of these different modulation formats and how to match them to communications needs and circumstances.

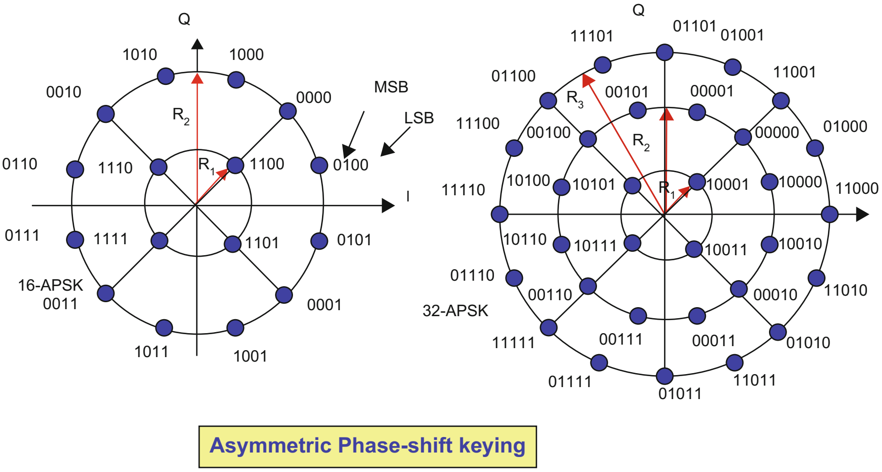


Fig. 7 Constellation diagrams for 16APSK and 32APSK

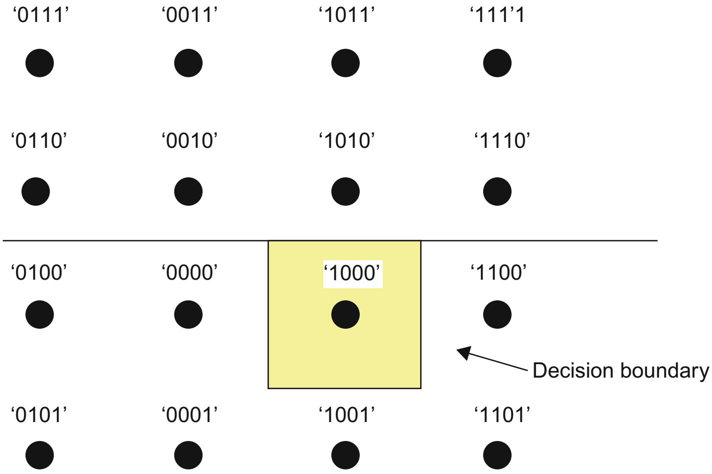


Fig. 8 16-QAM constellation diagram

Performance of Modulation Schemes

The design of a satellite link is based upon the level of performance required. The performance is usually described in terms of a bit error rate (BER) at which acceptable service is possible such as one error occurring in one million bits (often indicated as a BER probability of 1 in 10 $^{-6}$). The characteristics of the link are then matched to this BER in terms of the carrier to noise (C/N) level required to achieve this performance. An alternative to C/N is E_b/N_o, which is a measure of the energy per bit to noise power spectral density ratio (the noise in 1 Hz bandwidth). The two are related as indicated below.

$$C/N = E_b/N_o \,.\, \frac{f_b}{B}$$

where

f_b is the channel data rate and B is the channel bandwidth.

In logarithmic form (dB), the relationship is:

$$C/N_{\mathrm{dB}} = 10\log_{10}(E_b/N_o) + 10\log_{10}\left(\frac{f_b}{B}\right)\ \mathrm{dB}$$

If E_b/N_o is expressed in dB, then the above equation becomes:

$$C/N_{\mathrm{dB}} = E_b/N_{o\mathrm{dB}} + 10\log_{10}\left(\frac{f_b}{B}\right)\mathrm{dB}$$

The above analysis assumes that the noise environment of the satellite link can be described as Additive White Gaussian Noise (AWGN) that is a satellite link in which the only impairment to communication is noise with a constant spectral density and a Gaussian distribution of amplitude. This approach does not take account of fading, frequency selectivity, interference, nonlinearity, or dispersion. However, it produces an adequately simple model which is useful for gaining insight into the underlying behavior of a system.

E_b/N_o can be related to the energy per symbol per noise power spectral density (E_s/N_o) by:

$$E_b/N_o = E_s/N_o \cdot \frac{1}{\rho}$$

where E_s is the energy per symbol in joules and ρ is the nominal spectral efficiency. Spectral efficiency is defined as the number of bits per second that can be transmitted within 1 Hz. For a number of years, a typical spectral efficiency in satellite communications was 1 bit/s per Hz, but today, efficiencies are more typically in the range of 2.5 bits/s per Hz and as can be seen in Table 1 could range as high as 4.5 bits/s per Hz with 32APSK digital modulation schemes and with a very low level of noise and interference. E_s/N_0 is also commonly used in the analysis of digital modulation schemes.

They are related by:

$$E_s/N_o = E_b/N_o \,.\, \log_2(M)$$

where M is the number of modulation symbols or phase states.

The lower the operating E_b/N_o, the lesser the power that is required from the transmitters, or alternatively, the more noise or interference that can be tolerated. Figure 9 shows such an effect for typical uncoded modulation.

Table 1 DVB-S2 performance at the QEF threshold

DVB-S2 MOD	FEC rate	Spectral efficiency (bps/Hz)	Ideal E_s / N_o dB	IDEAL E_b / N_o dB	Including modem margin E_b/N_o dB	M
QPSK	1/4	0.490243	−2.35	0.75	1.95	4
QPSK	1/3	0.656448	−1.24	0.59	1.79	4
QPSK	2/5	0.789412	−0.30	0.73	1.93	4
QPSK	1/2	0.988858	1.00	1.05	2.25	4
QPSK	3/5	1.188304	2.23	1.48	2.68	4
QPSK	2/3	1.322253	3.10	1.89	3.09	4
QPSK	3/4	1.487473	4.03	2.31	3.51	4
QPSK	4/5	1.587196	4.68	2.67	3.87	4
QPSK	5/6	1.654663	5.18	2.99	4.19	4
QPSK	8/9	1.766451	6.20	3.73	4.93	4
QPSK	9/10	1.788612	6.42	3.89	5.09	4
8PSK	3/5	1.779991	5.50	3.00	4.20	8
8PSK	2/3	1.980636	6.62	3.65	4.85	8
8PSK	3/4	2.228124	7.91	4.43	5.63	8
8PSK	5/6	2.478562	9.35	5.41	6.61	8
8PSK	8/9	2.646012	10.69	6.46	7.66	8
8PSK	9/10	2.679207	10.98	6.70	7.90	8
16APSK	2/3	2.637201	8.97	4.76	5.96	16
16APSK	3/4	2.966728	10.21	5.49	6.69	16
16APSK	4/5	3.165623	11.03	6.03	7.23	16
16APSK	5/6	3.300184	11.61	6.42	7.62	16
16APSK	8/9	3.523143	12.89	7.42	8.62	16
16APSK	9/10	3.567342	13.13	7.61	8.81	16
32APSK	3/4	3.703295	12.73	7.04	8.24	32
32APSK	4/5	3.951571	13.64	7.67	8.87	32
32APSK	5/6	4.119540	14.28	8.13	9.33	32
32APSK	8/9	4.397854	15.69	9.26	10.46	32
32APSK	9/10	4.453027	16.05	9.56	10.76	32

Figure 10 indicates the performance of a range of modulations applicable to satellite links. The derivation of these curves is not presented here, but if further details are required, the reader is referred to chapter 7 of Pahlavan and Levesque (1995) and Proakis et al. (2004).

Filtering

In practice, there is a requirement to filter the signal at the modulator and the receiver. In general, the receiver will include a “matched filter” prior to the bit/symbol decision-making process. A matched filter maximizes output signal

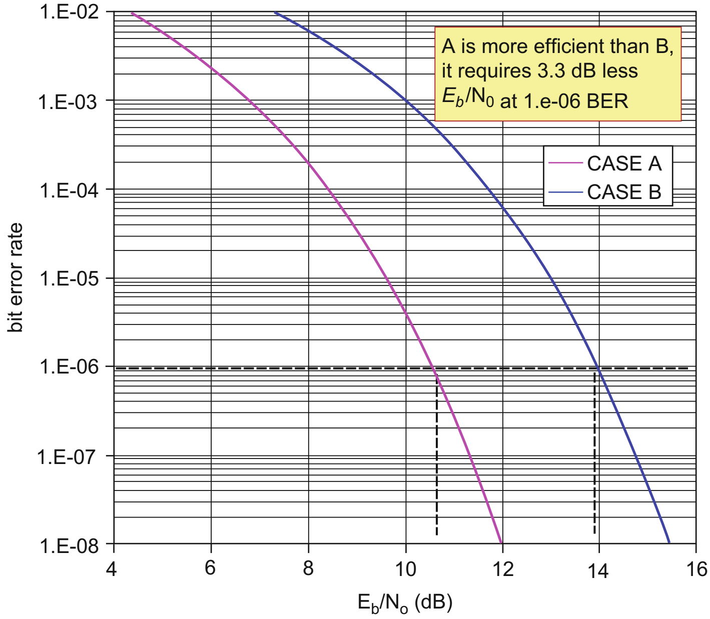


Fig. 9 Typical BER versus E_b/N_o curves

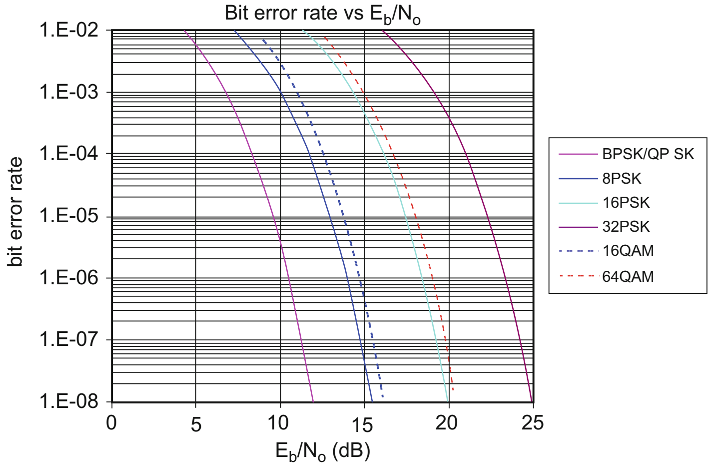


Fig. 10 BER versus E_b / N_o for a range of modulations appropriate to satellite links

to noise ratio for a particular symbol shape. The so-called Nyquist filters are usually adopted to constrain the spectrum to a sensible bandwidth. Filtering may introduce intersymbol interference (ISI), Nyquist filtering avoids this as much as possible.

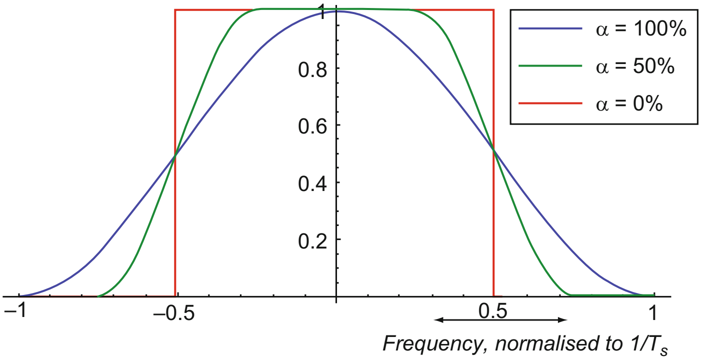


Fig. 11 Typical Nyquist filter shape

Nyquist filters have zeros in their impulse response at multiples of Ts and are usually raised-cosine filters. The frequency response of such a filter is given by the following formulas:

$$H(\omega) = 1 \text{ } \mathit{if} \text{ } 0 < \omega < \frac{\pi\,(1-\alpha)}{T_s}$$

$$H(\omega) = 0 \text{ } \mathit{if} \text{ } \frac{\pi\,(1+\alpha)}{T_s} < \omega$$

$$H(\omega) = \cos^2\left(\frac{T_s}{4\alpha}\left[\omega - \frac{\pi\,(1-\alpha)}{T_s}\right]\right) \quad \mathit{if} \quad \frac{\pi\,(1-\alpha)}{T_s} < \omega < \frac{\pi\,(1+\alpha)}{T_s}$$

where α is the roll-off factor which defines the filter sharpness and hence the bandwidth (Fig. 11).

One of the problems of filtering is the fact that the signal may no longer be a constant envelope and, therefore, more susceptible to any nonlinearity in the link. By nonlinearity, we mean variations in amplitude or phase introduced by the operation of an amplifier close to its maximum output such that they do not change in a linear manner.

Modulation Summary

- Modulation provides the mechanism of carrying the required data on an RF signal.
- A bit error rate needs to be chosen as it determines resource requirements and quality of performance offered.
- The main parameters of modulation schemes are bandwidth and power efficiency for a chosen E_b / N_o (or C/N) along with appropriate filtering.

- Performance should be determined in an appropriate noise environment. Additive White Gaussian Noise (AWGN) is often used to describe such noise.
- Multi-level modulation can increase bandwidth efficiency at the cost of power efficiency.

Coding

Communications links are susceptible to noise, and to achieve a high C/N (or E_b / N_o) implies a high signal power. In satellite communications, the RF power from the satellite is a key and valuable resource; consequently, the power is kept to a minimum while at the same time achieving the required performance in terms of C/N.

When power is at a premium, it is possible to trade power for bandwidth, provided the extra bandwidth is available. The tool for doing this trade is forward error correction (FEC) or "coding" as it is more commonly known. FEC operates by transmitting some redundant information in additional bits that can be used at the receiving end of the link to detect and correct errors. The more the redundant bits, the better the ability to correct errors but at the price of increased bandwidth as more bits are now being sent.

A measure of the improvement obtained with FEC is the so-called coding gain. This is the difference in required C/N for the coded case compared to the uncoded case at a specified BER. An example is given in Fig. 12.

Error correction codes can provide error-free communication through noisy channels provided the information rate does not exceed a specific limit, that is, the channel capacity as detailed by Shannon (the so-called Shannon limit). The Shannon limit

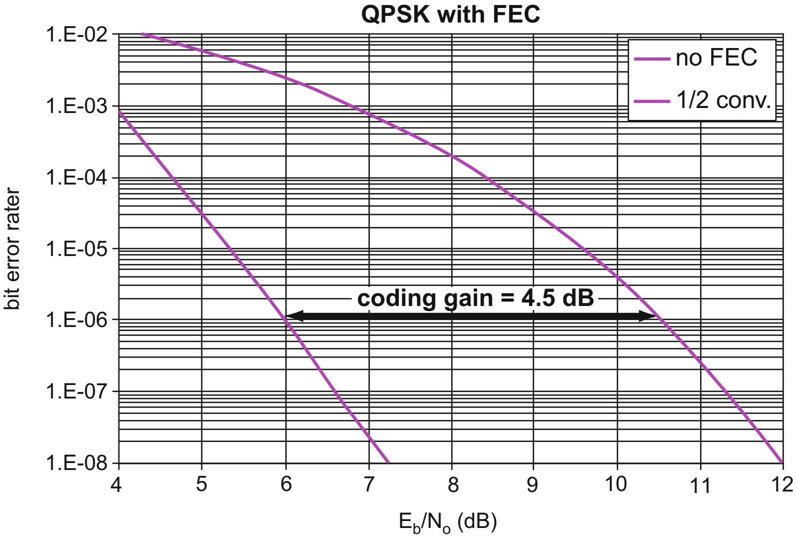


Fig. 12 Example of coding gain

defines the minimum E_b / N_o that can, in theory, provide error-free transmission for a given code rate in an AWGN channel, and the theory is based on random code structure and suggests that longer code blocks achieve better performance. However, long and random structure codes are difficult to decode, and practical techniques use an organized code structure to enable feasible decoding. Thus, the Shannon limit is difficult to achieve in practice, but improvements in digital technology have progressively made it easier to develop advanced codes with performance approaching the Shannon limit.

Error-correcting codes map a k-bit-long data word to an n-bit-long code word (where $k < n$), and $r = k/\ n$ is the FEC code rate.

In practice, two types of conventional FEC code approaches were adopted: the block codes and the convolutional codes. Both have advantages and disadvantages.

Block Codes

Block codes map directly a k-bit-long data word to a unique n-bit-long code word. For such codes, we find that the decoder complexity/memory grows exponentially with n. An important set of codes is known as the BCH cyclic block codes, which were invented by Hocquenghem and independently by Bose and Ray-Chaudhuri. The Reed–Solomon (RS) set of codes are a special set of BCH codes and are popular for outer coding.

Convolutional Codes

Convolutional codes have the advantage that optimal decoding complexity grows only linearly with the code word length but exponentially with the encoder's memory order. In a convolutional encoder output, bits are determined by the present input and previous input bits. The Viterbi algorithm is a particularly efficient tool for decoding the received signal and outputs the most likely data word.

Decoding Methods

There are several types of decoding methods applied to the demodulator output prior to the decoder.

Hard Decoding

- Makes symbol decisions directly from the received signal and maps the bit decisions as input to the decoder

Soft Decoding

- Determines symbol reliability values from the received signal using statistical approaches and maps these values as input to the decoder
- Offers better performance but with more computations

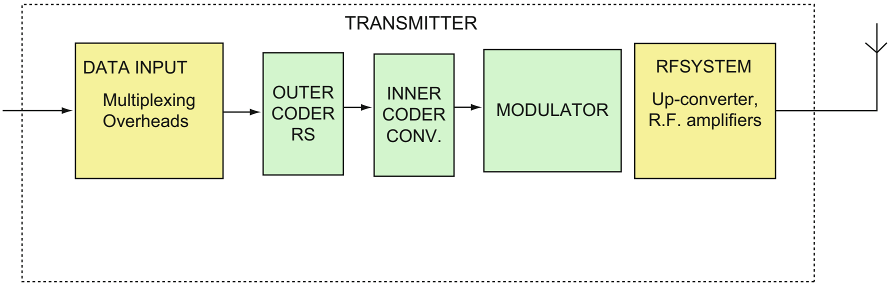


Fig. 13 Concatenated FEC coding

Serial Concatenation of Codes

There are significant advantages in adding two codes together in series, providing the additional computational complexity can be accepted. The two codes are normally called the inner code and the outer code (see Fig. 13). A good example employs convolutional coding (with Viterbi decoding for the inner code and a Reed–Solomon (RS) block code for the outer code). Such implementation exists for the INTELSAT IBS/IDR [2] services as well as the Digital Video Broadcast (DVB-S) standard [3]. Such an approach is useful to manage issues that arise from the fact that errors may be bunched together and, therefore, adversely impact convolutional codes used alone.

A two-stage concatenated coding scheme is usually used in the following manner:

- Inner code with soft decision decoding, targeted at random errors
- Outer code with hard decision decoding, targeted at bursty errors
- Outer code corrects inner decoding errors
- Useful for achieving very low BER (e.g., 10^{-11})

Concatenated "Reed–Solomon + Convolutional" coding was the state-of-the-art solution prior to the practical implementation of turbo codes in 1993.

Figure 14 presents a typical set of performance data for QPSK operation using a convolutional encoder with and without concatenated Reed–Solomon outer coding. It should be noted that typical performance is presented here. For more specific information, one can refer to manufacturers' data sheets.

When FEC is employed, the calculation of the various factors required in link performance assessment can be performed using the following equations.

---

[2] IBS/IDR stands for Intelsat Business Service/Intermediate Data Rate.

[3] Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for 11/12 GHz satellite services, ETSI EN 300 421 V1.1.

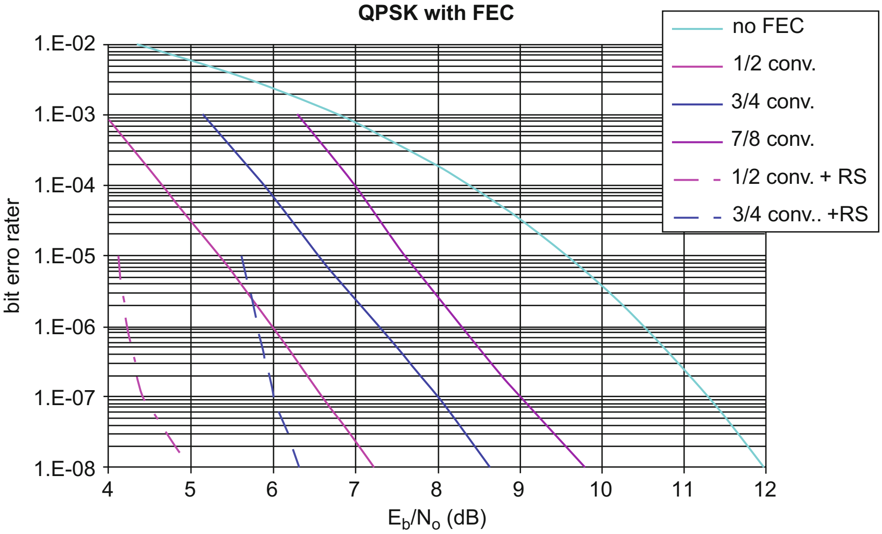


Fig. 14 Typical convolutional FEC and concatenated convolutional with Reed–Solomon performance

Determining the Required *C/N* from the Required $E_b\,/N_o$ and Vice Versa

$$C/N = \frac{E_b}{N_o} - 10\,\log_{10}(1+\alpha) + 10\log_{10}(m) + 10\log_{10}(r) \quad \text{dB}$$
$$\frac{E_b}{N_o} = C/N + 10\,\log_{10}(1+\alpha) - 10\log_{10}(m) - 10\log_{10}(r) \quad \text{dB}$$

where:

α =The filter roll-off factor (0 to 1)

r =The overall FEC rate, for example, for DVB concatenated Reed–Solomon and Viterbi ¾

$$r = 3/4 \times 188/204 = 0.6911$$

$(m = \log_2(M)\,)$ where M is the number of phase states (See Table 2)

Symbol Rate

$$\text{Symbol rate} = \frac{\text{Information bit rate}}{m \times r}$$

Table 2 M and m values for various modulation schemes

	M	m
BPSK	2	1
QPSK	4	2
8PSK	8	3
16APSK	16	4
32APSK	32	5

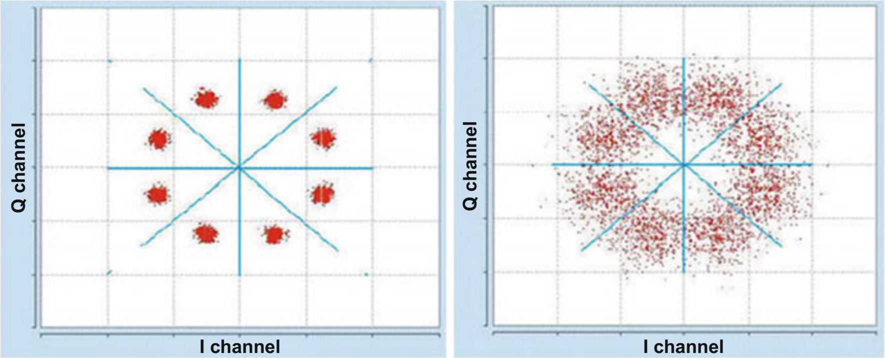


Fig. 15 Performance of an 8PSK channel with additive white Gaussian noise (*AWGN*) and convolutional rate 2/3 FEC. 8PSK FEC = 2/3: $E_b / N_o > 15$ dB (*left*), 8PSK FEC = 2/3 $E_b / N_o = 6.8$ dB (*right*)

Occupied Bandwidth

Occupied bandwidth $(B) = \text{Symbol rate} \times (1 + \alpha)$

Performance of a link may be visually observed using a plot of the received constellation diagram. Figure 15 depicts two such constellation diagrams with appropriate decision boundaries. The signal used in this example is an 8PSK signal with rate 2/3 forward error correction. When the E_b / N_o is high (left-hand plot), it is clear that very few data points cross into the incorrect decision region. However, when the E_b / N_o is low (right-hand plot), this is clearly not the case, and it would be tempting to observe that many errors are occurring. This is indeed possible, but the FEC takes care of these, recognizing that the constellation diagram is measured before the FEC takes effect (Fig. 15).

Evolution of Coding

In 1993, it was acknowledged that turbo coding was the first coding technique that can practically approach closely to the Shannon limit. Then in 1997, it was realized that low-density parity-check (LDPC) codes can also practically approach closely to the Shannon limit. Indeed, it transpired that LDPC codes had been first discovered by Prof. Robert Gray Gallager of MIT in the early 1960s but ignored by the coding research community as the implementation was not feasible at that time!

Turbo and LDPC codes use iterative soft decoding to enhance performance as compared to conventional FEC.

Low-Density Parity-Check (LDPC)

Defining features of LDPC codes are:

- Linear block codes.
- Sparse (low-density) parity-check matrix (more zeros than ones).
- Any two columns of the parity-check matrix H have no more than one nonzero entry in common.
- Performance close to Shannon limit.

The sparseness of the parity-check matrix of an LDPC code helps to guarantee a low decoding complexity.

LDPC is most effective for very large blocks and is employed in the well-known DVB-S2 standard with block lengths of the order of $n = 64{,}800$ (normal block) and $n = 16{,}800$ (short block). The long-code construction is consistent with Shannon's theorem which suggests that the performance of a code improves with its length. The use also requires the service to have a reasonable degree of "delay tolerance" as the processing of such large blocks may take some time.

Thus, large code blocks are very compatible with "delay-tolerant" broadcasting such as in DVB-S2, and LDPC has lower decoding complexity than turbo codes for very long code blocks, making it possible to approach closely to the Shannon limit.

The DVB-S2 standard [4, 5] uses irregular LDPC codes:

- Two code word lengths: $n = 64{,}800$ (normal block) and $n = 16{,}800$ (short block)
- Code rates: 1/4, 1/3, 2/5, 1/2, 3/5, 2/3, 3/4, 4/5, 5/6, 8/9, and 9/10
- Helps to support variable/adaptive coding and modulation

Figure 16 indicates the performance of the DVB-S2 LDPC. There are several points to be aware of when considering this figure. First is the fact that it uses packet error rate rather than bit error rate, and secondly it is presented in terms of E_s / N_o.

---

[4] See Note 1.

[5] Digital Video Broadcasting (DVB); User guidelines for the second generation system for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications, ETSI TR 102 376 V1.1.1.

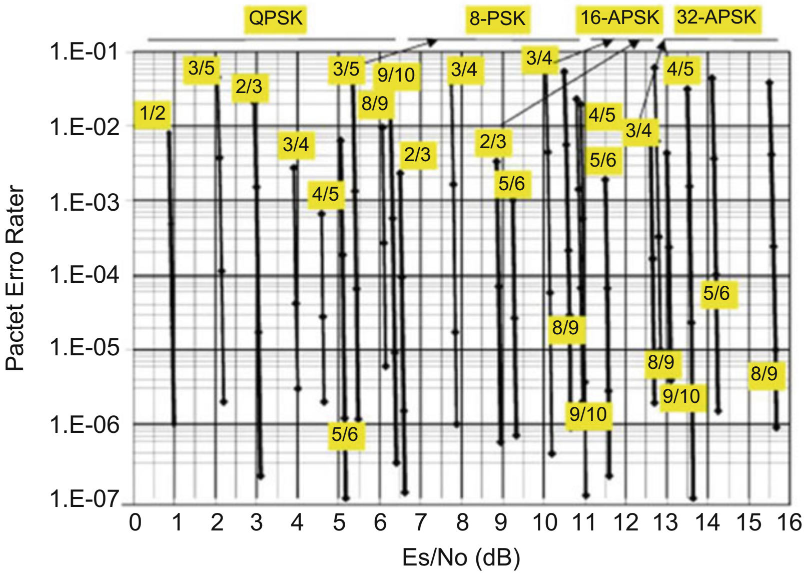


Fig. 16 Performance of LDPC + BCH codes over AWGN channel, $N = 64{,}800$ bits

A feature of such powerful codes is the fact that the range spanning excellent performance to failed performance spans a few tenths of a dB. Hence, it is more convenient to specify the performance in a table for a single error rate.

For the DVB-S2 standard, a quasi-error-free (QEF) performance is defined as "less than one uncorrected error event per transmission hour at the level of a 5 Mbit/s single TV service decoder," approximately corresponding to a transport stream packet error ratio PER $<10^{-7}$ before demultiplexer.

Table 1 indicates an example of the range of parameters relating to DVB-S2 with a block length of 64,800 bits.

Turbo Codes

The turbo encoder is constructed by the parallel concatenation of recursive systematic convolutional (RSC) constituent encoders. Recursive encoders are a key component of turbo codes. They enable a "code interleaver gain" factor. Performance improves with code interleaver length.

Turbo decoding relies on constituent soft-output trellis decoders interconnected in a "turbo" fashion with several iterations being performed to progressively improve the decision-making process. The error performance improves with the number of turbo loop iterations for which typically eight iterations are adequate. Parallel concatenated turbo codes typically suffer from an error floor at BER $<10^{-5}$; however, serial concatenated turbo codes do not suffer from this problem.

Table 3 DVB-RCS turbo-coded performance for PER = 10^{-7}

FEC	E_b / N_o (188 byte packets) dB	E_b / N_o (53 byte packets) dB
1/3	2.5	2.9
2/5	2.7	3.1
1/2	3.2	3.6
2/3	4.0	4.6
3/4	4.6	5.4
4/5	5.3	6.3
6/7	6.0	7.0

Turbo coding is used in the DVB-RCS [6, 7, 8] and DVB-SH standards as they tend to outperform LDPC for short code lengths.

Typical DVB-RCS turbo-coded performance is given as packet error rate or PER = 10^{-7}.

Table 3 presents an example of the E_b / N_o characteristics of the turbo-coded DVB-RCS return link.

Coding Summary

Error control coding mitigates the effect of channel noise under the right conditions. Forward error correction (FEC) coding is indispensable as satellite links are very often power limited and delay intensive, and therefore, FEC involves a trade-off between power and bandwidth.

The Shannon limit defines the optimum performance, but computational complexity/latency determines the achievable FEC performance that can actually be achieved.

Concatenation of convolutional encoders with Reed–Solomon block codes gives a robust implementation with reasonable performance.

LDPC and turbo codes are state-of-the-art FEC techniques. These techniques approach the Shannon limit in AWGN channels with practical implementation complexity. They, in fact, have been implemented in many recent satellite communications standards such as the latest DVB specifications.

---

[6] Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems, ETSI EN 301 790 V1.5.1.

[7] Digital Video Broadcasting (DVB); Interaction channel for Satellite Distribution Systems; Guidelines for the use of EN 301 790, ETSI TR 101 790 V1.4.1.

[8] Digital Video Broadcasting (DVB); Interaction channel for Satellite Distribution Systems; Guidelines for the use of EN 301 790 in mobile scenarios, ETSI TR 102 768 V1.1.1.

Multiple Access

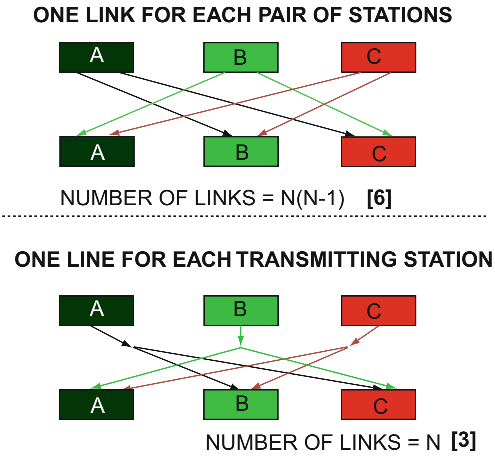
Introduction

Multiple access is the mechanism by which terminals in a satellite network communicate to each other. Figure 17 indicates the value of using some form of multiple access. In the upper part of the figure, each station sends a dedicated carrier to the stations it wishes to communicate with. This requires $N \times (N - 1) = 6$ links in the case shown. In the lower part of the figure, the case is depicted where a station combines traffic to all terminals (by multiplexing) such that it transmits only one carrier, the receiving station selecting the traffic destined for it. Such a configuration requires N links. If there were 100 terminals in the network, then the saving in number of carriers is very significant. There are many approaches to facilitate this multiple access approach and they are the subject of this section.

As the underlying telecommunications networks evolved to digital operation and the user networks transitioned from simple circuit-switched approaches (where a lot of the access functions were handled in the circuit switch) to packet transmission, there was a requirement to address multiple access in much greater detail. Hence, a requirement was established for some form of medium access control (MAC) that covered the key aspects of multiple access in a digital era.

MAC protocols can be classified in several dimensions:

- Static versus dynamic
- Distributed versus centralized
- Synchronous versus asynchronous



EARTH STATION TRAFFIC MATRIX

FORM	**TOA**	**TOB**	**TOC**
STATION A	---	t_{AB}	t_{AC}
STATION B	t_{BA}	---	t_{BC}
STATION C	t_{CA}	t_{CB}	---

Traffic is expressed in terms of required capacity (number of voice Channel or bit rate).
$\Sigma\, t_{IJ}$ = network capacity

Fig. 17 The value of multiple access

Popular MAC protocol classifications are:

- Fixed assignment
- Demand assigned multiple access (DAMA) (also called contention-less access or reservation access or scheduled access)
- Random access (also called contention-based access)

Access Techniques

- Single channel per carrier (SCPC)
- Frequency division multiple access (FDMA)
- Time division multiple access (TDMA)
- Code division multiple access (CDMA)
- Multi-frequency time division multiple access (MF-TDMA)
- Random access (RA) techniques for packet mode operation

Satellite Network Assignment Approaches

Basic multiple access approaches are depicted in the following figure (Fig. 18).

Fixed assignment: The frequency – time resource (i.e., RF channel) is shared between earth stations according to a scheme that does not vary with time.

Demand assignment: The frequency – time resource is shared between earth stations according to the demand from individual stations. If a terminal has no traffic demand at a particular instance in time, then no capacity is made available to it and others may be assigned to it (from a pool of resources). Such an arrangement is called demand assigned multiple access (DAMA) (or bandwidth on demand). DAMA is often associated with SCPC or TDMA including MF-TDMA.

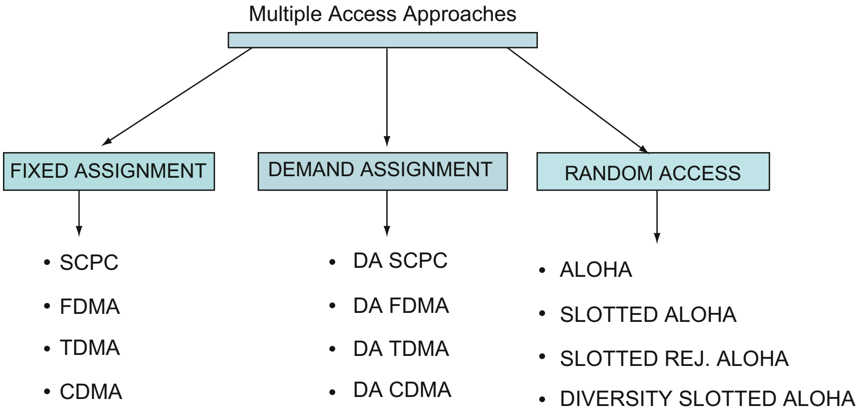


Fig. 18 Multiple access approaches

Random access: In random access, the stations or terminals randomly attempt to transmit a data packet. It is possible that two or more such random transmissions collide at the satellite resulting in failure to correctly receive them. This is handled by a process of retransmitting corrupted transmissions until the message gets through correctly. This can result in quite long delays.

Some random access protocols are:

- ALOHA, the basic contention-based access scheme
- Slotted ALOHA
- Selective reject ALOHA
- Diversity slotted ALOHA (DSA)

ALOHA is the basic contention-based access scheme described above.

Slotted ALOHA is similar but transmissions are restricted to specific time slots. This has twice the throughput capability of simple ALOHA before the system becomes unstable.

In ALOHA and slotted ALOHA, occasional collision among packets occurs, and on the average, a packet may have to be transmitted more than once before it is received correctly. This will introduce large packet delay in satellite slotted ALOHA systems, where each round trip propagation delay may be too large.

In selective reject ALOHA, larger packets are subdivided into sub-packets, and only damaged sub-packets are retransmitted.

DSA is a slotted ALOHA scheme in which, whenever a user generates a packet, it transmits k copies of the same packet. It is assumed that there exists some arrangement which allows a receiver to reject all but one correctly received copy of any packet. As the system may get many packets through without having to signal a retransmit event, it has good throughput with lower average delays.

Further Detailed Considerations

Single Channel per Carrier (SCPC) Access

The concept of SCPC is shown in Fig. 19.

FDMA Access

The concept of FDMA is shown in Fig. 20.

Advantages

- Use of existing hardware to a greater extent than other techniques
- Network timing not required

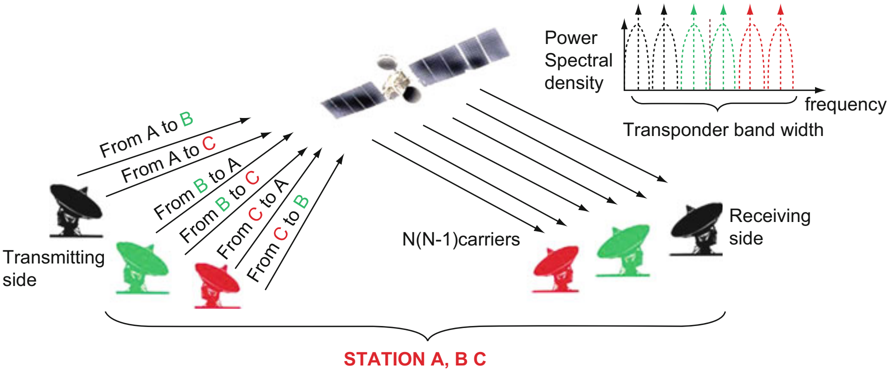


Fig. 19 The SCPC satellite access concept

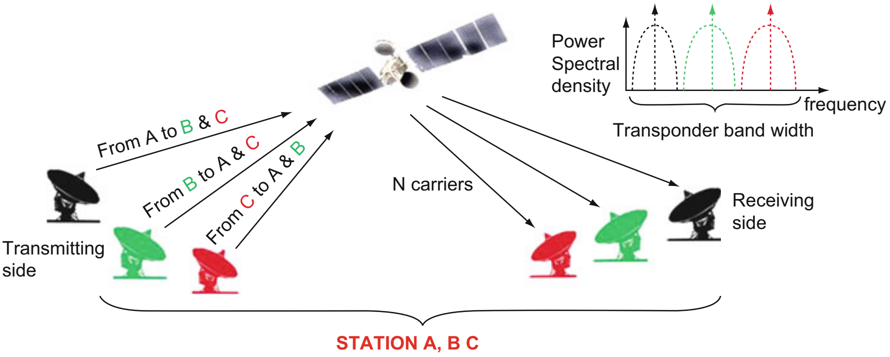


Fig. 20 The FDMA satellite access concept

Disadvantages

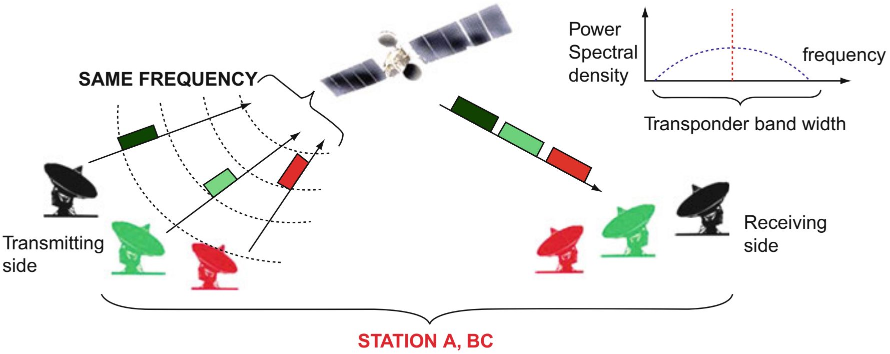
- As the number of accesses increases, intermodulation noise reduces the usable repeater output power (i.e., TWTA back-off). Hence, there is a loss of capacity relative to single carrier/transponder capacity
- The frequency allocation may be difficult to modify
- Uplink power coordination is most likely to be required

If demand assigned multiple access is adopted, then the resource allocated is the frequency band to use.

TDMA Satellite System

In a TDMA system, each earth station transmits traffic bursts, synchronized so that they occupy *assigned non-overlapping* time slots. Time slots are organized within a periodic structure called *time frame*.

The concept is depicted in Fig. 21.



- **A burst is received by all stations in the downlink beam and any station can extractits traffic from any of the bursts**
- **A BURST = link from one station to several stations (TDMA = one-link-per-station scheme)**

Fig. 21 The TDMA satellite access concept

Advantages

- Digital signaling provides easy interfacing with developing digital networks on the ground
- Digital circuitry has decreasing cost
- Higher throughput compared to FDMA when the number of accesses is large

Disadvantages

- Stations transmit high bit rate bursts, requiring large peak power.
- Network control is required.
  - Generation and distribution of burst time plans to all traffic stations.
  - Protocols to establish how stations enter the network.
  - Provision of redundant reference stations with automatic switchover to control the traffic stations.
  - Means for monitoring the network are needed.

Fixed and demand assigned FDMA and TDMA are shown in an alternative manner in Figs. 22 and 23, respectively.

Code Division Multiple Access

Such multiple access concepts adopt the principle of spread spectrum whereby the transmitter spreads baseband signal from bandwidth N to W by impressing the lower speed data on a much higher speed address code sequence.

W/N is the spreading factor (100 to 1,000,000), often called coding gain. The receiver de-spreads the received signal by "knowing" the proper address code. Any signal not generated with that code is not de-spread and remains as wideband noise (that slightly increases the noise that the terminal has to tolerate). Thus, received

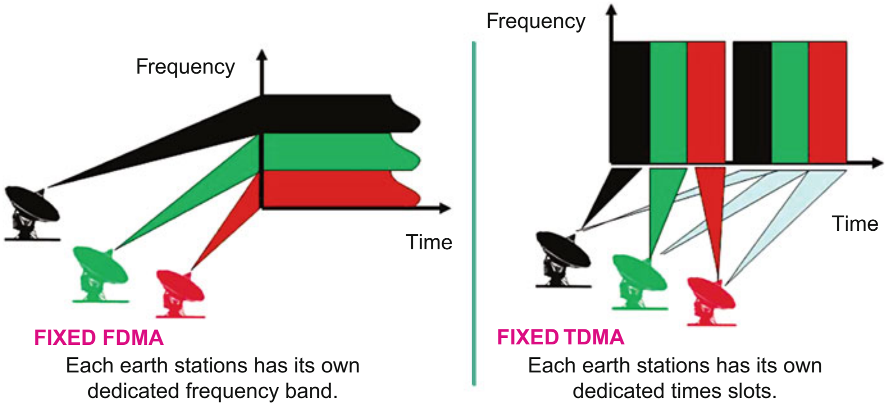


Fig. 22 Alternative depiction of FDMA and TDMA

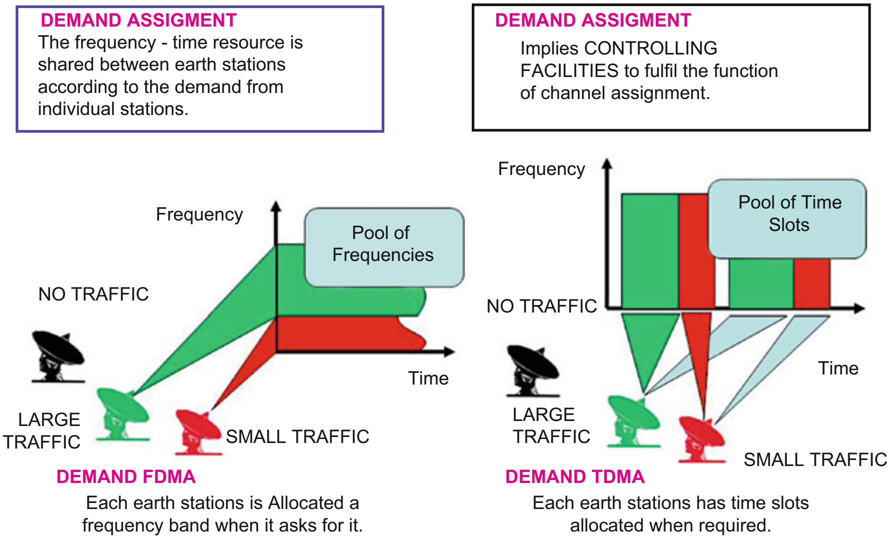


Fig. 23 Alternative depiction of FDMA and TDMA with DAMA

signals with other address codes and jammers are spread by receiver and act as noise. Addresses are periodic binary sequences that either modulate the carrier directly (*direct sequence systems*) or change the frequency state of the carrier (*frequency hopping systems*). The multiple access process is assured by operating the carriers on the same RF frequency but employing different codes for differing links. The codes are specially selected to have high decorrelation properties and are thus orthogonal codes. The concept of spreading and de-spreading is shown in Fig. 24 and that of CDMA in Fig. 25.

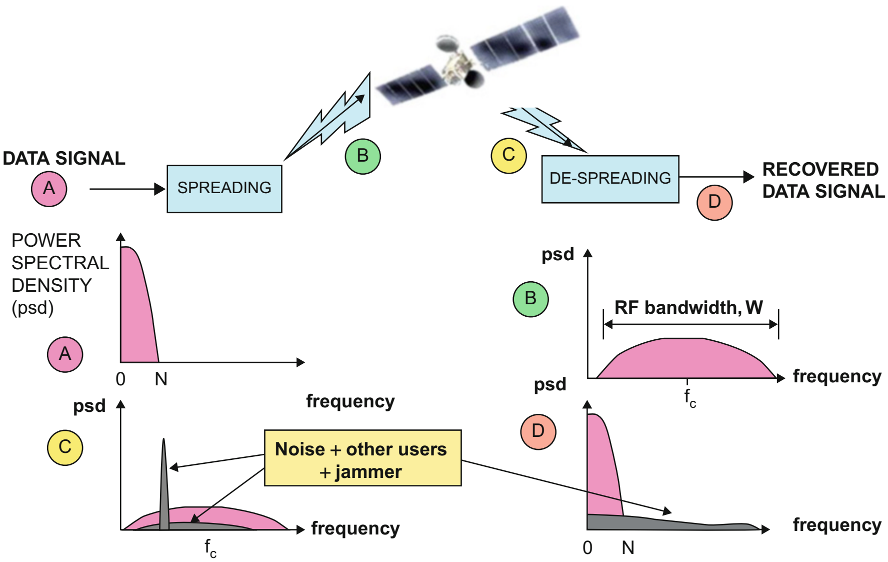


Fig. 24 The spread spectrum spreading and de-spreading concept

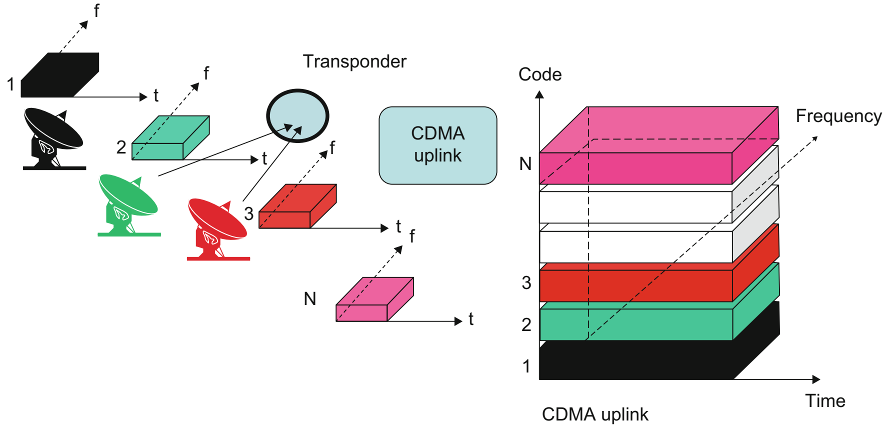


Fig. 25 The CDMA satellite access concept, each code being different (orthogonal)

MF-TDMA: Multiple TDMA Streams

When a large number of low-traffic terminals exist in a network, TDMA on the return link from the small terminals is not effective as the satellite transponder is not operated in a single carrier mode which avoids intermodulation effects. However, an effective configuration is to operate several small TDMA streams separated by

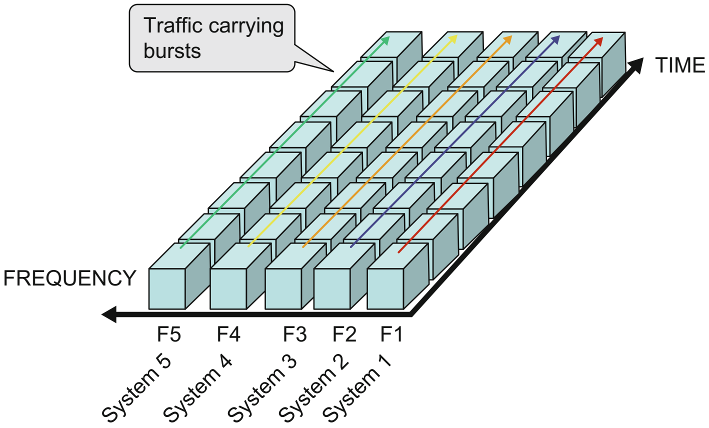


Fig. 26 Multiple TDMA streams (MF-TDMA)

frequency as depicted in Fig. 26. This is known as multi-frequency TDMA (MF-TDMA) and is used quite extensively in VSAT networks.

DAMA with MF-TDMA

In order to achieve a greater efficiency in small terminal satellite networks, the DAMA scheme can be used with MF-TDMA. By exchanging signaling packets, the process of call setup involves the negotiation among the earth station and a master control station (MCS) which controls the satellite network. The signaling packets are transmitted in the synchronization area which is a fixed portion in the MF-TDMA frame. Once the connection is established, a certain amount of memory and bandwidth is allocated to the new connection.

DVB-RCS: MF-TDMA

We consider here the DVB-RCS return link with MF-TDMA and DAMA. In this standard, there are four types of burst in the MF-TDMA:

- Traffic (TRF)
  - ATM (53 bytes)
  - MPEG (188 bytes)
- Acquisition (ACQ)
- Synchronization (SYNC)
- Common signaling channel (CSC)
  - Used by a terminal to identify itself during log-on

ACQ and SYNC bursts are required for accurately positioning the terminals burst during and after log-on. The concept is outlined in Fig. 27 where the bursts and their roles are depicted.

The DVB-RCS standard provides for the traffic bursts to change frequency as well as time slots in order to keep the spectrum loading balanced for interference advantages. Figure 28 shows the evolution of a five-frequency DVB-RCS MF-TDMA system over time. At the start of the displayed time frequencies, 3–5 are already using allocated time and frequency slots with existing traffic, while frequencies 1 and 2 are just preparing to handle traffic after requesting it via the slotted ALOHA CSC random access slot. It is of value to note that at any given frequency in the band used for MF-TDMA, there is only one carrier transmitting at a given instance, thereby providing a means to keep interference from the system at a low level.

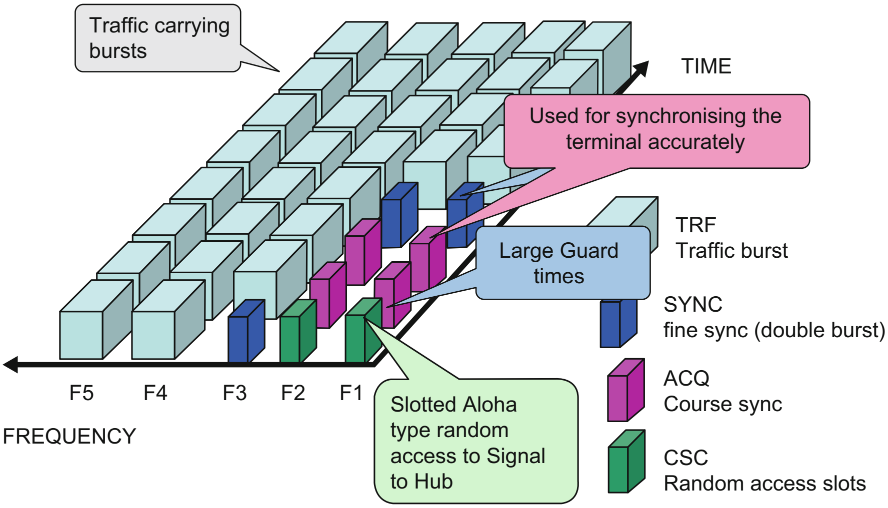
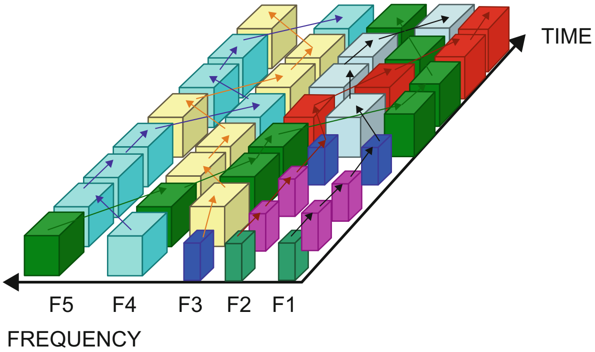


Fig. 27 The DVB-RCS MF-TDMA concept

Fig. 28 The DVB-RCS MF-TDMA slot use over time



Comparison of FDMA, TDMA, and CDMA

The comparison of FDMA, TDMA, and CDMA is presented in Table 4 terms of advantages and disadvantages.

The throughput of these various multiple access systems is of interest and varies with the number of accesses in the system. Figure 29 presents a depiction of throughput of the different approaches where 100 % throughput corresponds to the capacity of the system with just one access only.

Table 4 Comparison of FDMA, TDMA, and CDMA

Type of multiple access	Advantages	Disadvantages
FDMA	Network timing not required	Intermodulation products cause degradation and poor power utilization
	Compatible to a lot of existing hardware	Uplink power control may be required
TDMA	No mutual interference between accesses	Network control required
	Uplink power control not needed	Large peak power transmission for earth station
	Maximum use of satellite transponder power, most efficient	Being digital in nature interface with analog system is expensive
CDMA	Network timing not required	Wide bandwidth per user required
	Anti-jamming capability	Strict code sync. needed, lower capacity

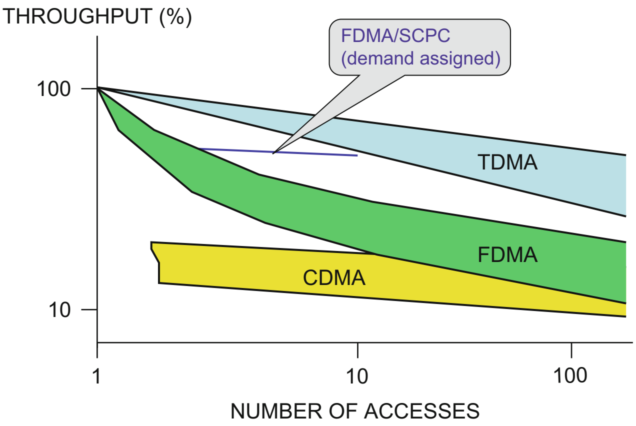


Fig. 29 Comparison of throughput for various access systems

Random Access Schemes

FDMA/TDMA/CDMA fixed-access schemes have been designed for circuit/stream traffic. Bursty data traffic, for example, packets, are more efficiently dealt with via random access schemes.

In random access, there are no permanent assignments. In this case, available resource is allocated when needed on a random basis.

The simplest system is ALOHA – where packets are randomly transmitted and if they collide with others are retransmitted with random time difference. ALOHA does not need synchronization, but the maximum theoretical throughput is a mere 18 %. A key impairment with such a scheme is the impact of the delay in getting information back to the transmission site that the packet was corrupted.

SLOTTED ALOHA confines transmission to slot boundaries and needs time synchronization, but the maximum throughput is increased to 36 %. It is used, for instance, in the signaling channel of DVB-RCS.

As the system rapidly becomes unstable as collisions build up, it is customary to operate below these maxima.

For variable length messages, we need to employ a more complex scheme, for example, selective reject ALOHA, which breaks long packets into sub-packets and only retransmits sub-packets that collide; this provides a throughput of ≈ 0.37 which is independent of message length.

Diversity slotted ALOHA (DSA). In slotted ALOHA, occasional collision among packets occurs; on the average, a packet may have to be transmitted more than once before it is received correctly. This will introduce large packet delay in satellite slotted ALOHA systems due to the large round trip propagation delay via the satellite. DSA is a slotted ALOHA scheme in which, whenever a user generates a packet, he transmits k copies of the same packet. It is assumed that there exists some arrangement which allows a receiver to reject all but one correctly received copy of any packet.

It has been found that multiple packet transmission gives better delay performance if the throughput is somewhat below its maximum. This technique is utilized in the IP over Satellite (IPoS), the European Telecommunications Standards Institute (ETSI TS 102 354), and the Telecommunications Industry Association (TIA-1008-A).

Figure 30 indicates in diagram form how the various random access schemes function.

A comparison of the performance of the different approaches is given in Fig. 31.

Selecting a Random Access Scheme

Select RA scheme for traffic type and delay/throughput. Take care to achieve stability.

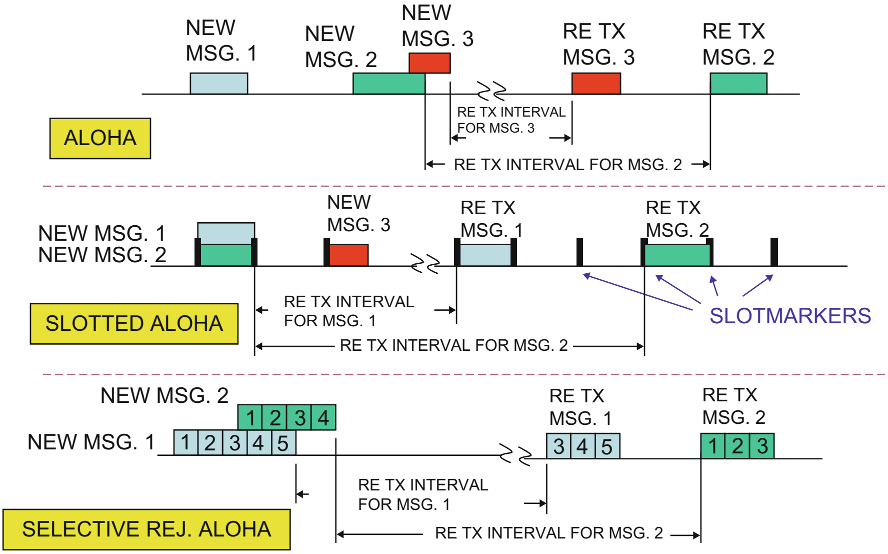


Fig. 30 Comparison of random access approaches

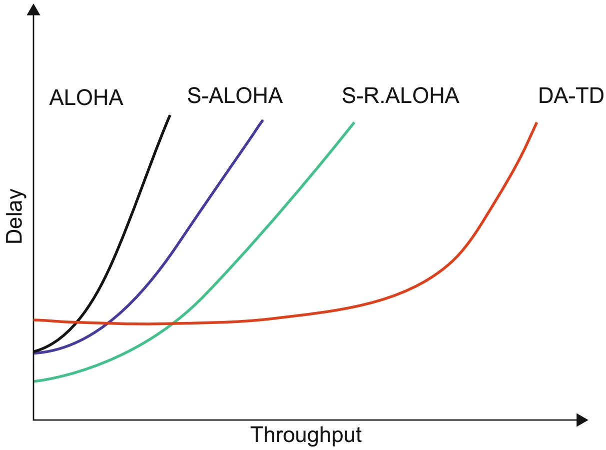


Fig. 31 Comparison of random access performance

- ALOHA: Short bursty traffic
- S-ALOHA: Short bursty traffic – better throughput (used for signaling in DVB-RCS)
- S-R ALOHA: Variable length, longer messages
- DSA: Diversity slotted ALOHA (used for signaling in IPoS)

Multiple Access Summary

- This section has outlined the various access methods used in satellite communications.
- For traditional circuit-switched voice and data, FDMA has been commonly adopted but is less used today.
- For packet-based services, TDMA and CDMA are used with the former, gaining in popularity.
- Demand assignment is commonly employed with SCPC, TDMA, and CDMA which significantly improves the capacity of the system.
- Multi-frequency TDMA (MF-TDMA) is popular for low-power multi-terminal uplink operation such as for DVB-RCS-type service.
- Random access methods have been outlined and compared. They are commonly used in signaling channels for MF-TDMA.
- Assignment and multiple access methods are key parts of any satellite network and are thus of critical importance.

Conclusion

This chapter has provided a background on how the information to be carried over a satellite link can be impressed on an RF carrier and then extracted at the far end. Furthermore, the adoption of forward error correction (coding) has been outlined to demonstrate its value in optimizing the overall link capacity.

Connectivity between stations in a network has been addressed in terms of multiple access schemes and their functionality and efficiency. The various alternative systems used in digital satellite communications in terms of modulation and demodulation, multiple access techniques, and forward error correction and coding are discussed in terms of their individual merits and their advantages and disadvantages.

Cross-references

- ▶ Ground Systems for Satellite Application Systems for Navigation, Remote Sensing, and Meteorology
- ▶ Satellite Radio Communications Fundamentals and Link Budgets
- ▶ Satellite Transmission, Reception, and Onboard Processing, Signaling, and Switching
- ▶ Telemetry, Tracking, and Command (TT&C)

References

K. Pahlavan, A.H. Levesque, *Wireless Information Networks* (Wiley, Hoboken, 1995). ISBN 0-471-10607-0

J.G. Proakis, M. Salehi, G. Bauch, *Contemporary Communications Systems* (Brooks/Cole, Pacific Grove, 2004). ISBN 0-534-40617-3

Further Reading

A. Burr, *Modulation and Coding* (Prentice Hall, Upper Saddle River, 2001)

I. Glover, P. Grant, *Digital Communications*, 2nd edn. (Prentice Hall, Upper Saddle River, 1998)

S. Haykin, M. Moher, C. Systems, *Communication Systems*, 5th edn. (Wiley, New York, 2010) (International Student Edition)

S. Lin, D. Costello, *Error Control Coding*, 2nd edn. (Prentice Hall, Upper Saddle River, 2004)

T. Moon, *Error Correction Coding* (Wiley, New York, 2005). doi:10.1002/0471739219

K. Pahlavan, A. Levesque, *Wireless Information Networks* (Wiley, New York, 1995)

J. Proaki, *Digital Communications*, 3rd edn. (McGraw Hill, New York, 1995)

M. Reza Soleymani, Y. Gao, U. Vilaipornsawai, *Turbo Coding for Satellite and Wireless Communications* (Kluwer, Boston, 2002)

B. Sklar, *Digital Communications*, 2nd edn. (Prentice Hall, Upper Saddle River, 2001)

P. Sweeney, *Error Control Coding* (Wiley, Chichester, 2002)

Satellite Transmission, Reception, and Onboard Processing, Signaling, and Switching

Bruno Perrot

Contents

Abstract

This chapter explains the technology that makes onboard processing (OBP) function as well as explores the new and important applications that communication payloads, based on onboard processing techniques, can effectively support. Further, it assesses the pros and cons associated with employing this technology in terms of performance, complexity, reliability, and cost. Satellite systems providing fixed and mobile services are evolving from bent-pipe payloads to more and more enhanced satellites with more and more capabilities and "intelligence." Thus one has seen the evolution of satellite capabilities to be able to achieve more and more functionally in space. We started with the so-called nonintelligent or bent-pipe satellites and then moved quickly to more flexible multi-points-type satellite services. Next, there was the transition to more

---

B. Perrot (✉)
SES, Betzdorf, Luxembourg
e-mail: bruno.perrot@ses.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_16

enhanced satellites with onboard switching, and then most recently there have been design innovations to bring true “intelligence to space.” This has been seen in the move toward highly capable satellites with increasingly “intelligent forms” of onboard processing (OBP).

This evolution involves moving from more efficient beam switching to actual processing of signals to enhance signal and remove attenuation affecting the uplink and thus partially overcome rain attenuation. The addition of so-called intelligent functions to the satellite that were once found only in terrestrial signaling and switching systems allows satellites to become more efficient and versatile.

In particular, this transition will allow the design and deployment of:

- Multibeam RF-IF switched transponder satellites (i.e., the ability to provide effective “beam switching” among satellite beams). This allows satellites to provide Physical Transport Layer Network Services that were once restricted to advanced terrestrial networks.
- And eventually there will be an evolution to advanced packet switched (“Data Switched” asynchronous transmit mode (ATM or ATM-like services). This will allow onboard processed multibeam satellite systems that provide specific and an increased array of network-level services.

Keywords

Asynchronous transfer mode (ATM) switching • Baseband processing • Bent-pipe satellites • Bit error rate • Intelligent satellite • Intermediate frequency (IF) • Multibeam antennas • Onboard processing (OBP) • Radio frequency • Solid state amplifiers (SSAs) • Traveling wave tube amplifier (TWTA)

Introduction

Onboard switching systems are designed to make more efficient use of a satellite communication network, especially those that employ multibeam technology that entails onboard switching to interconnect uplink and downlink beams with a high degree of efficiency. Onboard processing can also be used to reduce bit error rates for the uplink and downlink transmissions and to allow satellites to be optimized to provide a wider range of service applications more efficiently.

The added cost and complexity of including such a capability on a satellite, however, has for a number of years served to argue against investing in this technology. Industry players and satellite operators must, in effect, look at the “opportunity cost” represented by adding onboard capability to a satellite in terms of the extra mass, volume, and engineering and manufacturing costs. The cost, mass, volume, and reliability considerations associated with an OBP payload rule out the option of simply adding more conventional throughput capability that would have little incremental cost. The service flexibility and throughput efficiencies, the

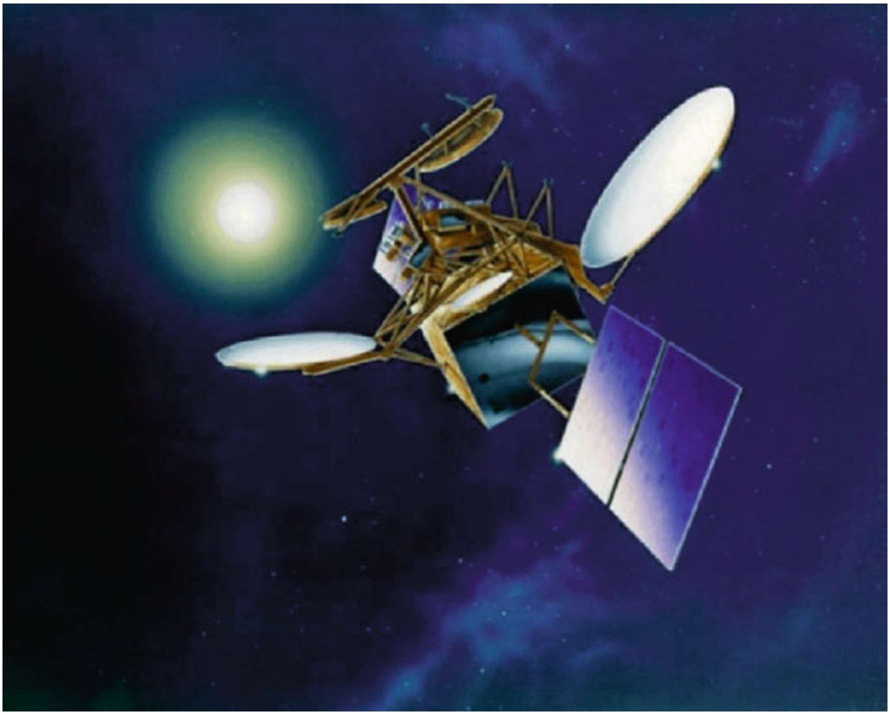


Fig. 1 The NASA developed Advanced Communications Technology Satellite (ACTS) (Graphics courtesy of NASA)

improved bit error rate capabilities, and the greater functionality of an "intelligent spacecraft" on the other hand have over time served to move telecommunications satellites – particularly those providing fixed and mobile services – in this direction over time. In earlier sections, it has been noted that experiments to demonstrate onboard processing have been in progress for some time. Most recently there have been the experimental satellite systems of Japan, including the WINDS and QZSSS spacecraft. It was the Advanced Communications Technology Satellite (ACTS) deployed by NASA over a decade ago that first demonstrated onboard processing for space communications and ATM networking (Fig. 1).

Services Applications

In a historical context, satellites have served to provide a connection between physical nodes of a network and not perform any of the higher level communications functions. Thus telecommunications satellites have in the past not participated as active nodes within networks and processed where or how information is routed. This is to say that the active signaling and processing of information at base band or

Table 1 OSI model levels – used in ATM switching according to X.200 recommendation

The seven layers	Function
#7: Application	Network process to application: actual content such as e-mails, video images, voice, and data
#6: Presentation	Data representation, encryption and decryption, convert machine dependent data to machine independent data: provides for such functions as encryption or data conversion
#5: Session	Inter-host communication: starts and stops sessions and creates the correct order
#4: Transport	End-to-end connections, reliability and flow control: ensures that the entire and complete message is delivered
#3: Network	Path determination and logical addressing: routes information to a particular location based on network address
#2: Data link	Physical addressing: routes data packets from node to node based on station addresses and the actual transmission mode
#1: Physical	Media, signal, and binary transmission: provides the physical channel to connect nodes in a network

Note that the layers 1–3 are named “Media layers” while layers 4–7 are “Host layers.”

intermediate frequencies were all accomplished within terrestrial switching centers. Satellites, in contrast, were simply employed to transmit higher frequency RF signals as essentially “cables in the sky.” They worked only at the physical layer of the seven-layer OSI Model as provided in Table 1.

As satellites have become more complex and as multibeam antenna systems are used to switch waveforms between a growing number of uplinked and downlinked beams, spacecraft have been designed to become, in effect much “smarter.” This added processing power and “intelligence” have allowed satellites to accomplish more complicated functions and provide for functionality above the physical level of the Open System Interconnection (OSI) model. Since the processing power was being added in any event, the thought evolved as to how one might add even more signaling capability to the satellite to accomplish even more complex tasks at ever higher levels within the seven-layer model of ATM-based communications. The same logic has also been applied to the more streamlined architecture of TCP/IP protocol as is used within the Internet.

The most important thought in this regard has been the idea that additional processing power on the satellite could serve to demodulate information (i.e., waveforms) back down to the base band or intermediate frequencies.

This could allow improvements to the signal, could reduce bit errors in the uplinking and downlinking of signals, or otherwise assist with digital service applications. The basic thought has been that “smart satellites” could ultimately help allow services sent via telecommunications satellite to be more competitive with terrestrial fiber optics and digital switching and signaling systems on the ground. It could allow satellites to deliver advanced services directly to mobile users or small office or home office (SOHO) users without the need of operating through terrestrial ATM or other modern digital switches.

In short, satellites with onboard switches can play a more active role within the digital networks of the future. Two kinds of services can thus be supported.

Physical Transport Layer Services

This type of service is the most basic. Physical Transport Layer Satellite Services are considered to be systems that are utilized to provide a physical layer interconnection between two network nodes, such as a fiber or terrestrial link would be used. This level of service provides node interconnections that are distance independent and readily available for remote service providers.

Figure 2 illustrates such a network application. Note that for these services the satellite is not an active network-switching node in the manner of an ATM switch node or similar terrestrial network nodes.

Statistical multiplexing and buffering are kept on the ground and service revenue increases due to their utilization are therefore realized by the network operators and users, not the satellite operator.

It is envisioned that Physical Layer Network Satellite Services can be implemented using a channelized transponder payload (see next section for details) configured by ground control of the payload switching configuration. Network users obtain satellite resources via a terrestrial control center that controls user access and system configuration. This demands that an additional earth-space-earth hop is required for user to control center communication. Thus Physical Layer Network Satellite Services (PLNSS) are appropriate for applications that have set-up, reconfiguration, and tear-down requirements compatible with fractional second timelines or longer.

These types of PLNSS operations therefore appear appropriate for leased line, dynamic assignment, and replenishment operational applications. Representative applications are listed in Table 2 at the end of this chapter. These applications also

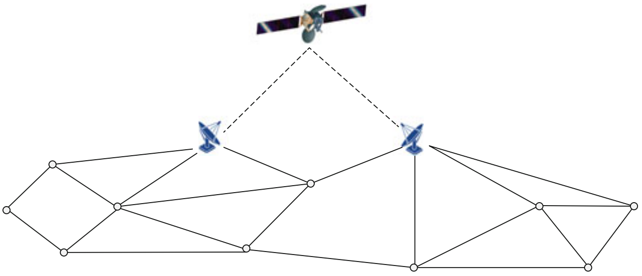


Fig. 2 Physical transport layer services

Table 2 Summary of implementation issues

Processor type	Applications	Advantages	Disadvantages
Channelized transponder processor			
Analog	Video conference News gathering Interactive video forward Two-way file transfer Software downloads VPN (larger bandwidth applications)	Switches arbitrary signals Switching flexibility increased compared to mechanical switches Minimum satellite power	No of cross-points drives complexity Limited by minimum channel bandwidth Variable bandwidth Mass, volume Multiple carrier D/L
Digital	Interactive audio Interactive video return link Games Data downloads Video conference Digital messaging Internet user Video streaming E-mail, fax (smaller bandwidth applications)	Switches arbitrary signals Full cross-point connectivity possible Small switching channel bandwidth possible Variable bandwidth possible	Power, mass, volume, thermal Multiple carrier D/L
De-mod-Re-mod processor			
Asymmetric	Interactive video forward Software downloads Data downloads (asymmetric BW applications)	Full cross-point connectivity possible TDM or packet routing possible to optimize throughput	Power, mass, volume, thermal Switches specific signals Single carrier D/L
Symmetric	Interactive audio Games Data downloads Video conference Two-way file transfer Digital messaging Internet user Video streaming E-mail, fax T1 virtual private network (VPN) (less than T1 applications	Full cross-point connectivity possible TDM or packet routing possible to optimize throughput	Power, mass, volume, thermal Multiple demodulators drives power requirements Switches specific signals Single carrier D/L

imply the use of satellite resources by a variety of networks, since channelized transponder implementations of physical layer satellite resources are independent of the waveform and relatively independent of the network protocols they are suited to varied applications.

Physical Layer Satellite Services implemented with a channelized transponder allow for modulation and protocol changes between users and over time. The satellite operator is therefore involved in a more traditional capacity and the necessary channel switching can be used to optimize channel capacity leasing opportunities.

Network Layer Services

Network Layer Satellite Services are services where the satellite is an active switching node within the network as illustrated in Fig. 3.

A Network Layer Satellite Network implementation provides network operations on the ground and on the satellite. The satellite operator therefore can benefit from the statistical channel loading and variable service-type multiplexing. Implementation of Network Layer Satellite Services requires the implementation of more complex De-modulation-Re-modulation (De-mod-Re-mod) payloads (see next section for details). This approach implies that all users must adhere to a fixed set of modulation and protocol formats compatible with the satellite. Changes between users and transition in modulation or protocol formats over time are difficult if not impossible to achieve. External network users are constrained by the satellite network implementation. The satellite operator is therefore effectively the network operator.

Satellite Implementation Issues

The design, engineering, and manufacture of a satellite with onboard processing thus leads to a number of key decisions involving the payload architecture, the antenna design, and the type of amplification sections that are utilized.

Table 2 below summarizes the advantages and disadvantages of each configuration and compares four different processor architectures.

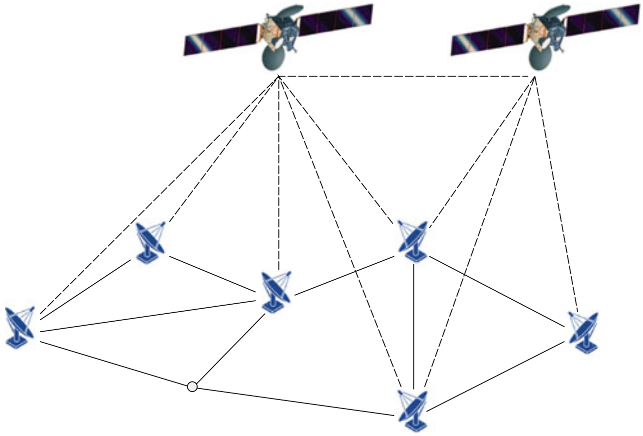


Fig. 3 Network layer services

Payload Architectures

Channelized transponder payloads filter received signals within a switchable bandwidth and connect the filtered signal to one or more downlink beams.

In contrast, De-modulator-Re-modulator payloads receive signals with specific uplink modulation formats, demodulate the received packets, process the packet addresses, route the received packets to the destination beams, and remodulate the packets with the specific downlink modulation format.

Physical Layer Network Satellite Services can be implemented with either Channelized or De-mod-Re-mod payload implementation. Since the channelized transponder implementation is simpler, the normal case would be for a channelized transponder payload to provide Physical Layer Services.

A De-mod-Re-mod payload could also be used for Physical Layer Services in a manner analogous to leased line services implemented by a switched network or virtual private network (VPN) operations. However, Network Layer Services, as used here, require the implementation of a De-mod-Re-mod payload. Baseband packet processing is required to implement the network node switching and routing functions.

Channelized Transponder Payload Implementation

Figure 4 summarizes the functional operation of a channelized transponder processor that assumes a digital implementation of the channelized transponder filtering and switching function. Any signal received within bandwidth A1–Z1 is down-

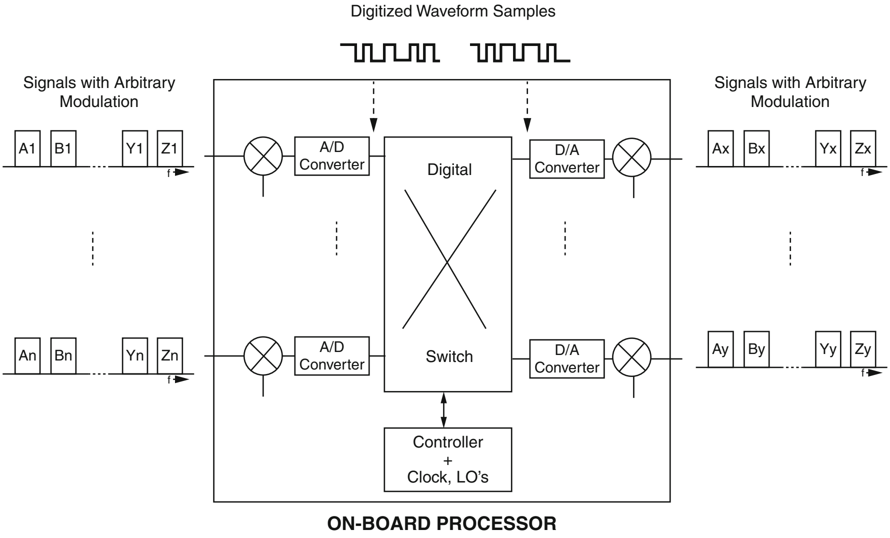


Fig. 4 Channelized transponder onboard processor architecture

converted to an intermediate frequency or baseband and digitally sampled. These samples are digitally filtered, stored, and routes to the switch port corresponding to the desired downlink beam. This routing may be accomplished by a simple readdressing of the stored digital samples within a common output buffer memory or by a more traditional digital switch implementation.

Figure 5, however, illustrates a conceptual block diagram for a channelized transponder payload hosting the processor described previously. Multiple uplink beams are formed by multiple feed apertures or phased array antennas. Multiple antenna apertures may be necessary due to feed packing, spillover, and beam shaping requirements. Current technologies support the implementation of received phased arrays for certain applications.

Signals received in each beam are amplified and down-converted before filtering and switching by the processor. Multiple processors may be utilized to minimize power dissipation and allow variable filter bandwidths. The output of the processor is a digitally routed and reconstructed version of the received waveform. Each output is up-converted to the appropriate downlink frequency and amplified. Similar to traditional transponders, channelized transponder payload amplifiers must be operated in back-off mode if multiple signals are present within the channel to avoid waveform degradation due to nonlinear effects. The resulting payload power efficiency is a significant consideration for channelized transponder payload applications.

Figure 5 also includes a multiple beam transmit antenna farm with a set of low noise amplifiers (LNAs) and traveling wave tube amplifiers (TWTAs).

The achievable power efficiency for solid state power amplifiers (SSPAs) is not currently high enough to be optimal to support transmit phased arrays. In future years this may be the case. Thus currently tube-based TWTA implementations are most likely indicated for the most cost-effective design, even though both options might be viable for lower powered systems. The technology in this area is in transition and the cost and performance equations are thus constantly changing. Additional system requirements that favor phased arrays such as beam reconfiguration or steering could mandate solid state amplifiers at least for some designs.

Channelized transponder processors and payloads can be implemented with currently available technologies. Receive multiple beam antennas can be implemented as passive arrays or active arrays that allow reconfiguration. Multiple beam systems require use of numerous low noise amplifiers (LNA) in receivers and down-converters. This generally requires one per beam. Technologies that enable power efficient implementation, interconnection, and packing of numerous components are required to support growth to a large number of beams. Channelized processor using surface acoustic wave (SAW) analog technology and digital technology has been developed for narrowband mobile satellite systems (MSS). These MSS technologies continue to advance. Larger bandwidth channelized processors are being developed utilizing the analog SAW technologies.

Current digital technologies can be used to implement channelized processors that take care of wider bandwidths. Digital processors often can more easily

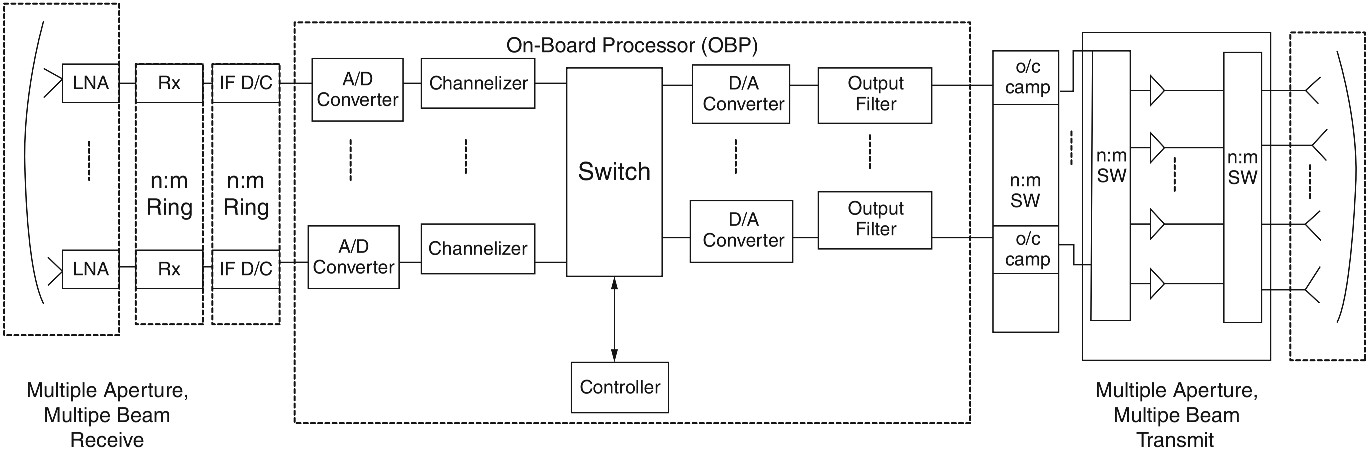


Fig. 5 Channelized transponder payload architecture

implement multiple or variable filter bandwidth than analog processors. Digital processors, however, require additional satellite power to achieve that degree of operational flexibility.

De-modulation-Re-modulation (De-mod-Re-mod) Payload Implementation

Figure 6 summarizes the functional operation of a digital implementation of a De-mod-Re-mod processor. Any signal received in channel A1–Z1 (See chart below) is down-converted to an intermediate frequency or baseband and digitally filtered, demodulated, and error correction coded. The demodulated packets are stored and routed to the switch port corresponding to the desired downlink beam. This routing may be accomplished by memory or by traditional digital switch implementation as discussed with respect to the channelized transponder functional description.

A conceptual block diagram for a De-mod-Re-mod payload which could utilize the processor, described above is shown in Fig. 7. Just as is the case for the channelized transponder, multiple uplink beams are formed by multiple feed systems working to conventional parabolic reflectors or phased array antennas. In both cases, the signals received in each beam are amplified and down-converted before filtering, demodulation, decoding, and switching within the processor.

The output of the processor is an encoded, re-modulated waveform. Each output is up-converted to the appropriate downlink frequency and amplified. In contrast to

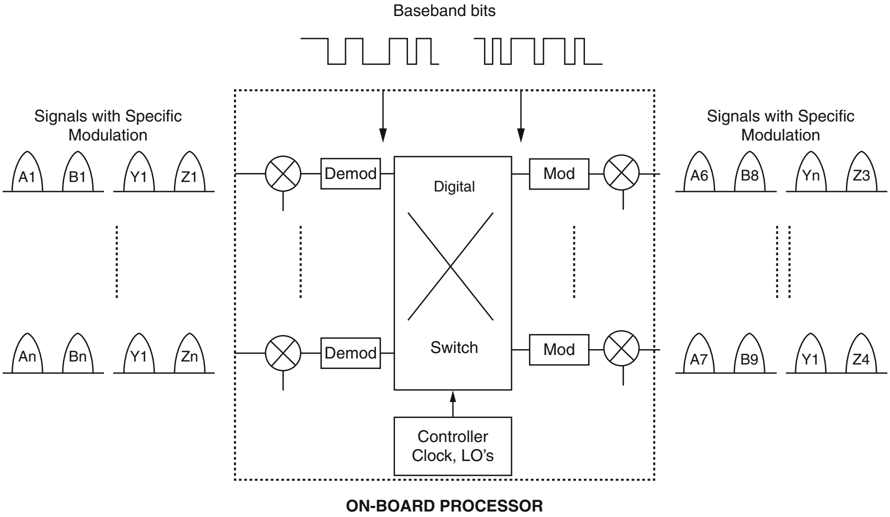


Fig. 6 De-mod-Re-mod onboard processor architecture

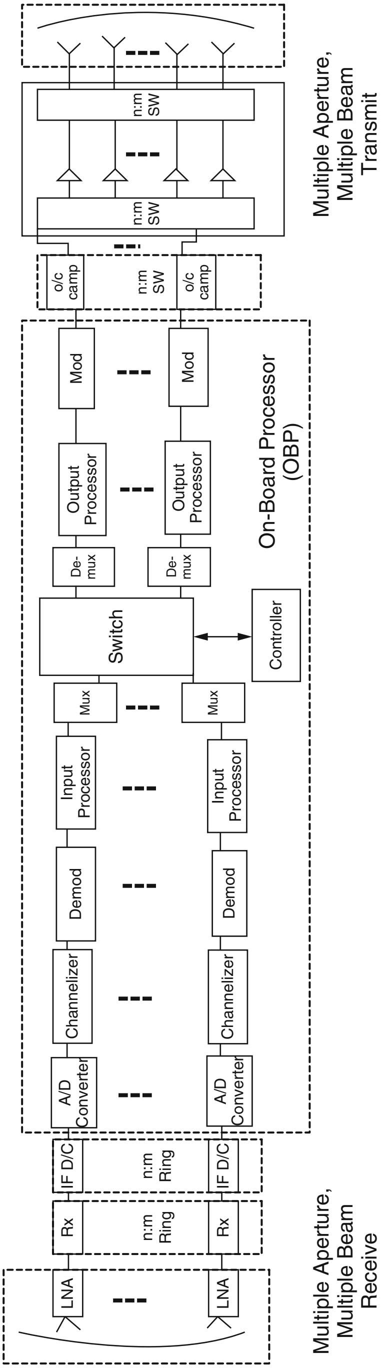


Fig. 7 De-mod-Re-mod payload conceptual block diagram

traditional transponders, De-mod-Re-mod payload amplifiers can be operated at saturation since the data from multiple uplink signals are combined at baseband by the switch and modulated onto one transmitted carrier. However, this power saving may be more than offset by the power required by the filtering and demodulation. This is true for the same reason that TWTAs are advantageous for a channelized transponder payload.

Figure 7 also includes a multiple beam transmit antenna farm with TWTAs. De-mod-Re-mod processors and payloads implementation are currently limited to available space qualified technologies. The power required for many proposed Ka-band multibeam, multichannel payloads is driven by the power required for the IF/baseband processors, de-modulators, and re-modulators. Due to important hardware quantities, technology enhancements that allow significant reduction in power of these components will be required before this type of payload come to widespread use. The state of all the other payload technology is similar to that described for channelized transponder payloads.

Implementation Summary

Channelized Transponder Processor

The channelized transponder payload can be implemented with analog filtering and switching or digital filtering and switching. The applications suited to these implementations potentially can be allocated based on bandwidth. Analog implementations may be more appropriate for channel bandwidths greater than 1 or 2 MHz. Digital implementations are generally more adequate for channel bandwidths below 1 or 2 MHz or applications where multiple or variable bandwidth are required. Processor power and thermal dissipation definitely are a spacecraft implementation dominant trade parameter.

De-mod-Re-mod Processor

Two De-mod-Re-mod implementations are discussed. Both are quite viable options. The asymmetric implementation includes fewer uplink beams and demodulators than downlink beams and modulators. Applications for this architecture could most likely include data distribution networks. The symmetric implementation includes an equal number of uplink beams and demodulators and downlink beams and modulators. Applications for this architecture include advanced Internet applications or networks that require symmetric data services. Again the processor power will be a dominant spacecraft implementation tradeoff consideration.

Conclusion

- Onboard switching system makes efficient use of a satellite communication network having onboard multibeam technology.
- Channelized transponder payload is adequate to Physical Transport Layer Services as a result of its simplicity.
- A demodulation-remodulation processor payload is more adapted to the Network Layer Services thanks to its flexibility and reconfigurability.
- Processor power and thermal dissipation are among the most dominant trade-off parameters for payload implementation on the spacecraft.
- Evaluation of advanced satellite network architectures must be based on solid business plan before advanced architectures are accepted and implemented.

Satellite Communications Antenna Concepts and Engineering

Takashi Iida

Contents

T. Iida (✉)
Tokyo Metropolitan University, Hachioji, Tokyo, Japan
e-mail: QZF04134@nifty.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_18

Abstract

The most critical component of a communication satellite is its antenna system. The purpose of this chapter is to show how the antenna works by describing the basic concepts related to satellite antenna pattern, side lobe, gain, and polarization so that satellite antenna systems are designed and engineered to meet specific requirements. First, the fundamental parameters such as antenna pattern, beamwidth, radiation power, gain, and polarization are introduced. Second, basic antenna such as linear wire antenna, horn antenna, reflector antenna, and microstrip antenna is described. Third, array antenna for scanning and hopping beams is described for its function, gain, and phased array. In the fourth, multibeam antenna is described in terms of its function and type. Finally, an antenna for optical communications system is introduced briefly.

Keywords

Antenna gain • Antenna pattern • Array antenna • Cassegrain antenna • Circular polarization • Effective area • Half-power beamwidth • Helical antenna • Horn antenna • Linear polarization • Main beam • Microstrip antenna • Multibeam antenna • Parabolic antenna • Phased array antenna • Polarization • Reflector antenna • Side lobe

Introduction

The most critical component of a communication satellite is its antenna system and supporting electronics. Without this communications subsystem, the satellite cannot fulfill its mission. The design of this system ultimately determines the communications capacity of the satellite. The satellite antennas and communications also determine the types of ground antennas or user devices that can access the satellite and the telecommunications services that can be provided. This chapter explains the fundamental concepts on which satellite antenna systems are designed and engineered to meet specific requirements. As spacecraft platforms have matured to allow the deployment of larger and higher gain satellite antennas as well as to point them with greater accuracy, it has been possible to launch more and more capable space communications systems. The engineering fundamentals are presented in this chapter, and the next chapter explains the evolution of these more sophisticated antennas and the resulting capability to launch higher capacity satellites that can work to ever smaller ground antennas, including handheld satellite phones. After reading these two chapters, it is useful to also read the chapters on ground antenna systems that connect with the satellite antennas.

Fundamental Parameters

Antenna Pattern

The key to a satellite antenna is the pattern, or the concentrated, beam that is emitted.

Figure 1 shows a typical pattern of antenna. Figure 1a shows a three-dimensional image, and Fig. 1b indicates a cutting edge of the antenna pattern at an angle of "ϕ" whose vertical axis shows radiation strength at the horizontal axis of "θ." The

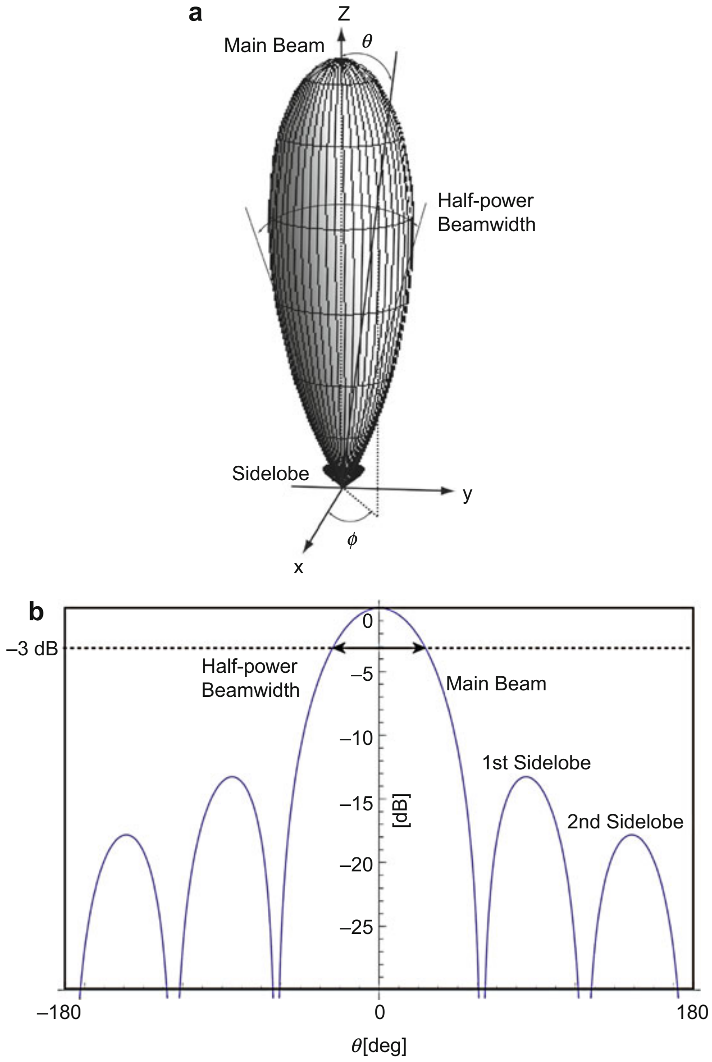


Fig. 1 Typical antenna pattern. (**a**) Three-dimensional image. (**b**) Two-dimensional image

direction of "z" shows a main beam and the directionality of antenna. The lobe other than the main beam is called a side lobe. The side lobe should be suppressed as much as possible because it may cause interference to or from other satellites and/or terrestrial communications system. In the case of an aperture antenna such as a parabolic antenna, if the aperture is illuminated by uniform amplitude and phase, the maximum gain is obtained for the given aperture, and the gain of the first side lobe is decreased by 13 dB from the gain of the main beam. This is to say that the first side lobe is 20 times less powerful than the main beam.

Beamwidth

The beamwidth is obtained from antenna pattern. Usually a half-power beamwidth is used, which is the angle of beamwidth at 3 dB decreased gain from the peak gain as shown in Fig. 1. The half-power beamwidth, $\theta_{1/2}$, of usual reflector-type antenna such as parabolic antenna is given approximately as

$$\theta_{1/2} = \alpha \frac{\lambda}{d} \text{ (deg.)} \tag{1}$$

where d is diameter of antenna and value of 65–70 is used as α.

Background Knowledge

Coordinate System

As noted earlier, a polar coordinate system (R, θ, ϕ) is used in the antenna technology usually. Figure 2 shows a polar coordinate system.

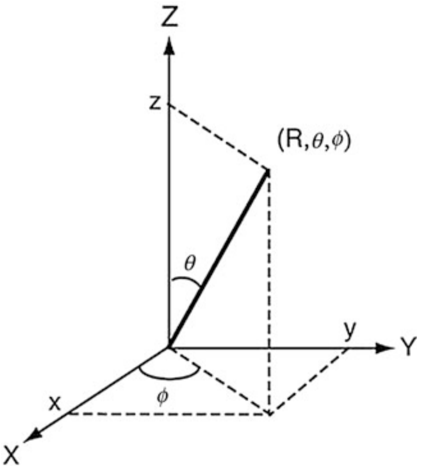


Fig. 2 Polar coordinate system

Solid Angle

When the area on a spherical body is *S*, the solid angle is defined by

$$\omega = \frac{S}{R^2} \tag{2}$$

where *R* is radius of sphere and whose unit is "steradian."

The relationship between a unit area and a unit solid angle is shown as follows: The unit area means $S = 1$in Fig. 3. This is indicated by solid angle as

$$\omega = \frac{1}{R^2} \tag{3}$$

This equation means the transformation of unit area to unit solid angle. The unit solid angle means $\omega = 1$ (steradian), that is, since

$$1 = \frac{S}{R^2} \tag{4}$$

the area per unit solid angle occupied on the surface of the sphere is given by

$$S = R^2 \tag{5}$$

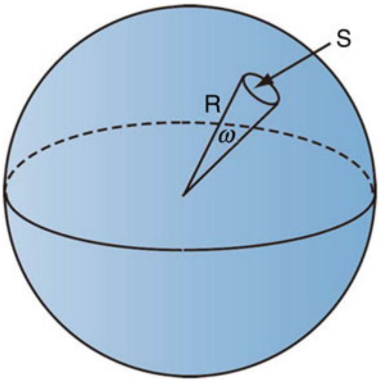
The solid angle of the entire sphere is, since the area of the sphere is $S = 4\pi R^2$,

$$\omega = 4\frac{\pi R^2}{R^2} = 4\pi \tag{6}$$

Radiation Direction

The electromagnetic field at the point (R, θ, ϕ) sufficiently far radiated by an arbitrary antenna positioned at the origin of the polar coordinate system in Fig. 2 is given by

Fig. 3 Solid angle



$$\left.\begin{aligned} E(R,\theta,\phi) &= \frac{e^{-jkR}}{R}\,U(\theta,\phi) \\ k &= \frac{2\pi}{\lambda} \end{aligned}\right\} \tag{7}$$

where λ is the wavelength and R is the distance.

$U(\theta, \phi)$ is called the direction function and is determined only by θ and ϕ independently of distance that is determined only by direction. It includes all the characteristics about the direction. Cutting $U(\theta, \phi)$by a plane and observing it is called antenna pattern.

Equation 7 is calculated from the famous Maxwell's equations that are the key basis for understanding electromagnetic transmissions. As for detailed solution from the Maxwell's equations, see Balanis (1997).

Radiation Power

The power flow on the surface of sphere with the sufficiently large radius of R is perpendicular to the spherical surface and directed to the outside. The power per unit area on the sphere, that is, power density, is given by

$$P(R,\theta,\phi) = \frac{(E(R,\theta,\phi))^2}{Z_0} = \frac{(U(\theta,\phi))^2}{Z_0R^2} \tag{8}$$

where Z_0 is performance characteristics of media which is called inherent impedance and is given in the vacuum by

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \cong 120\pi = 367.6[\Omega] \tag{9}$$

The power per unit solid angle, $F(\theta, \phi)$, is given by

$$F(\theta,\phi) = \frac{|U(\theta,\phi)|^2}{Z_0} \tag{10}$$

where $F(\theta, \phi)$ indicates the power strength to the direction of (θ, ϕ).

Gain

The antenna gain is defined as "a ratio of power per unit solid angle radiated to the arbitrary direction from the antenna to the power per unit solid angle radiated from the isotropic antenna derived by the same power as one of the antenna," where an isotropic antenna means an ideal nondirective antenna which radiates uniform

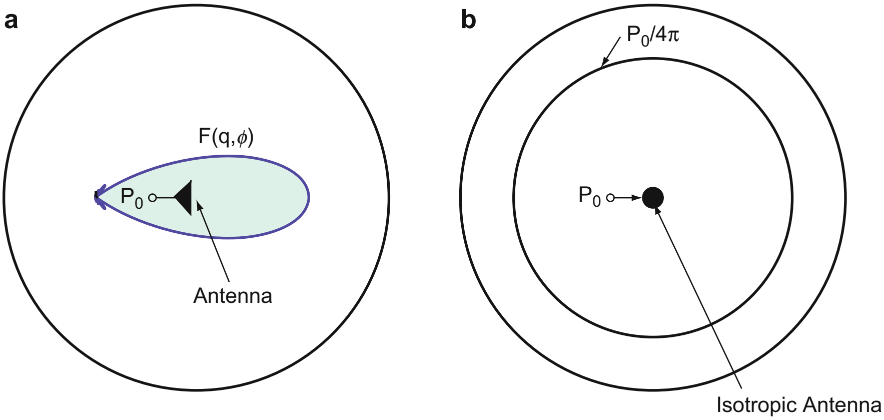


Fig. 4 Antenna gain. (**a**) Directive antenna. (**b**) Isotropic antenna

strength of electromagnetic field to every angle. The isotropic antenna is used as a reference of antenna gain. Given the power density $F(\theta, \phi)$ radiated from an antenna to the direction (θ, ϕ), gain $G(\theta, \phi)$ is given by

$$G(\theta,\phi) = \frac{F(\theta,\phi)}{P_0/4\pi} = 4\pi\frac{F(\theta,\phi)}{P_0} \tag{11}$$

where P_0 means a supplied power.

This can be considered as follows: When the power P_0 is supplied to the input terminal of antenna as shown in Fig. 4a, $F(\theta, \phi)$ means the power per unit solid angle of the radio wave radiated to the direction (θ, ϕ). Meanwhile, when the power P_0 is supplied to the input terminal of isotropic antenna with no loss as shown in Fig. 4b, all the power P_0 is radiated to the space because the isotropic antenna is no loss, that is, energy per unit solid angle is converted to $P_0/4\pi$, since the energy P_0 is radiated to every solid angles and this radiation is uniform for every direction. Therefore, according to the definition of gain mentioned above, Eq. 11 is a right expression. The expression of gain in Eq. 11 includes loss at antenna and it is also called power gain. The value of $G(\theta, \phi)$ at the maximum radiation direction (θ, ϕ) is called absolute gain and expressed in decibel [dBi], where the suffix “i” means isotropic antenna.

Reversibility of Antenna Characteristics

An antenna has the same characteristics of gain and pattern for using it as transmitting antenna and the reversibility is held.

Effective Area

Supposing that the power of incident electromagnetic wave per unit area is P and the effective area of antenna is $\mathrm{A_e}$ as shown in Fig. 5, the maximum power, W, taken from a receiving antenna is given by

$$W = P\mathrm{A_e} \tag{12}$$

The effective area $\mathrm{A_e}$ is given by using antenna gain G,

$$\mathrm{A_e} = \frac{\lambda^2}{4\pi} G \tag{13}$$

where λ is wavelength.

Examples of effective area of simple antenna are shown in Table 1. In addition, on the aperture antenna such as a horn antenna, ratio of actual aperture area A to effective area $\mathrm{A_e}$,

$$\eta = \frac{\mathrm{A_e}}{\mathrm{A}}, \tag{14}$$

is called the gain coefficient or aperture effectiveness. This is often also referred to as the efficiency of the antenna. From Eqs. 13 and 14, the antenna gain G having actual aperture area A is given by

$$G = \frac{4\pi}{\lambda^2} \eta A \tag{15}$$

Fig. 5 Effective area of antenna

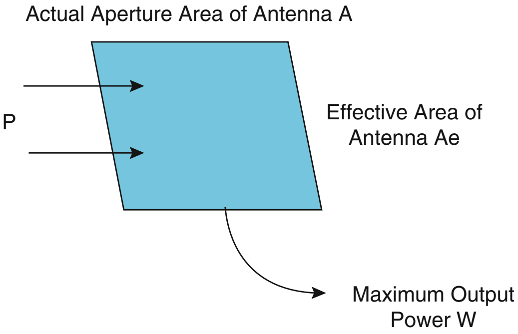


Table 1 Example of effective area of antenna

Antenna	Absolute gain	Effective area
Omnidirectional	0 dBi	$0.0796\ \lambda^2$
Half-wave dipole	2.15 dBi	$0.13\ \lambda^2$

If the actual aperture is a circle of diameter *d*, the antenna gain *G* can be given as

$$G = \left(\frac{\pi d}{\lambda}\right)^2 \eta \tag{16}$$

In the case of a parabolic antenna that is popular for satellite communication, the aperture effectiveness or efficiency is typically in the range of a low of 0.5 to a high of 0.7 or 0.75.

Polarization

The electric field and magnetic field of electromagnetic wave are directed toward a specific direction. This directional characteristic is called polarization. In the field of satellite communications, either linear or circular polarization is used to expand the effective use of the limited spectrum that is made available for satellite communications. By distinguishing between polarized signals by filters made for this purpose, one can reuse the same frequencies. Thus one can use horizontally and vertically "separated" signals to reuse the same spectrum, or one can distinguish between right-hand and left-hand circularly polarized signals.

Linear polarization involves discrimination between "wanted" and "unwanted" vertical and horizontal signals based on perpendicular coordinates. In this case, the electric field or wave form perpendicular to the ground is called vertical polarization. In contrast the electric field or wave form horizontal with the ground is called horizontal polarization. If a wave of horizontal polarization is combined with equal strength wave of vertical polarization by 90° phase difference, the combined wave has circular polarization whose electric field is rotated. Figure 6 shows the circular polarization. The circular polarization consists of right-hand circular polarization and left-hand circular polarization depending on its direction of rotation. As shown in Fig. 6b, the right-hand circular polarization is defined as that the rotation direction of electric field is clockwise looking from

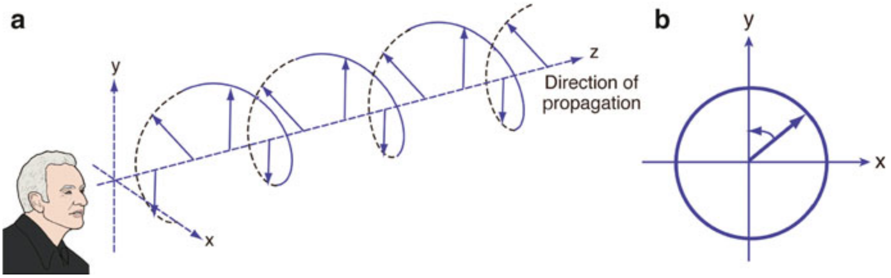


Fig. 6 Circular polarization. (**a**) Propagation of circular polarized wave plane (*Left circular polarization*). (**b**) Circulation of polarized looking from −z axis (*Right circular polarization*)

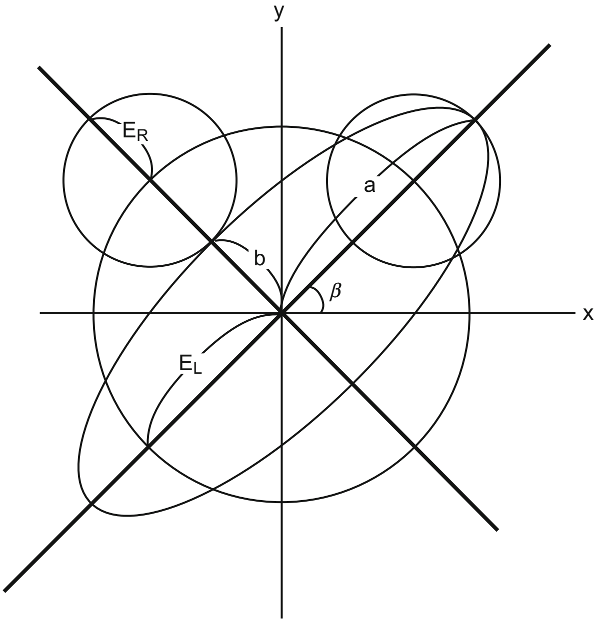


Fig. 7 Elliptical polarization

the opposite to the wave forward direction (−z); the left-hand circular polarization is defined as being vice versa. Figure 6a shows the left-hand circular polarization.

Generally speaking, since it is difficult to adjust the strength of electromagnetic waves to be equal or to adjust the phase to be 90°, the circular polarized wave has elliptical polarization. As shown in Fig. 7, the elliptical polarized wave can be decomposed into the component E_R of right circular polarization and component E_L of left circular polarization. These components can be used to define the following equations.

Ellipticity:

$$r = \frac{a}{b} = \frac{EL + ER}{EL - ER} = 20\log_{10}\left|\frac{EL + ER}{EL - ER}\right|[dB] \tag{17}$$

Circular polarization ratio:

$$\rho = \frac{E_L}{E_R} \tag{18}$$

Cross polarization discrimination:

$$XPD = 20\log_{10}|\rho| \tag{19}$$

where the relationship between r and ρ can be given as follows:

$$\left.\begin{aligned} r &= \frac{\rho+1}{\rho-1} \\ \rho &= \frac{r+1}{r-1} \end{aligned}\right\} \tag{20}$$

The ellipticity is also called axial ratio. The angle of major axis tilted from reference axis (X axis) is called tilt angle β as shown in Fig. 7.

Basic Antennas

Linear Wire Antenna

A dipole is a representative of a linear wire antenna and its structure is shown in Fig. 8. It is called a half-wavelength dipole whose length is $\lambda/2$. Its electric field pattern, E, is given by

$$E = E_0\ \cos\left((\pi/2)\cos\theta\right)/\sin\theta \tag{21}$$

where E_0 is the maximum strength of electric field. Figure 9 shows the antenna radiation power pattern, E^2, of half-wavelength dipole displayed in three-dimensional indication. Its gain, G_d, is given by

$$G_d = \frac{1}{\displaystyle\int_0^{\pi/2} \frac{\cos^2\left(\frac{\pi}{2}\ \cos\theta\right)}{\sin\theta}\, d\theta} = 1.64 (= 2.15dB) \tag{22}$$

The half-power beamwidth is 78°. It is an omnidirectional pattern whose gain is equal in all directions when it is taken of slice of the pattern with a plane normal to the dipole.

Fig. 8 Structure of dipole antenna

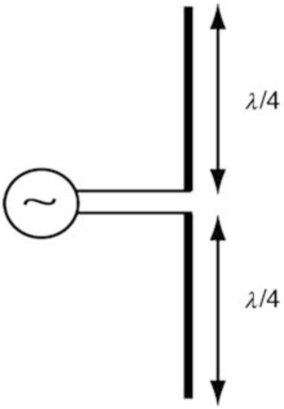
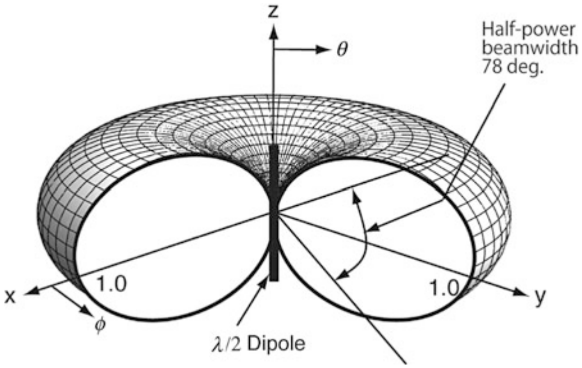


Fig. 9 Antenna pattern of half-wavelength dipole (three-dimensional image)



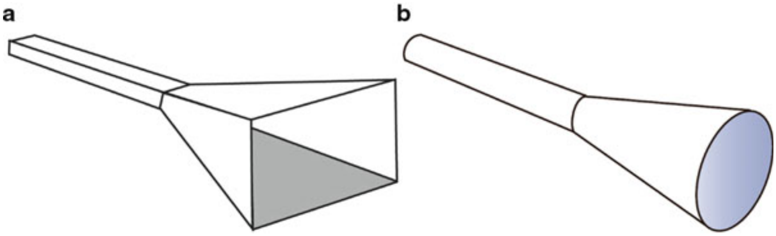


Fig. 10 Horn antenna. (**a**) Pyramid horn antenna. (**b**) Conic horn antenna

Horn Antenna

A horn antenna is often used as a primary emission element of a reflector-type antenna. In addition, it is used as an antenna itself when wide beamwidth is necessary. The horn antenna is often used as satellite-borne antenna since it has a proper beamwidth for global coverage that intended to look at the whole Earth from a geostationary satellite whose lookup angle is about 18°.

The horn antenna is categorized roughly into two kinds: a pyramid horn that widened rectangular waveguide and a conic horn that widened circular waveguide as shown in Fig. 10. Since the theoretical gain of a horn antenna coincides well to actually measured gain and it is strong structurally, it is often used as a reference antenna for measuring the gain of various antennas in the microwave frequency.

Reflector Antenna

The representatives of reflector antenna are as follows:

- Parabolic antenna
- Cassegrain antenna

- Offset parabolic antenna
- Offset Cassegrain antenna

The reflector-type antenna is often used as satellite communication as well as broadcast satellite receive antenna. A reflector antenna realizes high gain and low side lobe by converting the spherical wave radiated by a primary radiator to the plane wave. It is for these reasons that reflector antennas have become the predominant type of antenna used in satellite communications.

Parabolic Antenna

Structure of a parabolic antenna consists of a parabolic reflecting surface and a primary feed at a focus as shown in Fig. 11. Since the parabolic surface is a paraboloid of rotation, sum $(l_1 + l_2)$ of the distance l_1 from a focus to the reflector and distance l_2 from the plane normal to the antenna axis to reflecting point is constant; the spherical wave radiated at a primary feed put at focus is converted to the plane wave at the reflector. A horn antenna is used for the primary emission device. The gain of parabolic antenna has been previously given by Eq. 16.

Cassegrain Antenna

A Cassegrain antenna is a dual-reflector antenna which consists of a parabolic surface as a main reflector and a hyperboloid of revolution as a sub-reflector as shown in Fig. 12. Among two foci of a sub-reflector, one accords with the phase center of primary feed and another accords with a focus of the main reflector. The sub-reflector works as a converter of spherical waves for primary feed and main

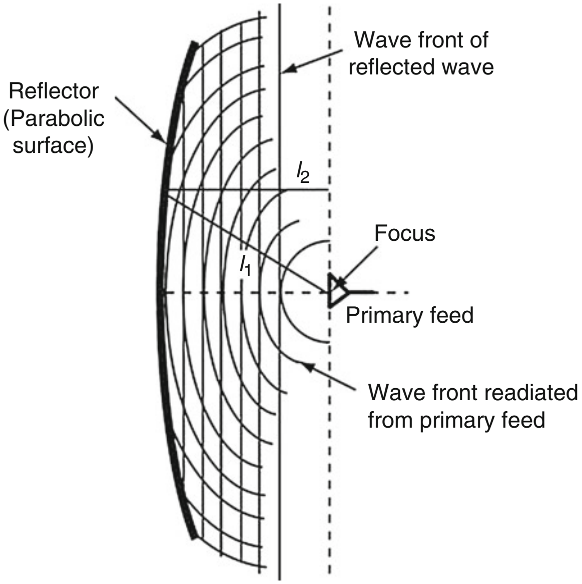


Fig. 11 Principle of parabolic antenna

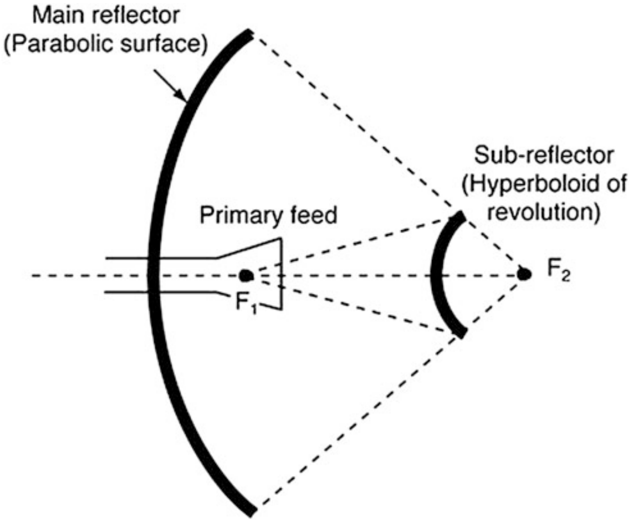


Fig. 12 Cassegrain antenna

reflector, and the main reflector is a converter between spherical waves and plane waves.

The characteristic of Cassegrain antenna that is different from a parabolic antenna is that it can accommodate a low-noise amplifier at the back of the main reflector. Also there is a margin to shorten the length of waveguide feed and thus to decrease the transmission loss. A Cassegrain antenna is used as an earth station antenna for satellite communications widely.

Offset Parabolic Antenna

Both parabolic antennas and Cassegrain antennas have a main reflector with a symmetrically rotated parabolic surface. This leads to the need for props supporting the primary feed and sub-reflector, and these must be positioned directly in front of the main reflector. This causes emission characteristic deterioration because of blocking the incoming and outgoing electric waves, increase of side lobes, and decrease of gain. In order to avoid these obstacles, an antenna that sets the primary feed and sub-reflector outside of the aperture is an offset parabolic antenna. This is accomplished by using only a part of a parabolic surface as a reflector.

An offset parabolic antenna is shown in Fig. 13. A low side lobe is possible to be established for this antenna because of its design architecture.

Helical Antenna

Figure 14 shows a helical antenna whose structure has a wounded conductor (helix conductor) in front of a reflector and a feeding point on the conductor. As for this

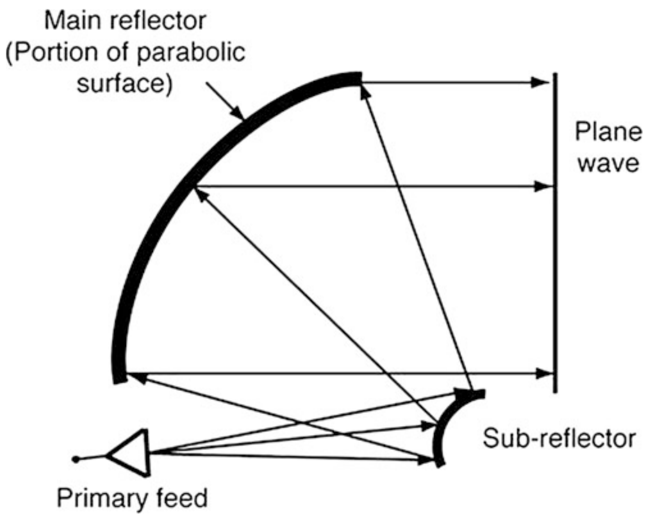


Fig. 13 Offset parabolic antenna

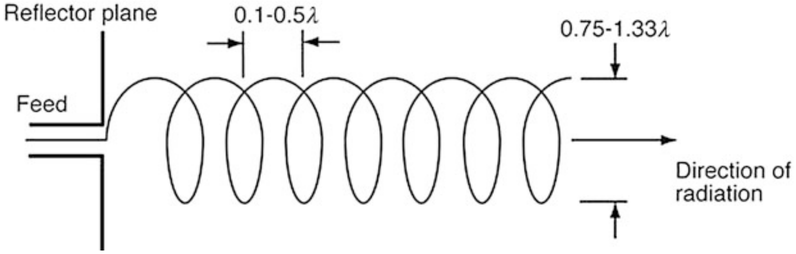


Fig. 14 Helical antenna

antenna, the electromagnetic wave is radiated in an axis direction when circumference length is 0.75–1.33 wavelength and pitch of spiral is 0.1–0.5 wavelength. When the reel number and the full length are increased, the antenna gain increases. In addition, when circumference length is small in comparison with a wavelength and the full length is around a wavelength, the electromagnetic wave is radiated in the direction perpendicular to the axis (i.e., an axis mode of operation).

Microstrip Antenna

A microstrip antenna has recently received a good deal of attention primarily in terms of user antennas. This approach has been applied to antennas for automobiles, receiving antenna of broadcast satellites, and aircraft-borne antenna.

Characteristics

The characteristics of the microstrip antenna are as follows:

- Thin structure
- Lightweight
- Simple structure
- Easy productivity
- Easy accumulation with semiconductor circuits
- Comparatively high gain for a simple antenna (about 7 dBi)

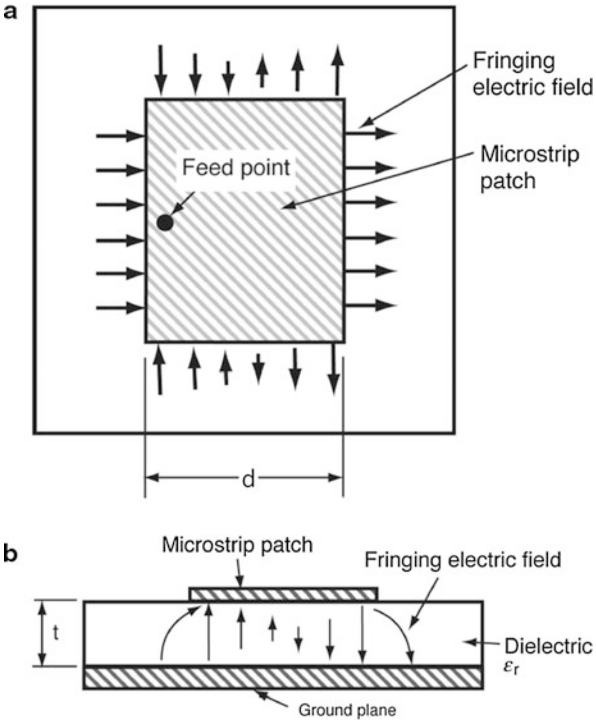
This antenna has the possibility to become widely used.

The gain of a microstrip antenna generally increases with the decrease of the dielectric constant. A deficiency of microstrip antennas, however, is the narrow frequency bandwidth it affords. So currently research is being directed at increasing the effective frequency bandwidth by various methods.

Rectangular Microstrip Antenna

It is generally considered that the resonance device formed as a rectangular-shaped open-type plane circuit on a thin dielectric substrate has low Q of resonance due to a loss of emission as shown in Fig. 15. The antenna that used this emission loss

Fig. 15 Rectangular microstrip antenna. (**a**) *Front* view. (**b**) *Side* view



positively is a microstrip antenna. Structure of resonance device is suggested with a circle, a triangle, and a pentagon other than the rectangular mentioned above in various ways.

The working principle of the antenna is shown in Fig. 15. The electromagnetic wave is emitted by leaking of an electric field formed at the border of the microstrip antenna. For example, radiation to the front direction of antenna is conducted by the leaking of an electric field at the right and left side of an antenna element. The leaking electric field at the top and bottom side of the antenna does not contribute to the radiation due to drowning out each other.

A resonance frequency is approximately given by

$$f = \frac{c}{2(d + t/2)\sqrt{\varepsilon r}} \tag{23}$$

where c is velocity of light, d is given in Fig. 15a, t is the thickness of the substrate, and ε_r is the dielectric constant of substrate.

Circular Microstrip Antenna

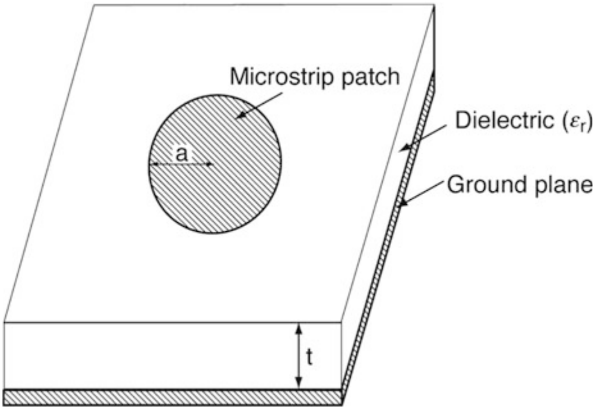
The structure of the circular microstrip antenna is shown in Fig. 16. The resonance frequency is given approximately by

$$f = \frac{1.841c}{2\pi\{a + (t/\pi)2\ln 2\}\sqrt{\varepsilon}} \tag{24}$$

where c is light velocity (a: see Fig. 16), t is thickness of substrate, and ε_r is dielectric constant of substrate.

The maximum gain is obtained at the front direction and a single direction pattern. As for a feeding system of microstrip antenna, there are microstrip feeding, pin feeding, and their combination (Iida 2000).

Fig. 16 Circular microstrip antenna



Array Antennas for Scanning and Hopping Beams

Function of Array Antenna

The antenna system which deploys an “array” of the same type of antennas is called an array antenna. In this case, an arranged antenna in the array is called an antenna element. The array antenna can have various kinds of antenna functions that cannot be conducted by a single emission element. The performance of an array antenna system depends on the kind of antenna elements, the arrangement method, and the way radiation is accomplished. The directivity of an array antenna is important. The particular method that is created to optimize directivity performance is called the directivity composition. When the combined directivity of the antenna elements is set for the array, the combined directivity is generally given by

$$\begin{aligned}(\text{Combined directivity}) = {} & (\text{Directivity of antenna element}) \\ & \times (\text{Arrangement directivity of omnidirectional antenna})\end{aligned} \tag{25}$$

This is one of the biggest functions of an array antenna. The array antenna with the following function is obtained by changing drive amplitude and phase of each antenna element of an array antenna:

- To get desired directivity
- To change width of main beam of emission directivity
- To suppress side lobes and to control their level
- To specify zero points of the emission directivity
- To get a desired gain

Another characteristic of an array antenna is possible to scan the main beam of emission directivity and do so three-dimensionally. Fixing an arrangement of an antenna element, the main beam can be pointed to an arbitrary direction of space by changing the driving phase of each antenna element. This is called a phased array antenna.

Directivity of Array Antenna

In an array antenna, all of the elements are excited simultaneously by dividing the feeding power. In Fig. 17, the directivity of array antenna $D(\theta, \phi)$ is given when N antenna elements arranged by equal space of d are derived by amplitude I_n and phase ϕ_n,

$$D(\theta,\phi) = g(\theta,\phi)\sum_{n=1}^{N} I_n e^{j\{\phi_n+(n-1)kd\ \sin\theta(\cos\phi+\sin\phi)\}}, \tag{26}$$

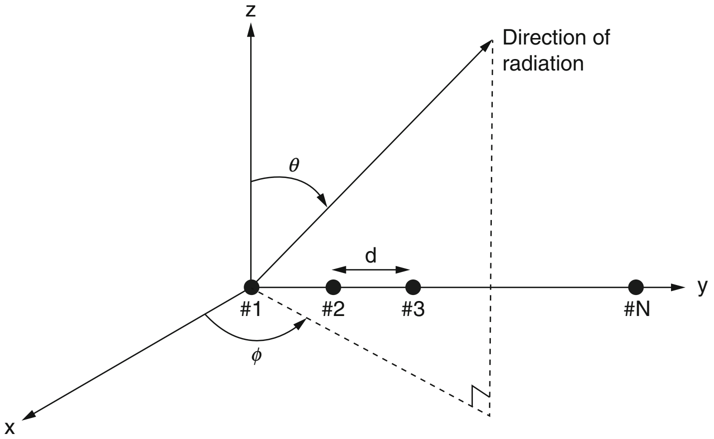


Fig. 17 Coordinate system of array antenna

where $k = 2\pi/\lambda$ and $g(\theta, \phi)$ is directivity of a single antenna element. This $g(\theta, \phi)$ depends on the antenna element. But the term of Σ is same, namely, the term of Σ indicates the characteristics of array antenna. This is called array factor.

Gain of Array Antenna

It is often used that many antenna elements are arranged to increase a gain of an array antenna. Supposing that there is no interaction at all between antenna elements, the gain of an array antenna, G_n, is n times of gain of an antenna element, G_e, in the case where the number of the antenna elements is n:

$$\frac{G_n}{G_e} = n \tag{27}$$

In other words, the gain increases in proportion to the number of antenna elements. However, actually, since there is mutual combination between antenna elements, the gain does not become n times. But Eq. 27 is used as an aim of a gain of array antenna as well.

Phased Array Antenna

Function of Phased Array Antenna

A phased array antenna is the antenna which changes feeding phase of each antenna element electronically and can scan the main beam of emission directivity three-dimensionally. In mobile satellite communications, it is used as an antenna of a vehicular side on the ground. The main beam changed its direction according to the

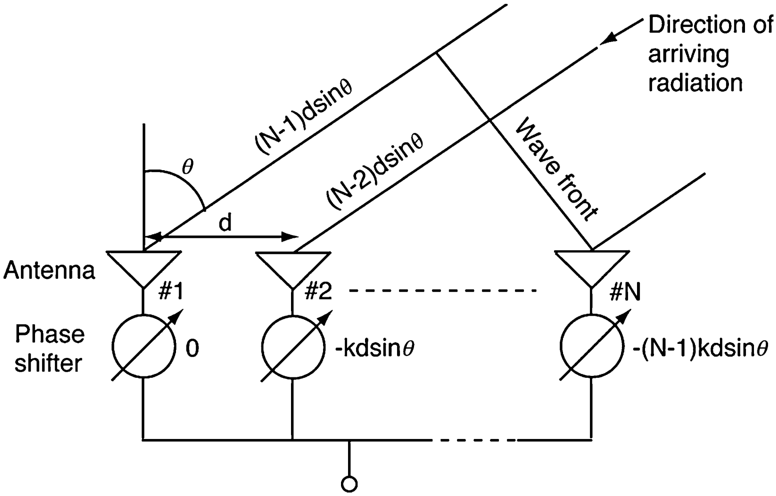


Fig. 18 Configuration of phased array antenna

direction of satellite to prevent the link from disconnecting when a position of a vehicle changed. It is also used as an antenna of the geostationary satellite (a relay satellite) side for inter-satellite communications.

Configuration of Phased Array Antenna

A configuration of phased array antenna is shown in Fig. 18. Phase-shifting device is installed at every antenna element, and it is controlled from its outside. The operation of the phased array antenna is explained for the reception of signal.

The equiphase plane (wave front) of incident wave from a certain direction at the Nth (#N) antenna element propagates a distance $(N-1)d\sin\theta$ from the wave front to the first (#1) antenna element. The phase of signal received at #N element advances $k(N-1)d\sin\theta$ during this propagation, where $k=2\pi/\lambda_g$ and λ_g is wavelengths in a feeding circuit. The signal at the #N element is delayed by a phase shifter. Namely, the phase shifter is set to $-k(N-1)d\sin\theta$ so that the phase difference with the signal received by the #1 element is zero. By such a processing operation the signal can be strengthened by calculating the optimum output.

Multibeam Antennas with Multiple Feed Systems

Function of Multibeam Antenna

A multibeam antenna is an antenna that has plural beams and plural input and output terminals to be able to transmit plural independent information as shown in Fig. 19. It enables increase of a gain by a spot of the beam, frequency reuse by the space division of the beam. Let us consider why a multibeam antenna is necessary for a

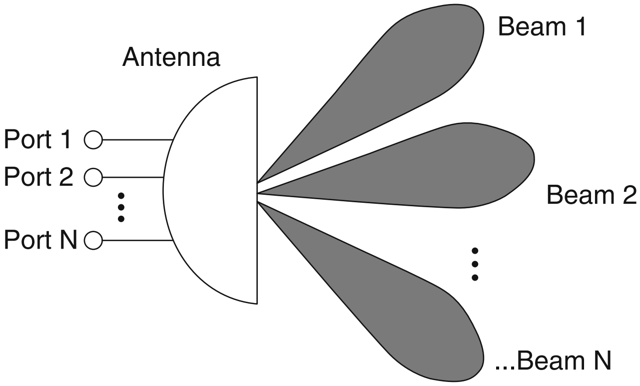


Fig. 19 Concept of multibeam antenna

mobile satellite communication, although it can also be quite effectively used for fixed satellite communications as well.

- Since it is difficult for a mobile vehicle to install a large-sized antenna and/or high-powered transmitter generally, an antenna of a big gain is necessary at the satellite side, but the aperture area A_e must be big to raise gain G from the equation indicated between gain and aperture area, Eq. 13. As for the antenna, the bigger the aperture area A_e is, the narrower the beamwidth, and it becomes spot beams from the relationship between aperture diameter d and beamwidth Eq. 16. In the case of mobile satellite communications, one needs not only high-powered spot beams to communicate with small user terminals, but also a lot of them in order to cover a wide service area. Thus, the solution is a very large aperture antenna with a multibeam feed system.
- A demand of many spot beams is satisfied by preparing for many single spot beam antennas. But many large antennas cannot be embarked on a satellite due to constraint of both weight and space; thus the multibeam antenna which can emit an independent spot beam of a plural number from an antenna is necessary.

Type of Multibeam Antenna

The multibeam antenna is categorized into the following:

- Reflector type
- Array type
- Reflector + array type

Reflector Type

In this case, multiple primary feeds (usually horn antennas) are installed in the neighborhood of the reflector's focus. This reflector-type antenna architecture allows high performance for few beams and with a simple structure. But when the number of primary feeds increases, a gap between the primary focus grows too wide, and the performance characteristics deteriorate. For this type of satellite antenna and feed system, offset parabolic antenna and offset Cassegrain antenna are employed.

Array Type

An array antenna has a characteristic to be able to operate with a high degree of directivity. This is accomplished by changing the placement and phase of each element.

Reflector + Array Type

This is a multibeam antenna that deploys a conventional parabolic reflector but then uses an array-type multibeam antenna at a focus as primary feed. There is no deterioration of each beam with this type of antenna, and this type of design can serve to make the feeding circuit loss small. This can also be the most cost-effective solution as well.

Antennas for Optical Communications Systems

It is necessary to perform acquisition, tracking, and pointing (ATP) to establish optical satellite communications (ATR 1995). The ATP optics includes an optical antenna (a telescope), an acquisition sensor, but also a rough coarse tracking sensor and a precise fine tracking sensor, and the mechanical elements are two-axis gimbals to control the antenna pointing direction drive device and a fast steering mirror actuator. These sensors and actuators form a dynamic system to assure link stability in transmitting and receiving a very narrow optical beam.

Figure 20 shows a Cassegrain telescope as a typical optical antenna. Usually, a primary mirror is constructed with a paraboloid shape and a secondary mirror with a

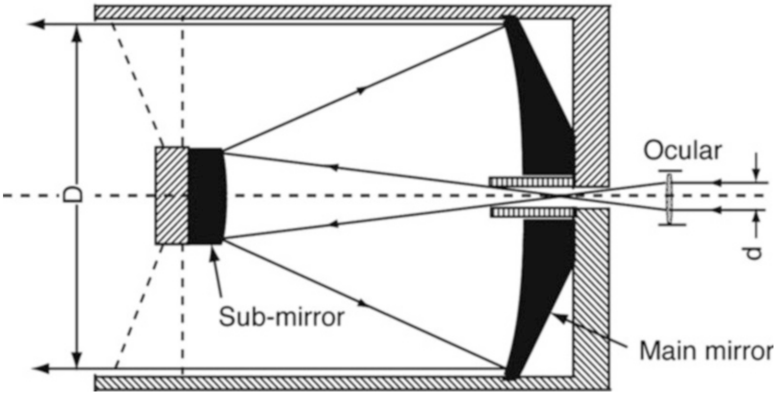


Fig. 20 Optical Cassegrain antenna

hyperboloid shape, and the laser light is led to an internal optical system through a hole which is made in the center of the primary mirror. If the aperture diameter of the main primary mirror (antenna) is D and the diameter of the internal beam is d and the magnification is defined as $M = D/d$, the precision requirement for the internal optical components and their alignment accuracy can be relaxed at the internal optics, because the angle deviation at the part of antenna aperture is magnified by M. However, the precision of lower than 1 μm is still necessary for the optical antenna to be used for inter-satellite links. Therefore, a material with low coefficient of thermal expansion coefficient, such as Invar (iron-nickel alloy) or Zerodur (glass ceramic composite material), has to be used.

The bigger the aperture diameter of antenna is, the bigger the antenna gain will be, but it is necessary to be careful in tracking accuracy accordingly because an antenna pattern becomes sharp. For example, a tracking accuracy of less than 1 μrad is necessary so that an antenna gain of 30 cm in diameter is made available. In addition, the transmitting laser beam can be approximated as a Gaussian beam that has a flat wavefront and a Gaussian amplitude distribution. This Gaussian beam degrades its energy/power because the outskirts of the beam are truncated by the finite primary mirror size or a body tube and the beam center is obstructed by the secondary mirror. This truncation and obstruction will limit the aperture efficiency of the antenna. In addition, wavefront error due to optical aberration or mirror surface inaccuracies will cause another degradation which is expressed with the quantity of the "Strehl ratio." It is necessary to increase or maintain the wavefront error less than 1/10 wavelength to keep the Strehl ratio better than -1.7 dB.

Conclusion

It is clear from the above descriptions of satellite antennas that there are a wide range of antenna types that can be used in satellite communications. Despite this diversity of antenna designs, parabolic reflectors with different types of feed systems, including phased array feed systems, are the most common satellite antenna systems today. This is because of the ability to achieve higher capacity, higher gain beams, and lower powered side lobes that create interference to other satellites and ground-based communications systems. Broadcast systems generally need more power to work to small dishes. Mobile satellite systems have evolved higher and higher spot beams using multibeam antenna technology. Fixed-satellite service satellites have employed similar satellite antenna technology as well.

Currently the greatest challenge in satellite antenna design is not the satellite antenna design itself, but in the design of maneuvering systems that allow the antennas on board medium Earth orbit (MEO) and low Earth orbit (LEO) satellites in constellations to avoid direct interference with GEO satellites and their associated ground stations as they pass through the equatorial orbital arc.

In the next chapter, the implementation of the antenna technology by satellite communications around the world will be addressed in more detail.

Cross-References

► Satellite Antenna Systems Design and Implementation Around the World

References

ATR, *Optical Inter-satellite Communication* (Ohmsha, Tokyo, 1995). in Japanese

C.A. Balanis, *Antenna Theory – Analysis and Design* (Wiley, New york, 1997)

T. Iida, *Satellite Communications – System and Its Design Technology* (Ohmsha/IOS Press, Tokyo/Amsterdam, 2000)

Satellite Antenna Systems Design and Implementation Around the World

Takashi Iida and Joseph N. Pelton

Contents

T. Iida (✉)
Tokyo Metropolitan University Hachioji, Tokyo, Japan
e-mail: QZF04134@nifty.com

J.N. Pelton
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_19

Abstract

In this chapter, the objective is to discuss the practical implementation of various types of satellite antenna designs over time and to indicate the current state of the art and future trends to develop even higher gain satellite antennas with greater efficiencies in terms of frequency reuse or higher capacity FSS or MSS type satellite systems. Although there continue to be smaller satellites that are launched for communications purposes, the antenna designs utilize the same technologies and concepts that are employed in larger scale satellites.

The evolution of antennas for satellite communications has generally conformed to the following historical pattern:

Low gain omni- and squinted-beam antennas

- Increased gain types of satellite antennas (horn type and helix antennas)
- Parabolic reflectors (including multibeam antennas with multiple feed systems)
- Deployable antennas (particularly for achieving more highly focused beams and support much high-gain multibeam antennas
- Phased array feed and phased array antennas
- Scanning and hopping beams
- Optical communications systems (initially for intersatellite links and interplanetary communications, but this type of technology might possibly be used for Earth to space systems in the future as well).

Examples of many of these types of satellite antennas will be presented in the following chapter. But first, the factors that have led engineers to design improved and higher performance antennas will be discussed and examined.

Keywords

Deployable Antenna • High-Gain Antenna • Horn Antenna • Isotropic Antenna • Mobile Satellite Services (MSS) • MSS with Ancillary Terrestrial Component (ATC) • Multibeam Antennas • Multifeed Systems • Off-Set Feed Antenna • Omni Antenna • Orbits–GEO, MEO, and LEO • Parabolic Reflector • Path Loss • Phased Array Antenna • Polarization • Three-Axis Body and Spin Stabilized Spacecraft

Introduction

In the preceding chapter the basic concepts related to satellite antenna patterns, interfering sidelobe transmissions, gain, linear and circular polarization, as well as different types of antennas with improved efficiency of performance were discussed. The discussion explained how satellite antenna designs have evolved to be more effective in producing higher gain, more effective reuse of available frequencies, etc. In this chapter, the objective is to discuss the practical implementation of these different types of satellite antenna designs over time and to indicate the current state of the art and future trends to develop even higher gain satellite antennas with greater efficiencies in terms of frequency reuse or higher capacity FSS (fixed satellite services) or MSS (mobile satellite services) type satellite systems.

Design Factors Driving the Development of Satellite Communications Antenna Systems

There are many factors that drive the development of new and improved satellite antennas. These include the need to reuse frequency bands because of limited spectrum allocations, the need to have antennas that can operate at higher frequencies with higher bandwidth, and the desire to deploy higher gain antennas while minimizing the required mass and volume.

The Need to Effectively Achieve More Useable Spectrum

Perhaps the prime design factor that has driven the research and development to design improved satellite antennas has been the need to make more efficient use of the available allocated spectrum for satellite communications. Today, fiber optic networks within closed cables have access to almost unlimited spectrum within the optical frequencies that have very broad spectra available. Satellites, which must transmit their signals within the open environment and largely within the UHF, microwave, and millimeter wave rf bands have much more limited spectra to utilize. Thus, there is a continuing need to develop satellite antenna technology that can use the available spectrum ever more efficiently. This primarily means finding ways to reuse the same spectrum many times over where ever possible.

Frequency Reuse Concepts

As just noted, the radio frequencies available for satellite communications are very limited resources. Effective utilization of this sparse resource, namely the band allocated to various satellite services, continues to be very critical to meet every rising service needs. There are fortunately now a number of frequency reuse technologies made available via antenna-related technologies that allow multiple types of frequency reuse in many types of communications satellites. Frequency reuse means to use the same frequency repeatedly. Reuse of the same frequency, however, usually causes interference to others trying to reuse the frequencies as well. If the same frequency is used in geographic areas that are sufficiently removed from one another, then such reuse becomes technically possible. Another method, separate from spatial separation, is to use different types of polarization techniques to separate the signals. This is, in a way, similar to the techniques used to separate “wanted and unwanted” incoming light such is accomplished by wearing polarized sun glasses. The prime methods of frequency reuse are thus either to separate antenna beams spatially (i.e., to transmit narrow spot beams that use the same

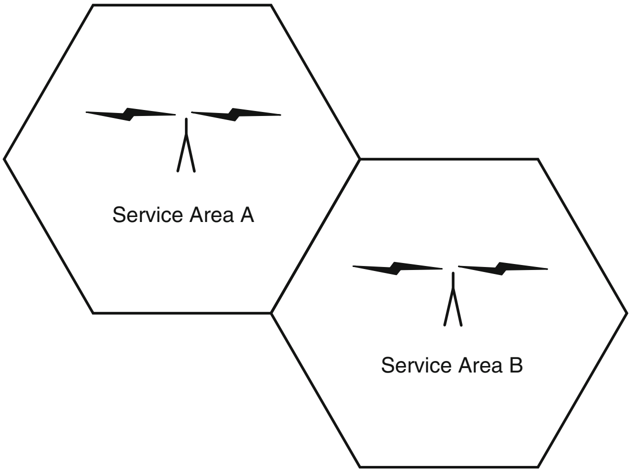


Fig. 1 Frequency reuse patterns using geographic separation of cells

frequencies to transmit signals to different geographic areas that are widely separate from each other) or to utilize orthogonal polarization of electromagnetic waves.

If a certain service area A and another service area B are located separately in the case of a terrestrial cellular communication system as shown in Fig. 1, the same frequency can be used over again.

The number of times that the same frequency can be effectively utilized can be increased with improved digital technology. This can be accomplished by dividing the cells into ever smaller service areas. In the case of satellite communications, the service area cannot be decreased to areas as small as the cells used in the terrestrial communication systems. This is because the service area is determined by antenna beam width of satellite. Since the satellite is in Earth orbit well away from the ground, the satellite beams spread much further than a terrestrial cellular beam.

However, the number of frequency reuses can be increased by squeezing the satellite beam width as small as possible. In this case, CIR (carrier to interference ratio), namely a ratio of interference power to carrier power, is a criterion for determining the degree of frequency reuse that can be reasonably obtained. If CIR is maintained more than 10 dB (i.e., the wanted signal is 10 times "clearer" than the unwanted signal), practical use can be possible, especially in the case of digital communication that are more tolerant of interference levels. If the CIR is 30 dB, the wanted signal is 1,000 times clearer.

The second method that can be used is, of course, the use of polarization. In this respect, there are two options – linear or circular polarization. In the case of linear or orthogonal polarization, the polarizer creates one signal that is irradiated in a horizontal plane and the other signal is irradiated in the vertical plane so the two signals can be clearly differentiated from one another when received. The other polarization alternative is the case of right hand and left hand circular polarization where the two signals are distinguished from one another by either rotating in right hand direction or rotating in the opposite left hand circular direction. CIR discrimination, from 27 to 35 dB, can be obtained by using polarization. Orthogonal discrimination is less expensive but does not provide quite the same CIR discrimination as circular polarization. The ratio of a component of the polarization to its opposite one is called XPD (cross polarization discrimination), and it is usually expressed in decibel (dB). The degradation of XPD due to rain fall must be considered in devising a communication link design (Iida 2000). In practical application, XPD must be within a regular value, namely the direction of antenna and polarization must be adjusted. This might be as much as 27 dB for a VSAT (very small aperture terminal) and perhaps more than 35 dB for the other larger and higher performance earth stations.

The Migration from Lower Frequencies Bands to Higher Frequencies VHF, UHF, SHF, and Now EHF

The spectra used for communication satellites have persistently migrated upward many orders of magnitudes in terms of the type of radio frequencies used for services over the past half century. Frequencies have increased from very high frequencies (VHF), to ultra-high frequencies (UHF), to super-high frequencies (SHF), and now to extremely high frequencies (EHF). The amount of spectra that is available in the very high frequency (VHF) and ultra-high frequency (UHF) bands is only a few megahertz. This is simply due to the laws of physics. There is less bandwidth available at these lower frequencies. Further, there are also many other competing uses for these bands from terrestrial services such as mobile communication via cellular networks as well as for radio and television. This competing use leads to problems of interference.

The solution for satellite communications has been to migrate upward to the higher frequency allocations in the microwave and millimeter wave bands. In the super high frequency (SHF) band, that is, from 3 gigahertz (GHz) to 30 gigahertz (GHz) and the extremely high frequency band, that is, from 30 to 300 GHz, there are many broadband frequency allocations of “rf” spectra for satellite communications. There is a 500 MHz allocation in the C band (6 GHz uplink and 4 GHz downlink) for fixed satellite services (FSS), plus another allocation in X-band for defense related communications (i.e., 8/7 GHz), plus another 500 MHz in the Ku band (14/12 GHz). Further, there is a very wide allocation of 1,000 MHz allocation in the Ka band (30/20 GHz). There is also a major allocation for the downlinking of direct broadcast satellite services at 18 GHz. For the future, there are also significant additional allocations in the Q, V, and, W bands at 38, 48, and 60 GHz. For mobile satellite services, there are smaller spectrum allocations in the 800 MHz band, plus allocations around in the 1,600/1,700 MHz bands, in the 2,000/2,100 MHz range, and then another allocation in the 2,500 MHz band. In addition, there is the ability to use the aforementioned FSS allocations as feeder links to uplink signals from MSS major earth station feeder networks. A detail listing of the bands that are utilized for FSS, MSS, and BSS services are provided earlier in this handbook.

The question that might spring to mind is why not just go to the wide frequency bands at the highest frequencies in the EHF bands if there are very wide bands of spectra available there? The problem is there are technical, operational, and financial reasons that argue against such a migration. The lower frequencies require lower cost and easier to manufacture equipment. Further transmissions are more tolerant of various types of interference and obstacles blocking the transmission path since the longer wave lengths do not have to be direct line of sight to be received. Also at the higher frequencies, rain attenuation and other atmospheric attenuation factors make it difficult to send signals when there are adverse weather conditions.

The migration from lower frequencies to higher frequencies has many specific difficulties associated with transitioning to the higher bandwidth allocations. There is the advantage of broader spectrum allocations on one hand but all the other factors in such a transition are adverse to moving up to higher spectra. This is to say that ground and satellite antenna systems and associated electronics are more difficult and expensive to build, rain attenuation can block signals, and any physical interference (such as can be anticipated in mobile communications satellite systems) will tend to block the signal.

In addition to all these considerations, there is the issue of required signal power at higher frequencies. The received signal power "*C*" that a satellite antenna collects through free-space propagation is given by using antenna gain formula as follows (Demers et al. 2009):

$$C = P_T G_T G_R \left(\frac{\lambda}{4\pi d} \right)^2$$

Formula for received signal power C

C is the received carrier power, P_T represents transmitted power, G_T represents antenna gain of the transmitter, G_R represents the antenna gain of the receiver, λ represents the wave length, and d represents the distance between the transmitter and the receiver. The antenna gain G is given by:

$$G = \eta \left(\frac{\pi}{D\lambda} \right)^2$$

Formula for antenna gain

In this formula, G represents antenna gain, η stands for efficiency, D represents the aperture of antenna, and λ represents wavelength.

As can be seen from these equation, the shorter the wave length and thus the higher the frequency, the larger the need for the received signal power, assuming the antenna aperture is constant. Additional power leads to additional design complexity and added cost to the satellite as well as increased launch costs.

In summary, if higher frequencies are to be used for satellite communications, there are the following challenges that must be faced:

- There is the need to develop totally new devices to operate at the very demanding frequencies and increasing small wavelengths.
- There accordingly tends to be an overall increase of cost in general for ground and space systems.

- At sufficiently high frequencies there will also be an increase in rain attenuation and other atmosphere degradations in signal strength.
- Due to rain attenuation and other such atmospheric phenomena, there will also be a need for higher link margin to maintain quality and reliability of service. The proportionate transmit power is also a factor in this regard.
- There is a considerable technical challenge of not only designing but manufacturing low loss antenna systems. Particular challenges include:
  - Necessity of increasing the surface accuracy of antenna
  - The need to have very small mesh in case of mesh structure that is exactly conformal to the parabolic shape to direct the signal to and from the feed system

It needs to be noted that these various challenges increase as one moves to higher frequencies not in a linear fashion but rather exponentially. Despite these difficulties, a great deal of progress has been made through R&D and experimental satellite programs so that operational satellites in the Ka band are now being implemented. Research and development over the past 10 years has now made it possible to design functional fine mesh deployable antenna for commercial use up through to even the Ka band (Demers et al. 2009).

Techniques for Improvement of Satellite Throughput

In light of the difficulties of going to the higher frequencies, a good deal of attention has been focused on ways to use the lower frequencies more efficiently for satellite communication. This literally means increasing the throughput of digital bits per available hertz within the lower band frequency allocations. Not long ago, a typical communications satellite would transmit about 1 bit per hertz. With improved modems and encoding (i.e., coder and decoders known as codecs), it is possible to achieve transmission efficiencies on the order of 2–2.5 bits per hertz, and the very latest coding techniques have even increased performance up to the level of 4–5 bits per hertz. This is equivalent to expanding the spectrum allocation by a similar amount. As always, there are tradeoffs and technical difficulties to be overcome. Advanced encoding such as turbo-coding only works well in a relative low noise or interference environment. The use of intensive encoding such as 8 bit, 16 bit, or even 32 bit encoding can be successful in a low noise and essentially clear sky condition. Thus, some advanced encoding systems can be stair stepped down to lower throughput efficiencies in the case of heavy rain and higher levels of rain attenuation.[1]

---

[1] http://www.**lascom**.or.jp/member/iomver/iomver4_303.pdf

Differences in Satellite Antenna Designs for GEO, MEO, and LEO Orbits

The path loss characteristic of communications satellites is one of the most critical issues. The spreading out of the signal from the time it leaves the satellite until it reaches the ground station or user terminal on an aircraft of ship decreases the satellite's telecommunications performance and throughput capabilities. Since a low earth orbit (LEO) satellite can be up to 40 times closer to Earth than a GEO satellite, the relative performance in terms of the flux density can be up to 1,600 time (40 × 40) greater, if the satellite antennas and power of for the GEO and LEO satellites were exactly the same. The path loss is of course due the spreading circle of the signal strength. Since the area of a circle is πr^2 then the spread of the signal over the transmission path must be calculated by squaring the height of the orbit.

The varying look angles or masking angles for LEO satellites in a global constellation, however, must be taken into account. When this is considered and the relative maximum transmission distances between a GEO satellite at its longest transmission path is compared to the longest transmission of a LEO satellite, then the relative lengths of the path distances shrinks to about 25–1. This means the relative advantage of a LEO satellite to GEO in terms of path loss shrinks to perhaps 625 times (25 × 25). This is a complicated way of saying that a LEO satellite can have on the order of a 28–30 dB advantage over a GEO satellite due to path loss consideration.

As always, there are a number of other things to consider here. When it is recognized that three satellites give global coverage from a GEO orbit but it takes on the order of 50 or so LEO satellites to give complete global coverage at all times, then the seeming relative advantage begins to disappear. This type of comparison is also based on the satellite power and antenna systems' gain being exactly the same. It is possible of course to increase the power and high-gain antenna to add a relative 25dB (or more) of transmission gain to GEO satellites. By making this engineering change to a GEO satellite, by adding power and high-gain antennas, the path loss disadvantage can be compensated for especially now that very large high-gain fine mesh deployable antennas can be deployed on GEO satellites. The case of the MEO satellite and its relative path loss is, of course, somewhere in the middle between GEO and LEO satellites. Thus, for a MEO satellite, it is a simple matter of calculating the height of the constellation's orbit. Then one again computes the path loss that come from the spreading circle of the signal strength that comes from a medium orbit satellite. The technical specifics of these considerations are provided below.

GEO Systems

The free-space propagation loss L (represented in decibels) is given in the formula below. In this case, L is equivalent to P_T/C. The calculation for C (which is the value for received signal power) was provided earlier. For GEO systems, $GT = 1$ and $GR = 1$.

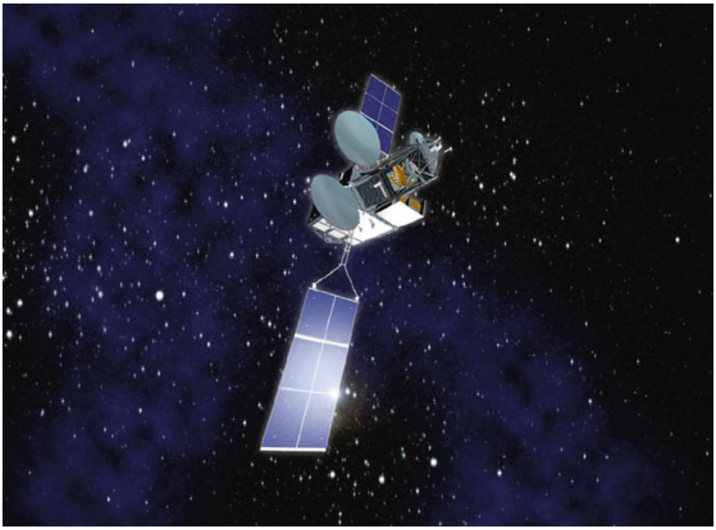


Fig. 2 The Ka-band WINDS satellite with conventional and phased array antennas (Graphic courtesy of JAXA)

$L = \frac{(4\pi d)^2}{\lambda^2}$ or *in a more convenient form* $L(\mathrm{dB}) = 32.44 + 20\log f(\mathrm{MHz}) + 20\log d(\mathrm{km})$.

Formula for determining free-space propagation loss (L): Since the distance between a GEO satellite and an earth station is around 38,000 km, the propagation loss at 20 GHz in the Ka band is about 210 dB. Thus, the requirement of link budget is severe. This can be overcome as noted above by adding high power to the GEO satellite and extremely high-gain antennas many meters in diameter.

An example of a GEO orbit satellite communications system is the WINDS experimental satellite, which is shown in Fig. 2.

In the case of the WINDS satellite, a 5 m aperture earth station antenna is needed for the highest broadband speeds of about 1.2 Gbit/s communication, a 2.4 m class can support 622 Mbit/s speeds, while a 1.2 m class antenna allows 155 Mbit/s throughput, and the smallest 45 cm class antennas can still support 1.5–6 Mbit/s speeds (Fig. 3).

A transportable class WINDS 2.4 m aperture antenna is shown in Fig. 4.

In addition, GEO orbit systems when utilized for mobile satellite service (i.e., MSS communications) lead two additional considerations that add to the need to larger satellite antennas than is the case for a fixed satellite service (FSS) satellite like WINDS. First of all, lower frequencies are used for mobile communications because radio waves with longer wavelengths are more tolerant of obstacles being in the way such as trees, the roof of a car, or truck, etc. But lower wavelengths mean the antennas must be larger to focus the longer wavelength signals more accurately. Also the receiving antennas that consumers are using must be quite small since they are intended to be completely mobile. Thus, these user transceivers cannot be effectively equipped with a tracking capability. To accommodate the need for

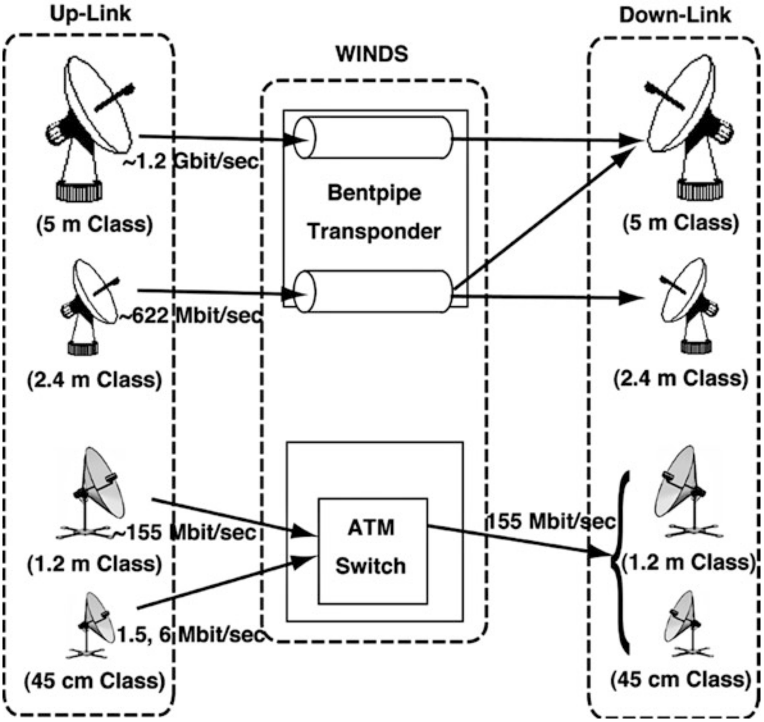


Fig. 3 The various classes of earth stations required for broadband services (Graphic courtesy of NICT of Japan)

small user devices and thus a low gain antenna, the satellite antenna in space, especially from GEO orbit, must be much, much larger to compensate.

In fact, the GEO satellite antennas for mobile communications today can be anywhere from 12 to 20 m in diameter (or in the range of 40–65 ft). In the case of the Engineering Test Satellite-8 (ETS-8) of Japan, the deployable downlink mobile satellite antenna was 17 m by 19 m in size (see Fig. 5 for a graphic of the ETS-8 Satellite and its large deployable antennas for uplinking and downlinking in the MSS bands).[2]

Since the ETS-8 GEO orbit antenna system is so high gain, this satellite is able to work with an S-band hand-held terminal as shown in Fig. 6. Its size is 58 mm (W) × 170 mm (D) × 37.5 mm (H) and weight is 266 g (or about a half a pound) without battery. The terminals designed for use with the latest MSS systems such as Thuraya, Light Squared, Terrestar, and Inmarsat have handsets that are comparable or even smaller in size.

---

[2] The ETS-8 Project http://www.jaxa.jp/projects/sat/ets8/index_e.html

Fig. 4 A transportable 2.4 m Ka-band satellite that supports 622 Mbit/s (Graphic courtesy of NICT of Japan)

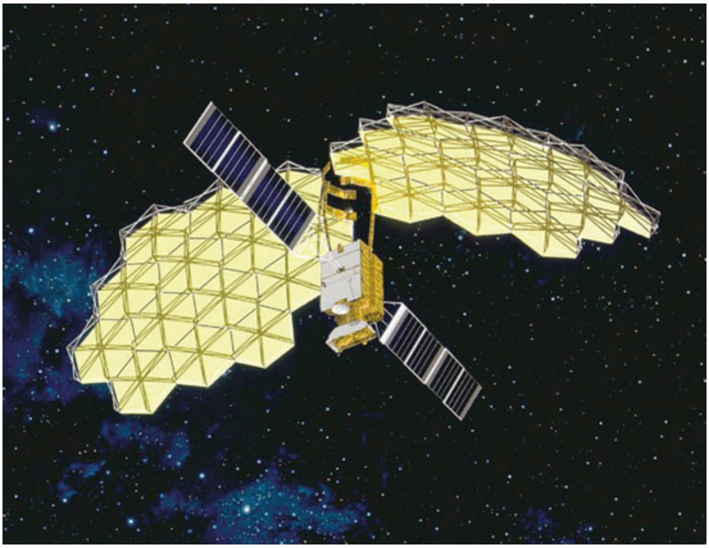


Fig. 5 The ETS-8 satellite with large deployable antennas for mobile satellite service (Graphic courtesy of JAXA)

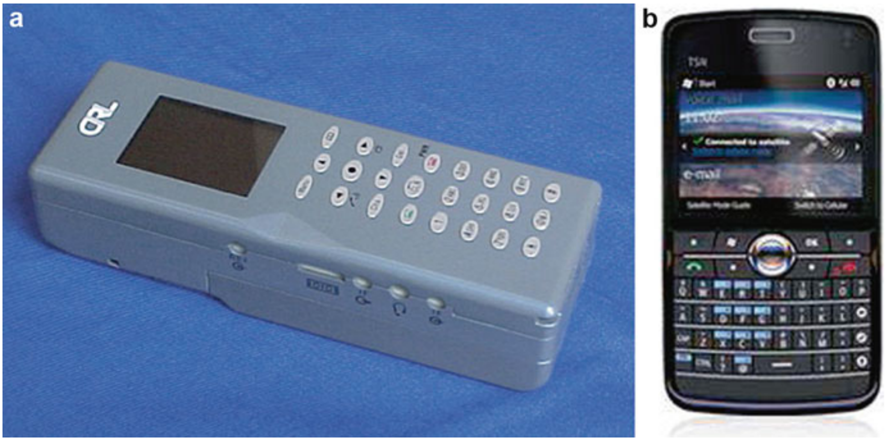


Fig. 6 (**a**) The hand-held terminal designed for the ETS-8 satellite. (**b**) The ISAT phone pro that operates with Inmarsat satellites

MEO Systems

As for the MEO (medium earth orbit) systems, the distance between a satellite and earth stations is usually in the range of 8,000–20,000 km (5,000–12,500 miles). These MEO orbits are typically chosen to avoid the most intense radiation of the Van Allen belts. Thus, LEO and MEO constellations are carefully designed for deployment either below or above the parts of the Van Allen belts with the most intense levels of radiation.

The propagation loss for MEO orbits is in the range of 189–201 dB in the Ku-band frequency (14 GHz). This propagation loss is smaller than the case of GEO satellite. But it is still very large. Therefore, antennas of both satellite and ground station often tend to have tracking capability. It is important that the MEO satellite antennas, its onboard electronics, and its solar cell power array employ protective shielding against the Van Allen radiation belts that extends in various zones to an altitude beyond 10,000 km. The following point must be considered carefully in terms of designing an antenna for MEO systems:

- Designing an antenna of MEO satellite must consider carefully the optimum orbit. This means both assessing levels of radiation from the Van Allen belts and the increased path loss that higher orbits entail.
- In addition to the antenna system design, one must also give special consideration to the design and protection of the computer control system plus all of the satellite electronics devices. It also means applying silica coating for the solar cells to prolong their life against the hazards of Van Belt radiation.

LEO Constellations

LEO satellite constellations often are deployed from an altitude of about 650–1,200 km (i.e., from about 400–750 miles). The frequency bands utilized for

Table 1 Typical LEO systems and their characteristics

Name of system		Iridium	Globalstar	Orbcomm
Number of satellites		66	48	34
Altitude		780 km	1,100 km	785 km
Inclination		90°	52°	45°
Intersatellite link (ISL)		Yes	–	–
Store-forward com		–	–	Yes (business to business (B2B) data service)
Voice com		Yes	Yes	No
Data com		Yes	Yes	Yes
Latitude of service coverage		Global	±60°	±50°
Frequency band	User link	L band	L band (up), S band (down)	VHF band
	Feeder link	Ka band	C band	VHF band
	ISL	Ka band	–	–
Type of satellite borne antenna		Phased array	Phased array	Cross Yagi
Type of antenna of hand-held terminal		No-tracking and built-in extendable antenna	No-tracking and built-in extendable antenna	Whip

MSS services are typically in the UHF, the L band, or S band. In the case of LEO constellations, propagation path losses are on the order of 170–180 dB. The much smaller path loss in LEO orbits tends to make it easier to design satellite handset for users that do not require any tracking. An example of typical LEO system is shown in Table 1.

As shown in Table 1, the intersatellite link of Iridium system is a unique feature. This capability reduces the number of gateway stations, while Globalstar and Orbcomm need many gateway stations in the world because there are many satellites in the constellation to be tracked and commanded. Iridium and Globalstar can provide both voice and data communication, while Orbcomm can provide only data communication with store-forward communication. Since the service area depends on both the inclination and altitude of a constellation orbit, communications using Globalstar or Orbcomm will typically suffer from a bigger propagation loss than using Iridium.

Technical and Economic Challenges in Designing Satellite Antennas

The demands of designing efficient satellite antenna systems to meet the special needs of effective telecommunications from GEO, MEO, and LEO orbits have driven the satellite industry forward over the past 50 years. Overriding the demands to meet the requirements unique to the particular orbits have been the prime desire to design satellite antennas and associated communications electronics systems that were higher in overall performance. This has meant finding ways to design very stable platforms to all the precise pointing of high-gain antennas. It has meant finding ways to build larger and more capable antennas for ever higher frequency bands that could be constructed of lighter and more cost-effective materials. It has also meant finding ways to design "deployable antennas" that could not only be very high gain but could be launched within the constraints of the fairings of available rocket launch systems. At the same time, there has been a drive to increase reliability and to support longer lived satellites while driving down the cost of manufacturing and testing of these systems. In all the results are impressive. The most advanced satellite antennas of today when compared to the omni antennas of the very first experimental communications satellites are on the order of a million times more capable when examined in terms of gain efficiency and lifetime.

The Evolution of Satellite Antenna and Communications Systems

The following history of the evolution of satellite antenna systems for telecommunications services can be summarized as having the following main elements:

- Design of spacecraft platforms that can stabilize more sophisticated antenna systems and allow more accurate pointing of narrower transmission beams
- Design of higher gain antenna systems
- Design of lighter weight and less massive high-gain antennas
- Design of more reliable antenna systems in terms of test, deployment, and operation
- Design of systems to support intersatellite links
- Design of antenna systems that can operate at ever high frequencies, that is, up to Ka band and beyond into the EHF frequencies and overcome the effects of environmental attenuation
- Design of improved feed systems that can allow the creation of a very large number of spot beams that can be interconnected via on-switching and board processing
- Design of satellite antenna and electronic systems that can work interactively with terrestrial telecommunications systems and/or support multiservice satellite capabilities from a multipurpose bus
- Design antenna systems to support a wide range of telecommunications and data relay requirements for a wide range of application satellites including remote sensing, satellite navigation, satellite meteorology, as well as scientific satellites
- Design satellite telecommunications systems capable of supporting interplanetary communications

Phase One: The Earliest Phase of Satellite Communications with Omni Antennas

One of the world's first artificial telecommunications satellites such as the 1962 low earth orbit Telstar (Fig. 7) contained an omni antenna. This omni antenna was deployed on the Telstar because there was no stabilization or pointing system that could accurately point this satellite toward Earth as would be required by a higher gain antenna. Most of the irradiated power from this small satellite was thus uselessly transmitted into outer space. Any attempt to significantly increase the telecommunications throughput of a communications satellite would clearly need to find a way to concentrate the irradiated energy toward Earth rather than sending electronic signal in every direction with equal effect (Fig. 7).

By the time of the launch of the Early Bird satellite (i.e., the Intelsat I (F-1)), the Hughes Aircraft designers, led by Dr. Harold Rosen, concluded they could "squint" the antenna beam so that a signal could be somewhat directed toward Earth while circling in geosynchronous orbit. This was more efficient than a full omni-beam antenna but still a great deal of the energy was lost into outer space (Fig. 8).

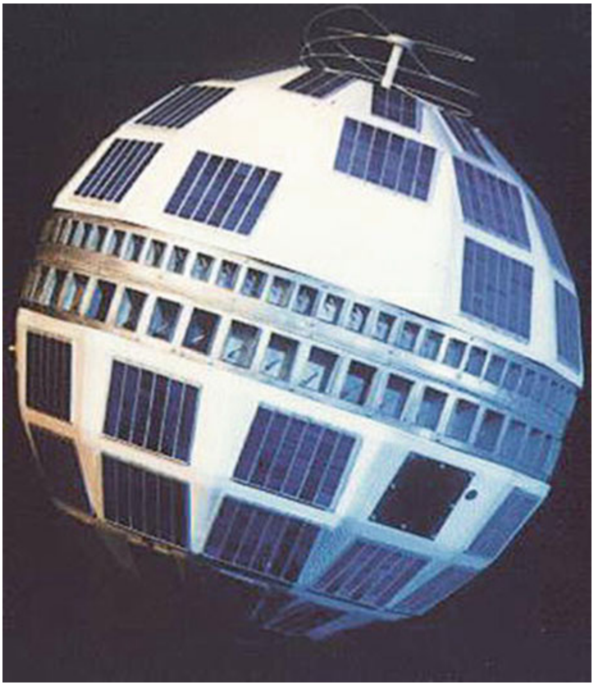


Fig. 7 The Telstar satellite designed by Bell labs and launched by NASA (Photo courtesy of NASA)

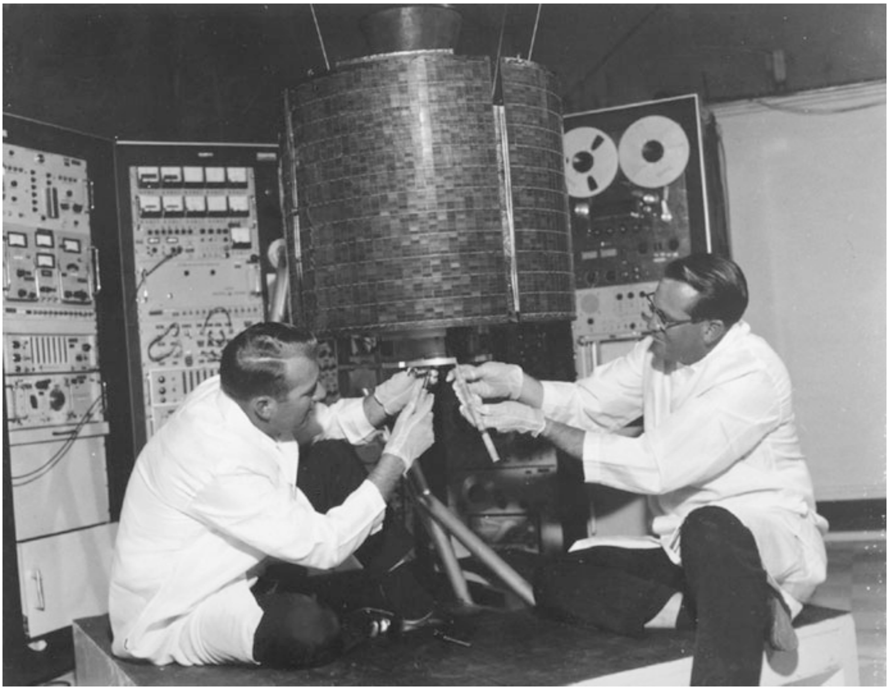


Fig. 8 The Early Bird satellite launched by Intelsat in 1965 improved performance via a squinted-beam omni antenna (Graphics courtesy of Comsat legacy foundation)

Phase Two: Three-Axis Stabilization and Higher Gain Satellite Antennas

The next step in the evolution of satellite communications antenna performance came by deploying a “despun antenna” that could be constantly pointed toward the Earth. This was achieved by developing a “spinning” spacecraft with sufficient rotational speed and angular momentum to maintain a constant vertical stabilization just like a spinning top. The critical aspect of the design was to have the antenna and electronics inside the spacecraft to spin in the reverse direction from the outside spacecraft body that contained the power and fuel for the firing of jets. Since the rotational speed of the spacecraft on the outside could be exactly matched to the interior antenna, the effect was to create a system that could continuously point to the Earth in all three axes.

The horn antenna system on the Intelsat 3 satellite (Fig. 9) spun at 60 revolutions a minute and the spacecraft matched this speed into opposite direction. The result was a reasonably high-gain antenna constantly and steadily illuminating the earth. The combination of higher power and higher gain antenna system allowed this satellite to achieve a capacity of 1,200 duplex voice circuits plus two color television channels. This capacity of this satellite when launched in the late 1960s in

Fig. 9 Intelsat 3 satellite manufactured by TRW with "Despun" horn antenna (Graphics courtesy of Comsat legacy foundation)

effect duplicated the entire world's capacity for overseas communications at that time. The bearing in the Intelsat 3 that allowed the spinning antenna to operate in this mode unfortunately froze in the first Intelsat III satellite to be launched successfully. Refined engineering of the diameter of the bearing and a new lubricating system allowed the remaining satellites in this series and hundreds of three axis spinners that followed to operate successfully. In fact, it was this type of satellite that allowed transoceanic services across the Atlantic, Pacific, and Indian Oceans to create the first truly global satellite communications system to be established in mid-year 1969 and allowed global television coverage of the first Moon landing in July 1969 (Fig. 9).

Phase 3: The Creation of Higher Gain Parabolic Reflector on Communications Satellites

Now, that three-axis spinning satellite communications systems had been demonstrated to work in geosynchronous orbit, the next step was to upgrade from horn antennas to parabolic reflector antennas on commercial satellites. The Intelsat 4 and 4A satellites, which were manufactured by the Hughes Aircraft Company, featured such parabolic antennas in the 1970s (Fig. 10).

Once again the addition of more power (as a result of much larger solar arrays and drop-down circular panels) plus significantly higher gain parabolic antennas

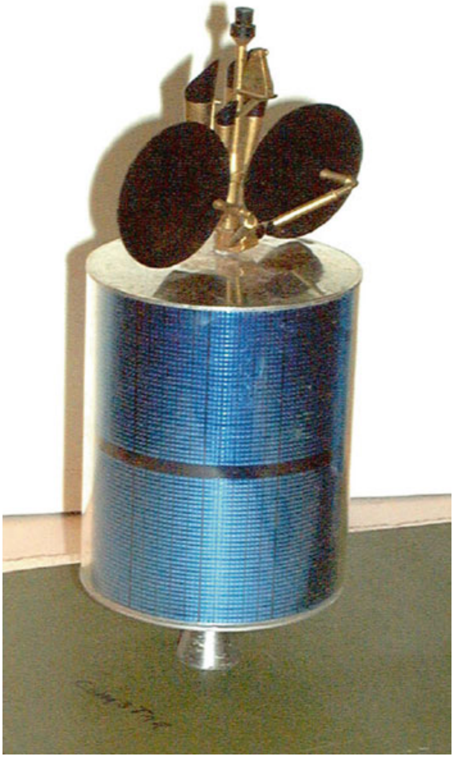


Fig. 10 The Intelsat 4 with high-gain despun antennas (Graphics courtesy of Comsat Legacy Foundation)

allowed another significant boost in communications capacity to support transoceanic communications.

Phase 4: The Advent of Three-Axis Body-Stabilized Communications Satellites

The "despun" satellite worked well for pointing 1 or even 2 m parabolic satellite antennas toward Earth, but if one wanted to deploy even higher gain antennas that could create very tightly pointed beams to targeted areas, this design had limitations. Further, the spinning design with the solar cells on the outside of the drum meant the solar cells were being illuminated by the sun only about 40 % of the time.

The answer that was developed at the Jet Propulsion Lab for interplanetary exploration missions was to replace the spacecraft spinning like a top and instead put momentum wheels that spun much faster inside the spacecraft body to create a platform that revolved around the Earth that remained stable in all three axes. These momentum or inertial wheels because they rotated at speeds up to 5,000 rpm could be much smaller than the overall spacecraft since mass times velocity determines momentum.

Fig. 11 The Intelsat V designed by space systems/Loral featured three-axis body stabilization (Graphics courtesy of Comsat Legacy Foundation)

This three-axis stabilized spacecraft with solar arrays extended from the box-like body offered a number of advantages. It could be pointed with an accuracy of 0.5° or less and thus host very large antenna systems. The resulting spot beams could be held to their optimum location with much less variation in gain. The extended solar arrays could "see" the sun, except during eclipse periods, all the time and the arrays could be tilted to get maximum exposure. In practice, this design has also proved quite reliable, and magnetically suspended momentum wheels have essentially no friction and thus no mechanical wear out or lubrication issues. This design for communications satellites was commercially deployed in the 1980s with the Intelsat V and other spacecraft (Fig. 11).

Phase 5: Service Diversification and Alternative Satellite Antenna Design

The last 30 years has been a time of diversification with many new types of satellites being designed for fixed, mobile, broadcast, and defense related services. These different services operated in many different frequency bands. This required different types of satellite antennas to meet system users' needs. These various requirements led to the creation of different types of antennas on the satellites – that generally grew bigger. But on the ground or on the oceans or in airspace, the user antennas and receive only terminals have grown smaller and smaller. NASA's Application Test Satellite 5 (ATS-5) demonstrated some 35 years ago the feasibility of deployable antennas for broadcast and mobile service applications. This technology has evolved further and further and quite rapidly as well. Deployable antennas, such on those on the experimental ETS-8, and on operational satellites like Inmarsat 5, Terrestar, and Light Squared satellites, are clearly operating effectively despite their large reflector sizes.

Also leading the evolution toward large space based satellite antennas was the Communications Test Satellite – a joint project of NASA and the Canadian Space Agency. This satellite also demonstrated new mobile satellite service capabilities and the feasibility of very small aperture terminals (VSATs) in remote locations such as the Amazon. A series of experimental vehicles in Europe (known as the experimental communications satellites (ECS) series and Maritime Experimental Communications Satellites (MAREC)) plus the Japanese Experimental Test Satellites helped to advance communications satellite antenna design. Innovative operational programs such as the Marisat satellite design to provide communications services to the US Navy and commercial maritime shipping and the maritime packages on some of the later Intelsat V satellites also demonstrated not only a host of new satellite applications but a wide range of new satellite antenna designs including the idea of having more than one communications payload and sets of antennas on one satellite.

In general, three types of satellites and antenna systems evolved. (1) There are broadcast satellites for BSS service that have smaller antennas for broader coverage. These spacecraft are equipped with very large power systems so that they could operate with very small receiver dishes. (2) Mobile satellites for MSS service have evolved toward spacecraft with very large antennas because they operated at the lower frequencies and new to work to very small mobile antenna units including hand-held satellite telephone units. This has led to these types of spacecraft operating with a large aperture multibeam antenna that can create a large number of beams which can be interconnected via onboard switching. (3) Finally fixed satellite service (FSS) spacecraft also have moved toward higher power and larger antennas to support not only very small aperture antennas (VSAAs) but even ultra-small aperture antennas (USAAs) especially as these services migrated to Ku band and even to Ka band. These space craft also retain smaller aperture antennas for global and zonal coverage.

The most demanding designs for broadcasting and mobile satellite systems often benefit the design of FSS satellites that do not have the most stringent requirements. Today, there are satellites that tend to be evolving toward "multipurpose" busses that can meet a variety of requirements if simply equipped with the right antennas to service the right spectrum bands that can be fitted to the right class of spacecraft body and power system suited to the various service needs. Just as automobiles are designed to be built on various sized platforms, the satellite industry has evolved in the same direction. Thus, one takes a solar power array system and supporting battery system, takes an appropriate spacecraft body and stabilization system, and then fits it with the various types of antennas need to meet one or perhaps several missions. Dose this mean that the evolution of satellite technology is essentially complete and most types of antennas need to meet future needs already invented? The answer is clearly no. The next generation of satellite antenna technology is currently being invented.

Future Satellite Antenna Technology

One of the keys to the future may well be phased array satellite antennas or phased array feed systems that can create a very large number of beams off of a multibeam antenna reflector. This type of technology is currently being developed on experimental satellites, and phased array antennas have actually been used on the Iridium mobile satellite system.

Phased Array Satellite Antennas of WINDS Satellite

An example of a Ka-band active phased array antenna (APAA) is the one that was designed for the broadband Internet satellite WINDS (Wideband Internetworking Demonstration Satellite). This satellite operates in the 28 and 18 GHz bands. The WINDS satellite has been shown previously in Fig. 17.2. The WINDS satellite performs as a very broad band Ka-band satellite with high-speed transmission capability of in the gigabit/s range. This satellite can operate to a single location at a maximum speed of 1.2 Gbps. The system can operate as a bent pipe or it can use onboard ATM (Asynchronous Transfer Mode) switching and multibeam antennas with a high-speed scanning capability (Yajima et al. 2007).

The APAA can scan independently each two electronic beams of transmit and receive and covers almost whole regional area of the satellite's outlook as shown in Fig. 12. The antenna control method enables consecutive wave modes and SS-TDMA (Satellite Switched Time Division Multiple Access) mode to reach various sites (Yajima et al. 2007).

The active phased array antenna pictured in Fig. 13 consists of 128 antenna elements that can be flexibly used to create beams for satellite transmission. Each element consists of a high density RF module.

The main performance characteristics of the WINDS APAA are shown in Table 2. The block diagram of APAA is shown in Fig. 14.

The existence of such a large number of antenna elements allows the flexible creation of many beams, not through conventional physical beam forming off of a multibeam antenna but through the creation of beams through electronic processing.

The WINDS satellite represents state-of-the-art capabilities for a broad band phased array satellite antenna on a geo satellite.

Phased array satellite antennas have been used on commercial and defense communications satellites with lesser active elements and at lower orbits. The Iridium satellite constellation, launched in 1997 and 1998, operates in low Earth

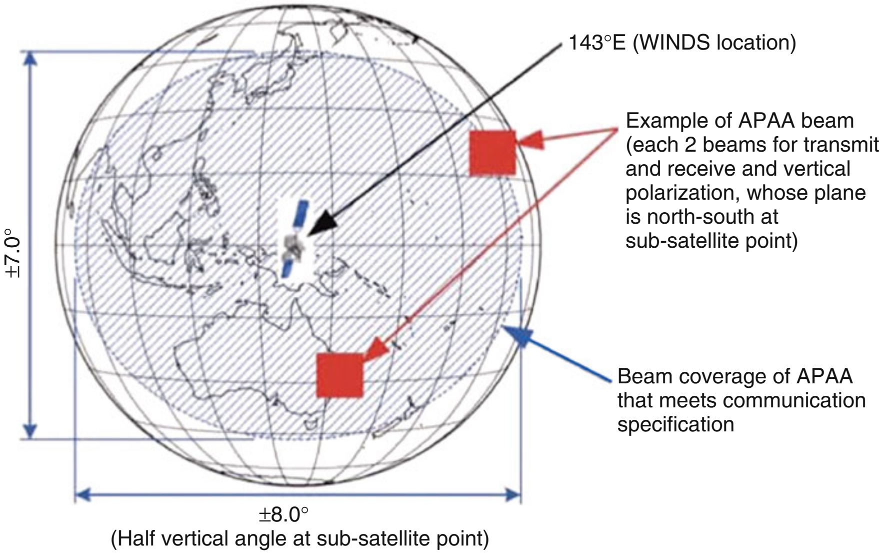


Fig. 12 Active phased array beam forming capability of the WINDS satellite (Graphic courtesy of the NICT of Japan)

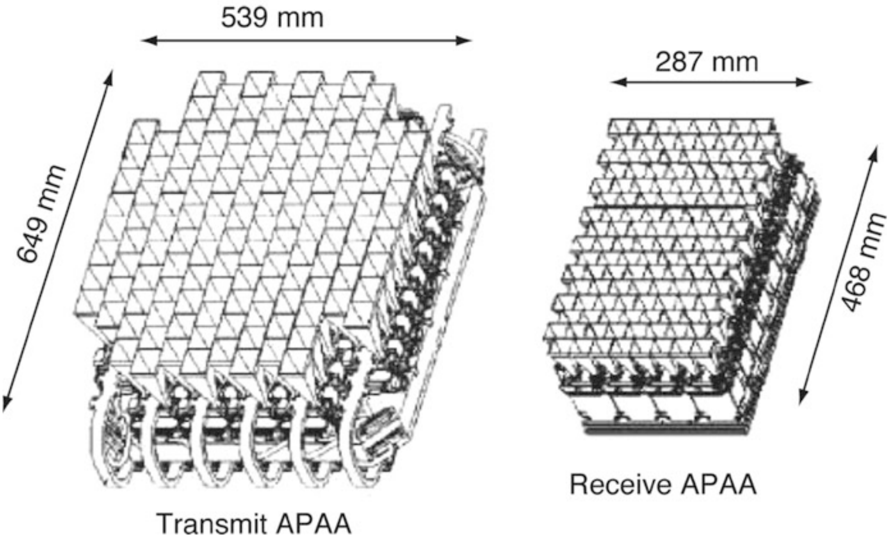


Fig. 13 Phase array elements in the WINDS satellite design (Graphics courtesy of JAXA)

orbit and its mobile services are provided in the L band. In order to achieve the 48 cell coverage required to generate the 48 beams projected by each Iridium satellite, the design called for three separate panels that were able to simultaneously transmit or receive 16 beams.

Table 2 Performance of WINDS active phased array antenna (APAA)

Item	Unit	APAA	
		Transmit antenna	Receive antenna
Style of antenna		Directly emitted phased array antenna	Directly emitted phased array antenna
Size of antenna	mm	1,510 × 990 × 1,530	
Weight of antenna	kg	183	
Aperture of array	mm	649 × 539	287 × 468
Frequency band	GHz	18	28
Frequency bandwidth	GHz	1.1	
Number of elements		128	128
Polarization		Linear polarization	
Scan range of beam	deg	Within ellipse of (θ,ϕ) (see Fig. 17.12)	
		Long axis: $\theta \leq 8$ Short axis: $\theta \leq 7$ $\phi = 0 \sim 360$	
Effective isotropic irradiated power (EIRP)	dBW	$\geq$54.6 per carrier	
		$\geq$52.1 per 2 carrier	
G/T	dB/K		$\geq$7.1
Number of bit of phase shifters	bit	5	5
Operation mode		SS/TDMA mode	
		Continuous mode	
Timing of beam scanning	ms	2 (SS/TDMA mode)	
Consumption power	W	$\leq$750	

These three phased array elements are arranged to be activated by a Butler-matrix-type beamformer that can simultaneously generate the 16 required shaped beams with minimal losses in gain. These systems that were manufactured by the Raytheon Corporation have proved not only able to generate clearly differentiated beams but with high reliability with many of these satellites operating more than 12 years in orbit and a survival record of over 600 satellite years of operation for the combined constellation. This is of particular note since the satellite antennas as well as the entire spacecraft were subject to much accelerated testing. These satellites have another feature that will be discussed later, namely the use of intersatellite links (ISLs) to allow the entire global constellation of 66 satellites plus spares to be managed by only two control stations. These RF based ISLs operate at 23 GHz and each satellite has four such ISLs. Two of the ISLs are used to connect to the satellites in "front" and "back" of the North–South latitudinal orbit, and two are used to connect to the satellites to the "sides" in parallel orbits in the longitudinal direction (Fig. 15).

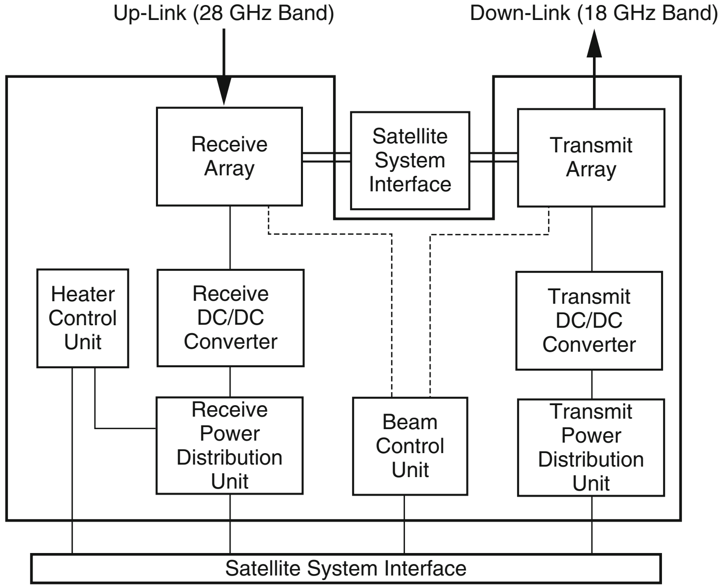


Fig. 14 Block diagram for the WINDS satellite transmit and receive system (Graphic courtesy of NICT of Japan)

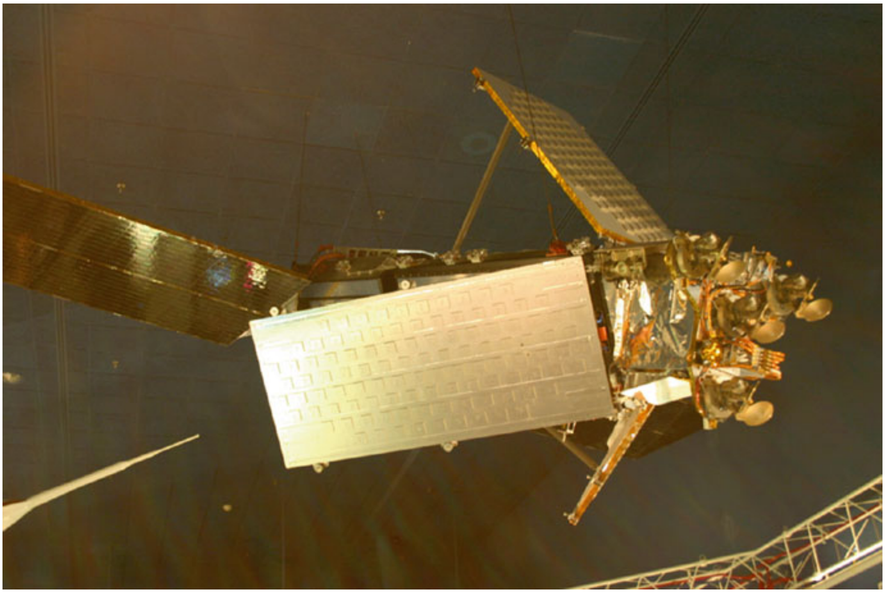


Fig. 15 An iridium satellite with three phase array panels and ISL antennas shown (Graphic courtesy of iridium)

Phased Array Antenna for the Quazi-Zenith Satellite System (QZSS)

Yet another example of a phased array antenna is one onboard QZSS (Quasi-Zenith Satellite System) as shown in Fig. 16 along with the major specifications (Jono et al. 2006). This satellite is in an unusual orbit that is like a geosynchronous orbit but inclined at a 45° angle to the orbital plane. This satellite carries experiments for space navigation services. The unusual orbit when populated by three satellites can provide extremely favorable look angles for coverage to the Japanese islands that allow service even in areas with extensive high rise buildings such as in downtown Tokyo.

The phased array antenna on the Quasi-Zenith Satellite (QZSS) operates in the L band. In the case of the active phased array antenna is an L-band helical antenna with 7 inner elements and 12 outer elements. The inner elements provide antenna gain with right circular polarization and the 12 outer elements form the antenna beam (Furubayashi et al. 2008).

In future years, there will likely be more phased array antennas and multibeam reflector antennas with phased array feed systems as improved engineering and economies of scale allow such antennas to be designed and built at lower cost.

Fig. 16 Graphic of the quasi-zenith satellite (Graphic courtesy of JAXA)

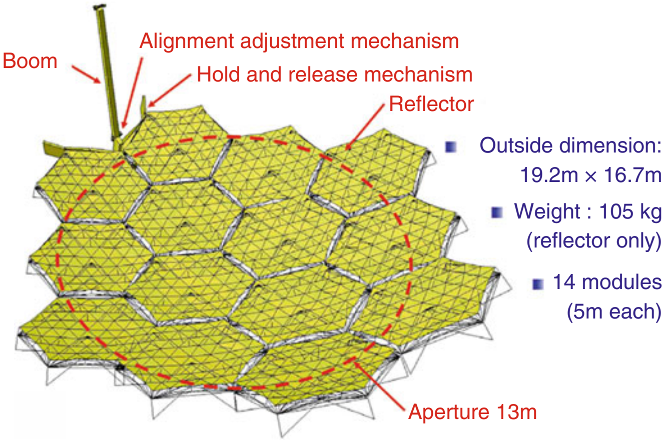


Fig. 17 Highly conformal large-scale deployable mesh antenna

Large Deployable Antennas for Mobile Services

As noted earlier, one the most important elements for achieving truly high-gain satellite antennas in space is the need to develop deployable satellite antennas. This is because much larger aperture antennas can be achieved with greatly reduced weight using a deployable and unfurlable design. Such a design allows very large aperture devices to be stowed compactly within the nose fairing of a launch system that might be 4–5 m in diameter. Deployable antennas in the range of 12–20 m in diameter could not otherwise be launched into orbit unless they were folded up and then deployed in space. The challenge has been not only to develop systems that can unfurl or deploy in space but to use extremely fine mesh and precision frameworks so that the deployed antenna can still conform to the parabolic shape with great precision to accommodate Ku- and even Ka-band frequencies (i.e., 12–30 GHz) as well as in lower S- and L-band systems (Fig. 17).

Although 18 m (or 60 ft) antennas represent the current industrial state of the art for the commercial communications satellite industry, there are apparently classified missions that have created unfurlable antennas that are up to 100 m in diameter that have been developed and deployed such as the reported National Security Agency NROL-32 mission.[3]

In addition to the challenge of designing antennas that unfurl or otherwise unfold in space, there is also the difficulty of creating a complex feed system that might create hundreds of beams off the surface of such a large reflector. In designing for the future, a phased array feed system may be the only way to create such a complex feed system within a condensed enough space. In the case of the Inmarsat 4 satellite that operates with a 9 m deployable antenna there is a 120 element helix cup array

---

[3] http://www.liquida.com/page/13583910/

which was designed and built in Europe by EMS. This feed system is designed to create over 220 spot beams and some 20 zonal and broader coverage beams in the L band in the 2,500–2,600 MHz range. The design of the array feed system in this case emphasized the virtual elimination of passive intermodulation (PIM) products that would create interference among the large number of beams created.

The Antennas of Terrestar and Light Squared

The latest commercial technology in the field of extremely large satellite antennas that deploy in space are represented by the two US-based mobile satellite operators known as Terrestar and LightSquared (formerly SkyTerra and prior to that Mobile Satellite Ventures). These satellites have gigantic satellite antennas that are 18 m in diameter with fine gold-plated mesh parabolic reflectors. These satellite antennas were designed and built by the Harris Corporation (Terrestar) and Boeing (Light Squared), respectively. These satellites both have extremely fine contoured deployable mesh antennas designed to provide the maximum gain in the assigned S-band frequencies. These deployable mesh reflectors are by far the largest commercial satellite antennas ever deployed. These satellites are also different in that they are part of an integrated space and terrestrial telecommunications network for mobile communications. Each of these operators will invest on the order of $7 billion in space and ground assets to provide blanket mobile cellular services for the entire continental United States, Canada, Puerto Rico, Hawaii, and Alaska. These satellite providers are offering what is known in the United States as mobile satellite service with an ancillary terrestrial component (MSS-ATC). The concept is to deploy and operate terrestrial cellular mobile services including video services via 4 G standards in urban areas but have a dual mode satellite and terrestrial phone that will switch to satellite service as one makes the transition to rural or more sparsely populated suburb or exo-urban areas. These companies intend to provide phone, text, data, and video services. The gigantic structural framework for the 18 m Terrestar satellite antenna was deployed in space in July 2009. This unfurlable gold mesh antenna is shown in its “stowed configuration” prior to launch in Fig. 18.

The Solaris Corporation of Dublin, Ireland, has been licensed in a number of European countries to provide a similar hybrid MSS and terrestrial service in Europe that will, like the US hybrid systems, provide e-mail, text, voice, and video services using a combination of terrestrial and satellite networks. This will, like in the case of the US-based MSS-ATC services, require the launch and deployment of truly large 18 m (60 ft) satellite antennas to achieve the amount of gain to provide video to small dual mode 4 G satellite/cellular phones.

Optical Communications Systems

The rapid gains in terrestrial communications in the past two to three decades has come with the transmission of messages utilizing the light spectrum and sent through

Fig. 18 Framework for the 18 m terrestar mobile satellite antenna (Graphics courtesy of Harris corporation)

fiber optic cable. The very high quality light fiber carries modulated and encoded messages with very little attenuation and the enclosed cable prevents interference. Since the light frequencies are much higher than radio waves and thus their wavelengths very much smaller this allows extremely high throughput capability. The development of high quality blue and green lasers that are at even higher frequencies than red wavelength lasers opens up even more capacity for the future. One pathway to the future, in terms of higher capacity satellites, would thus seem to be the use of modulated light waves. The problem, in terms of transmission of modulated light signals via satellite, is that atmospheric conditions can easily interrupt these signals. In clear sky conditions, signals can be transmitted from Earth to an optical telescope and vice versa. In the case of cloud cover, rain, snow, or other conditions, where the light signal can be blocked, the transmission will not go through. Optical transmission above the atmosphere can be of very high quality. Since the speed of light is fastest in a vacuum and encounters no interference, the application of light signals for either intersatellite links or for interplanetary or cislunar links is a very logical application, and both applications are under development in experimental R&D programs. The following discusses experiments that are being carried out using optical waves for satellite-based communications.

Optical Antenna of OICETS Satellite

One example of an onboard optical antenna is that of the Japanese Optical Intersatellite Communication Engineering Test Satellite (OICETS) (Jono et al. 2006). The experimental OICETS was launched in 2004. This small satellite had a weight of 570 kg, and it was launched into a circular and nearly polar circular orbit with an inclination of 98°. It conducted successfully the optical intersatellite communication experiment with ESA's ARTEMIS satellite. The data transmission rate for the OICITS satellite was 50 Mbit/s (Fig. 19).

The mission equipment for OICITS was called as LUCE (for Laser Utilizing Communication Equipment) as shown in Fig. 20.

The LUCE consists of an optical telescope on Earth and corresponding optical and electronic parts in the satellite. The optical part on Earth was a two axis gimballed, high-gain optical antenna with an inner part. This inner optical part consisted of optical elements including a high power output semiconductor laser as well as a high sensitivity signal detector. The major specification is shown in Table 3.

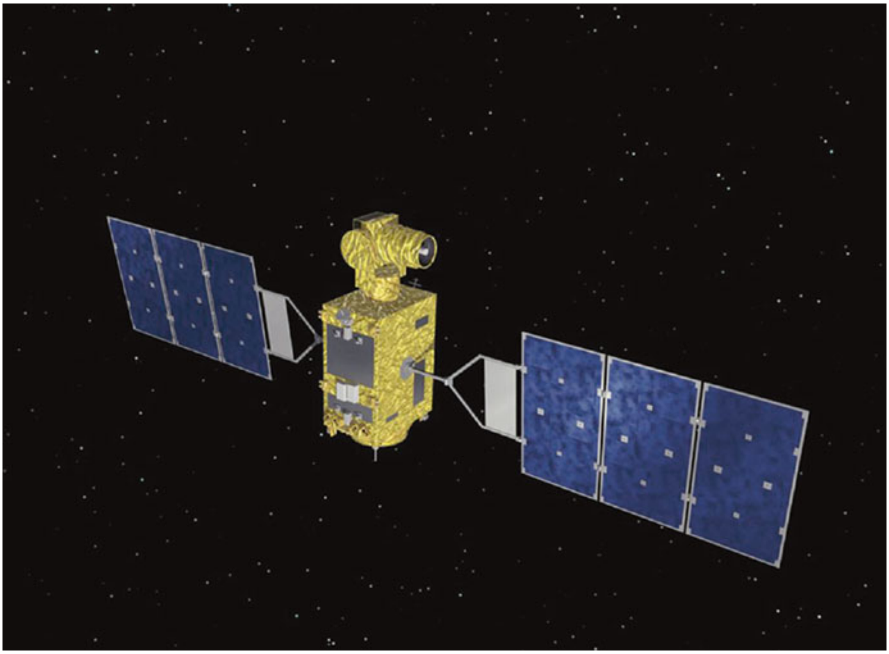


Fig. 19 The experimental optical communications satellite (OICETS) (Graphics courtesy of JAXA)

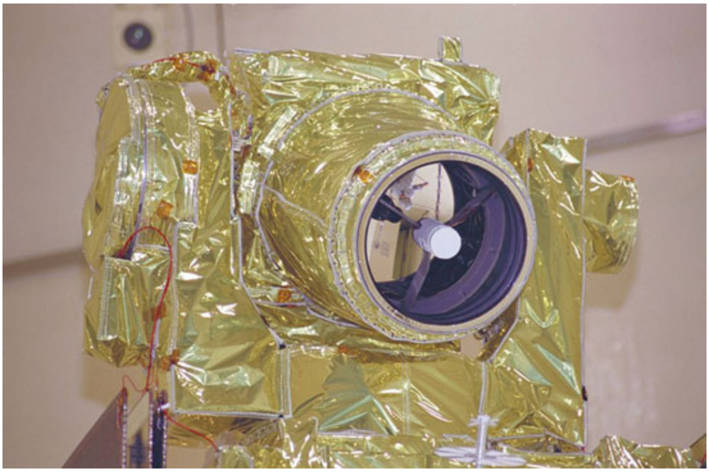


Fig. 20 The laser utilizing communications equipment (LUCE) of the OICITS satellite (Graphics courtesy of JAXA)

Table 3 Performance of OICETS optical antenna

Structure	Cassegrain-type telescope
Wave front error	Less than $\lambda/20$ rms within 1 mrad field of view ($\lambda = 847$ nm)
Magnification	20
Effective diameter	26 cm

European Optical Antennas

The Artemis satellite data relay satellite, developed by the European Space Agency, has several payloads. This experimental satellite was eventually deployed in geosynchronous orbit in late 2001 using its ion engine to reach the desired altitude. This was accomplished over a period of several weeks after its initial deployment resulted in a lesser altitude elliptical orbit. The Artemis was designed and manufactured by Astrium to support data relay between low earth orbit satellites and geosynchronous orbit. The key payload on the Artemis satellite was the SILEX Semiconductor-laser Intersatellite Link Experiment (SILEX) that can send and receive signals between Artemis and lower earth orbit satellites. Specifically SILEX has been used for experiments to link to an airplane in flight and to connect to the Japanese OICETS satellite. Most significantly, it was used once a day to connect with the French SPOT-4 remote-sensing satellite to obtain real time images from the ground or to empty the memory of this satellite for relay to the ground via conventional Ka-band radio frequency transmissions.[4]

---

[4] Artemis Satellite Reaches Geostationary Orbit, ESA News www.esa.int/export/esaCP/SEMCVY1A6BD_index_0.html

The SILEX payload is based on gallium aluminum arsenide (GaAlAs) laser diode as the transmitter and a photodiode detector, with a 25 cm aperture, fork-mounted telescope that can be used for establishing links with satellites in lower orbits. This system supports a data rate of 50 Mbps. The SILEX payload weighs about 160 kg and uses 150 W of power. Before the Artemis satellite ended operations, over 1,000 successful laser transmissions were achieved between Artemis and other spacecraft and/or aircraft.[5]

Improved Large-Scale Satellite Systems with Large Reflectors with Phased Array Feeds and Nano Satellite Arrays

There is some thought that the continued development of high capacity and low cost fiber optic systems could in time make satellite technology obsolete. This, in fact, does not seem to be the case. This is because of two important reasons. One essential reason is that fiber optic technology does not support mobile communications nor does it efficiently support broadcast and multicasting operations or services to remote and sparsely populated areas.

Secondly, satellite technology has a large number of new concepts still to be developed. There are ideas about how very large-scale passive reflectors combined with phased array feed systems to increase the capabilities of satellites even beyond the nearly 20 m antenna systems deployed with the Terrestar and LightSquared MSS-ATC systems. There are further ideas that thousands of nanosatellite elements might be deployed to form a very large "virtual satellite" in the sky that might be square kilometers in size. Some of these concepts are described in chapter "▶ Satellite Communications: Regulatory, Legal, and Trade Issues" on the future of satellite systems.

Conclusion

Currently, the emphasis with regard to RF-based satellite antennas seem to be to deploy highly efficient antennas to support the broadband spacecraft known as high throughput satellites, typically using the Ka-band frequencies. The ViaSat 1 and 2, the Intelsat EPIC, and the Hughes Network Systems Jupiter satellites have ten times or more the digital throughput of more conventional FSS satellites. The other option that is being explored is deployment of what are sometimes call "MegaLEO" satellite constellations that may have 800 to even thousands of small satellites that achieve increased system capacity – not by large and sophisticated multibeam antennas with many hundreds of beams – but rather by deploying a very large

---

[5] Silex: The First European Optical Communication Terminal in Orbit, www.esa.int/esapub/bulletin/bullet96/NIELSEN.pdf also see Silex update in Space Reference, "SILEX: More than one thousand successful optical links," http://www.spaceref/com/news/viewpr.html?pid=17298

constellation of satellites with lesser capabilities. This approach, however, raises questions as to orbital debris, higher risk of orbital collision, and interference between satellite services provided from GEO and those provided from LEO constellations. The driver of LEO constellations is to find ways to reduce transmission latency and to be better optimized for Internet services – especially in underserved areas without fiber optic networks in place.

Today the diversity of approach to achieve spacecraft throughput is only growing wider. Options involve antennas transmitting in higher frequency bands, improved coding efficiency that allow more bits per hertz, FSS services being provided by GEO satellites in competition with MEO and LEO constellations of great size, and even plans for laser-based satellite rings that change basic spacecraft architectural concepts. In short, improved and more efficient spacecraft RF antenna designs, coupled with on-board processing and more efficient coding, represent only a part of the equation in future satellite system design. These other options are to be discussed in more detail in the following chapters.

Satellite antenna and ground transceivers and user terminals are key to the future development of various types of satellite communications services. New technology, improved digital processing and coding is being rapidly developed and implemented around the world and innovations in satellite antenna design is one of the most important sources of this global development.

Cross-References

► Satellite Communications Antenna Concepts and Engineering

References

Y. Demers, A. Liang, E. Amyotte, E. Keay, G. Marks, Very large reflectors for multibeam antenna missions, in *15th ka and Broadband Communications, Navigation and Earth Observation Conference* (Cagliari, 2009)

T. Furubayashi, O. Amano, T. Takahashi, H. Noda, E. Myojin, M. Kishimoto, Test result of on-board L-band antenna assembly for high accuracy positioning experiment system, in *52nd Association Conference on Space Science and Technology*, *3F04* (2008), http://www.jaxa.jp/projects/sat/qzss/index_e.html

T. Iida (ed.), *Satellite Communications –System and Its Design Technology* (Ohmsha Ltd/IOS Press, Tokyo/Amsterdam, 2000)

T. Jono, Y. Takayama, K. Ohinata, N. Kura, Y. Koyama, K. Arai, Demonstrations of ARTEMIS-OICETS inter-satellite laser communications, in *24th AIAA International Communications Satellite Systems Conference (ICSSC), No. AIAA 2006–5461*. San Diego, California (2006)

M. Yajima, T. Kuroda, T. Maeda, M. Shimada, T. Hasegawa, S. Kitao, K. Hariu, Ka-band active phased array antenna. Spec. Issue WINDS NICT Rev. **53**(4), 49–55 (2007) (in Japanese)

Satellite Earth Station Antenna Systems and System Design

Jeremy E. Allnutt

Contents

Abstract

This chapter reviews the design and operation of user antennas for satellite communications – for fixed, mobile, and broadcast services. This review includes simple dipole antennas and progresses to Yagi-Uda antennas and then on to high-gain parabolic reflector antennas that are the most commonly used in satellite communication systems. The trade-off between antenna gain and beamwidth is

---

J.E. Allnutt (✉)
Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA, USA
e-mail: jallnutt@gmu.edu

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_20

explored in detail. The key differences in the design process for very small aperture terminals (VSATs) and large earth stations are explained. The influence of blockage on whether to choose offset-fed antennas over on-axis fed antennas is seen to be key. This is particularly true for small aperture antennas with diameters of less than 100 wavelengths. Frequency reuse through dual-polarization operation is presented, with the different system advantages of dual-linear and dual-circular operation set out. The impact of the choice of modulation on the power margins required for a given bit error rate (BER) is seen to be significant. Noise temperature contributions from the atmosphere, from the ground, and particularly from lossy-feed runs that reduce antenna performance are explored. Reducing the feed losses is key to the design of very large earth station antennas. The difference in the impact of noise temperature on the uplink and the downlink is explained, and the differences between antenna design and performance with regard to fixed-satellite service, mobile satellite service, and broadcast satellite services are noted. Finally, some additional aspects of earth station designs that are affected by the environment, both meteorological and interference, are discussed.

Keywords

Antenna systems • Bit error rate (BER) • C-band • Depolarization • Directional antennas • Ka-band • Ku-band • Micro-terminals • Modulation • Optical systems • Rain attenuation • Satellite earth stations • Site shielding • UHF • USAT • V-band • VSAT

Introduction

All telecommunication systems need to have a transmitter at one end of the path and a receiver at the other in order to complete the link. This is true whether audio frequencies are used (as in human speech, with a mouth transmitting at one end and an ear receiving at the other) or radio frequencies are used (e.g., a cell phone at one end and a base station at the other). The material between the transmitter and the receiver is referred to as the *transmission medium* and in some cases as the *propagation channel*. In order to successfully send the required message over the transmission medium, an efficient mechanism needs to be used to *launch* the message from the transmitter into the medium. Such a mechanism is an antenna.

Basic Antenna Concepts

Dipole and Monopole Antennas

An antenna is simply a device for taking energy, usually electrical energy in the form of either a current or a field, and transferring that energy as efficiently as possible from the physical structure in which it originated (e.g., a wire or a waveguide) out into the transmission medium, which is often air or space. Antenna engineers call it "launching" the field. The simplest antenna for launching a field is a dipole antenna (see Fig. 1).

Dipole literally means two poles. Usually the "poles" consist of electrical conductors, one of which is attached to the source that is to be radiated and the other to a ground or earth. The outer sheath of a coaxial cable is usually connected to the ground: thus, a simple dipole antenna (see Fig. 1) can consist of one wire attached to the center conductor of the coaxial cable and the other to the sheath. The wires are bent to form a straight line normal (i.e., at 90° angle) to the coaxial cable, and the greatest energy is radiated normal to the long axis of the dipoles, as shown in Fig. 1.

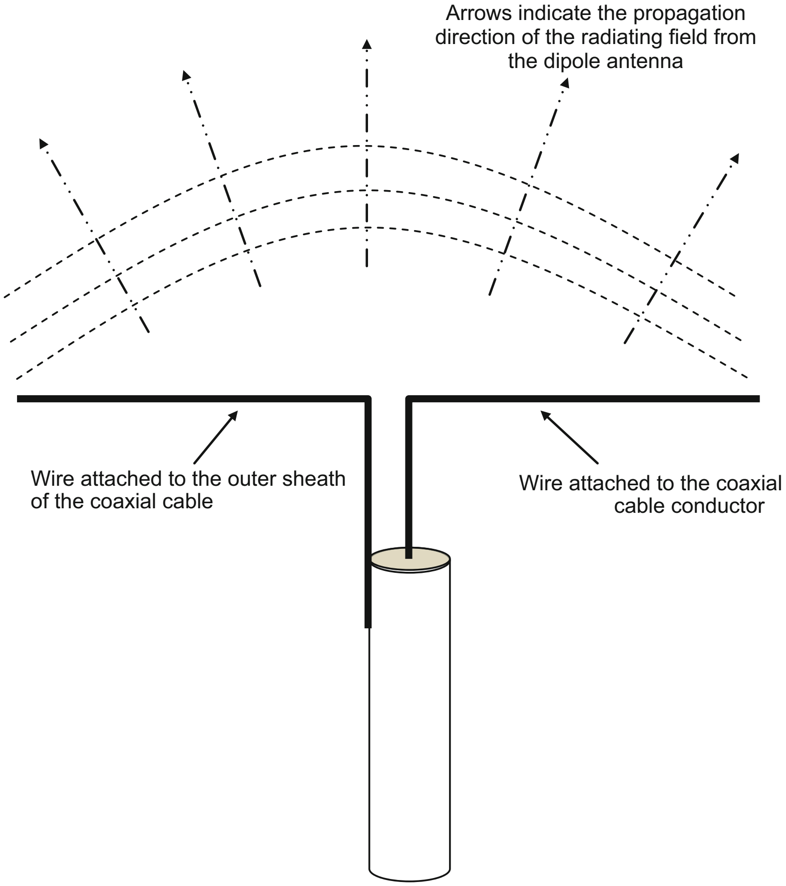


Fig. 1 Schematic of a simple dipole antenna. The example above shows a simple dipole antenna that is attached to a coaxial cable. The inner conductor of the coaxial cable is attached to one of the wire antennas and the coaxial cable's sheath to the other. Together, the two wire extensions – the dipole antenna – create a radiation field around the antenna that propagates away from the antenna, with a maximum approximately normal to the orientation of the dipoles. Most antennas that operate at frequencies below about 2 GHz tend to use just one wire extension, rather than two, and so are more commonly referred to as monopole antennas (see Fig. 2)

To make dipole antennas radiate as efficiently as possible, the length of the wires should be a submultiple of a wavelength of the signal to be radiated, usually half a wavelength or a quarter wavelength. In many cases, the wire attached to the earth of the coaxial cable is omitted, and only the central core of the coaxial cable is used to radiate. Since just one radiating element is used, this is usually referred to as a monopole antenna (see Fig. 2), and the antenna is left straight in what is called a *whip* antenna. (First used at very high frequencies (VHF) – 30–300 MHz – the monopole antennas were long and flexible, looking like a whip.)

In Fig. 2a, the monopole antenna can be seen radiating a field that is like ripples on a pond when a stone is dropped in: electromagnetic energy moves outward from the antenna with ever-increasing diameter. If the field can be measured at any point on any one of these circles radiating outward from the monopole antenna, we would

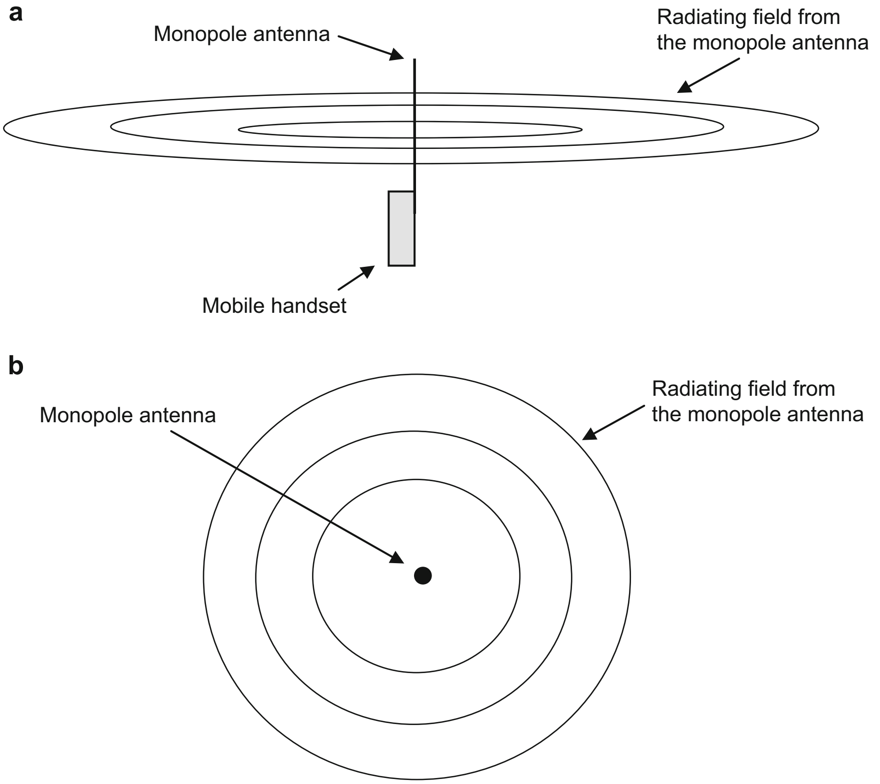


Fig. 2 (**a**) Schematic isometric depiction of the field radiating outward from a monopole antenna. The monopole antenna of a mobile handset is radiating equal power in every direction about its long axis. In modern mobile radio handsets, the monopole antenna is formed within the case of the radio unit or is part of the structure carrying the operator (e.g., an automobile). (**b**) Plan view schematic of the field radiating from a monopole antenna. In this plan view of the radiating field from a monopole antenna, it is clear that the energy radiating out from the monopole antenna has equal power in each direction that is normal (i.e., at right angles) to the long axis of the monopole antenna

find that the energy, or more properly the *power flux density*, is the same in any direction from the monopole point of origin. This is depicted schematically in Fig. 2b. An antenna that creates such a uniform field in any direction is called an *isotropic antenna*. Isotropic antennas are very popular for mobile handsets and electronic personal assistants, whether for terrestrial or satellite services, since they do not require the user to know in which direction the antenna needs to be pointed to pick up a signal. This versatility, however, comes at a very high price: much of the radiated energy is wasted since most of it does not reach the intended antenna at the other end of the link. To improve on this, the radiating element – the antenna system – must increase the radio energy radiated in the desired direction or received from the desired direction when compared with all the other possible directions. This enhanced performance antenna will now have a preferred direction in which to transmit and receive radio energy. The increase in energy radiated on the preferred direction is referred to as antenna gain.

The Concept of Antenna Gain

To understand the concept of antenna gain, it is important to know that a directive antenna does not create energy: it just focuses it in the desired direction. This is illustrated in Fig. 3.

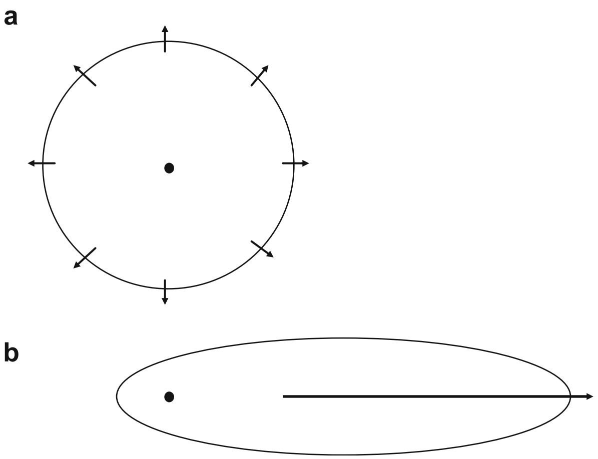


Fig. 3 Concept of preferred radiation direction. In (**a**) and (**b**), an antenna (shown as a *black dot*) is radiating energy. (**a**) The antenna is isotropic, and it radiates energy equally in every direction. The *arrows* show the radiation direction, and their length indicates the amount of power radiated in that direction. It can be seen that energy is radiated equally in each direction. (**b**) There is a preferred radiation direction, shown by the *long arrow*. The radiated energy can be thought of as being like air in a balloon. The balloon can be squeezed to give a nonuniform shape, and this is essentially what has happened in (**b**). Different antenna types "squeeze" the energy in different ways and with differing degrees of focusing or *gain*

The simplest antenna that provides energy in a preferred direction is the Yagi-Uda antenna, named after its two Japanese inventors. The Yagi-Uda antenna is essentially a half-wavelength dipole that has a slightly larger dipole behind it to act as a reflector and smaller elements in front to provide additional gain in the forward direction. A three-element Yagi-Uda antenna is shown in Fig. 4.

A Yagi-Uda antenna is most often employed at frequencies in the UHF (ultrahigh frequency) band, which is from 300 MHz to 3 GHz. At these frequencies, the physical size of this type of antenna is convenient: In the lower bands, the antenna dipoles would become large and somewhat clumsy, while in higher bands, the antenna dipoles would become small and have high losses. Table 1 illustrates the different wavelengths for the popular communication bands.

The *gain* of a Yagi-Uda antenna, and that of any other antenna, is the ratio of the energy transmitted in the particular direction specified to that which an isotropic antenna would provide. By definition, the gain of an isotropic antenna is unity, since it has no preferred radiation direction (see Fig. 3), and the uniform radiation pattern can be thought of as creating a sphere of radiated energy around the antenna. Antenna gain can be expressed as an analog ratio or as a decibel value. The analog ratio is simply the arithmetic value the gain exceeds that of an isotropic antenna, the gain of which we

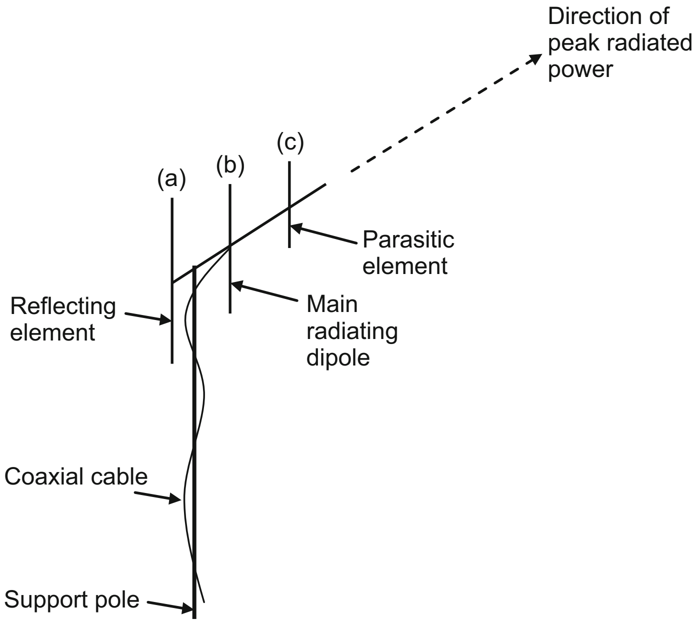


Fig. 4 Simple Yagi-Uda Antenna. The simplest Yagi-Uda antenna, sometimes just called a Yagi Antenna or abbreviated to Yagi, consists of a half-wave, radiating dipole with a reflecting dipole behind it and a parasitic element in front of it. The lengths of the dipoles, and the separation between them, are carefully chosen to maximize the peak radiated power, which is highest in the direction shown. Increasing the number of parasitic elements in front of the main radiating dipole will increase the forward gain and slightly narrow the beamwidth

Table 1 Illustration of frequency versus wavelength for typical communication bands

Frequency	Wavelength	Half-wavelength	Quarter-wavelength
890 MHz	33.71 cm	16.85 cm	8.43 cm
1.5 GHz	20 cm	10 cm	5 cm
6 GHz	5 cm	2.5 cm	1.25 cm
14 GHz	2.14 cm	1.07 cm	0.54 cm
30 GHz	1 cm	0.5 cm	0.25 cm
200 THz	0.00015 cm	0.000075 cm	0.0000375 cm

A frequency of 890 MHz is close to that used by terrestrial mobile wireless systems and over-the-air TV broadcasting services. Satellite mobile services operate around 1.5 GHz (L-band), while the uplinks for fixed-satellite services are at 6, 14, and 30 GHz (C-band, Ku-band, and Ka-band). Free-space optical communications are usually in the infrared region, close to 200 THz

have seen is unity. If the antenna now has some value of gain that exceeds unity in one direction, the radiated energy can no longer be described as a sphere. It is more like a cone of energy symmetrical about the preferred radiation direction, in some respects like the beam of a flashlight. Usually when one speaks of the gain of an antenna, it is normally the maximum gain that is being referenced. This is the gain along the electrical axis, or *boresight*, of the antenna. Since the gain of an isotropic antenna is unity (i.e., there is no preferred radiating direction), the analog gain of any antenna is simply the ratio of the energy radiated in a given direction to that of an isotropic antenna. For example, if the amount of energy radiated in the preferred direction is ten times that which an isotropic radiator would transmit in the same direction, then the analog gain is $10/1 = 10$. For a Yagi-Uda antenna, the maximum analog forward gain is between about 15 and 100, with the higher gain being for an antenna with many more than just three dipoles. The more dipoles there are, the longer the antenna is and the more bulky it becomes. The beamwidth of a Yagi-Uda antenna – the angular dimension between points in the beam either side of boresight where the radiated energy is half the maximum – is fairly broad. Typical values are 18–25°.

Analog values of gain between 15 and 100 that we saw above for the Yagi-Uda antenna are common numbers met in everyday life and so are easy to work with. However, when these analog gain numbers start to get large – values exceeding 10,000 being common for large earth station antennas – satellite engineers resort to the decibel notation to describe antenna gain. Converting an analog value into a decibel value is straightforward. There are two steps. First a logarithm to the base 10 is taken of the analog value, and then this logarithmic number is multiplied by 10 to arrive at the decibel value. An example is shown below for an antenna with an analog gain of 2,000.

- Step 1 – Gain of 2,000 is first converted into a logarithm: $\log_{10}(2{,}000) = 3.3$.
- Step 2 – Logarithm of 3.3 is converted to a decibel value: $3.3 \times 10 = 33$ dB.

Hence, an analog gain of 2,000 is equivalent to a decibel gain of 33 dB. Decibel units are normally written as “dB” and so the decibel gain of this antenna = 33 dB. Antenna systems have become ubiquitous in all aspects of modern life. Small Inmarsat “M” terminals, essentially briefcases with their lids used as an antenna, are employed for satellite news gathering (SNG), and almost everyone has seen video reports from the field using these terminals. UHF and VHF antennas uplink geophysical information (water flow rates, temperatures, humidity, etc.) to low earth-orbiting satellites for onward distribution to meteorological organizations. Anyone who has swiped their credit card at a gasoline station to pump gasoline has (probably without even knowing) used a Ku-band VSAT link to the credit card center for ID verification and billing. USAT antennas are everywhere on tens of millions of homes receiving TV from direct broadcasting satellites. More recently, two-way Internet links at Ka-band using micro-terminals operating in what has become known as SOHO – small office/home office – are becoming widespread. And of course, there is the ultimate USAT, a handheld mobile communication handset that can operate to either low earth orbit (LEO) satellite constellations like Iridium and Globalstar or to geostationary earth orbit (GEO) satellites like Inmarsat. As we saw earlier, a mobile wireless handset usually has unity gain, which has to be compensated for either by having the satellite in LEO, to reduce the path loss between the satellite and handheld device, or by having the satellite use a huge, deployable antenna (probably 12 m in diameter) if the satellite is in GEO. Whenever there is a need to have high gain in a satellite communication link, parabolic reflector antennas are used to focus the energy.

Parabolic Reflector Antennas

A parabola is a geometric shape that has two focal points: one is at infinity and the other is close to the parabola. The concept is illustrated in Fig. 5.

Two principal parameters describe the performance of a parabolic reflector: the gain of the antenna and the *beamwidth* of the antenna. The beamwidth of an antenna is normally the angular distance between two points on the beam either side of boresight where the power has dropped by half (or 3 dB in decibel units) from the maximum. The terms *half-power beamwidth* and *3 dB beamwidth* are used interchangeably. Both the boresight gain and the 3 dB beamwidth of an antenna are directly related to the aperture diameter of the antenna “D,” the wavelength of the signal “λ,” and the efficiency of the antenna “η” as (Allnutt 1989, 2011):

$$\text{Gain} = \left(\frac{\pi D}{\lambda}\right)^2 \times \eta \tag{1}$$

$$\text{Beamwidth} = 1.2 \times \left(\frac{\lambda}{D}\right) \text{radians} \tag{2}$$

Antenna peak forward gain and the 3 dB beamwidth are also closely related, as is shown in Eq. 3:

$$\text{Antenna gain} = \frac{30,000}{(\text{3dB beamwidth})^2} \tag{3}$$

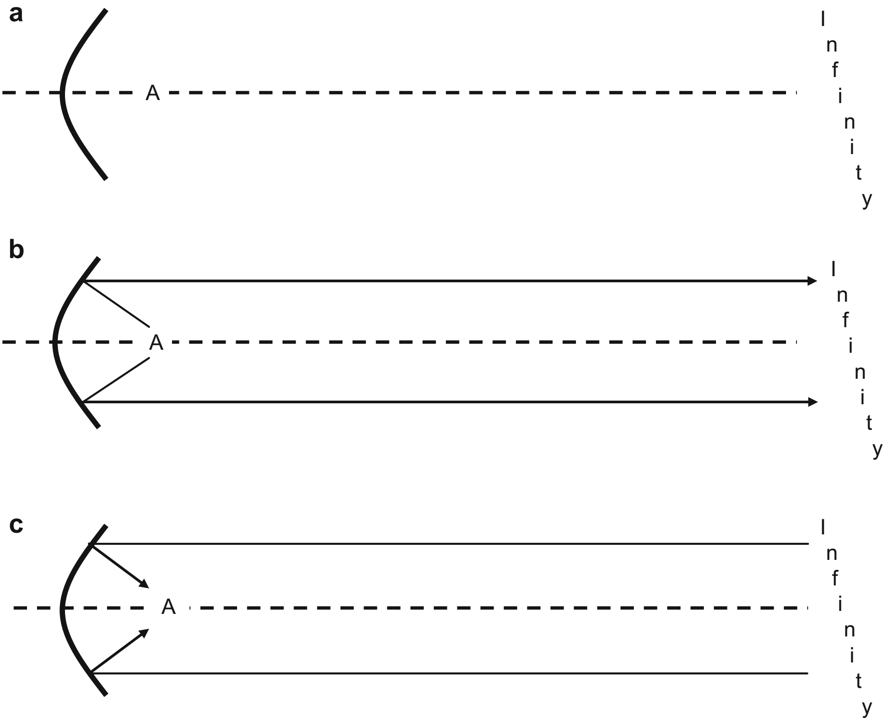


Fig. 5 Illustration of a parabolic antenna. A parabola has two focal points, one at infinity and the other close to the parabola (**a**). If a signal is radiated from focal point "A," then the energy will "bounce" off the surface of the parabolic reflector and be transmitted outward parallel to the antenna axis to infinity (**b**). Likewise, if energy arrives from infinity along the antenna axis, it will arrive at the focal point "A" after being reflected off the parabolic antenna (**c**). To capture the energy at point A or to radiate it, a device known as a *feed* or *feed horn* is used. Coaxial cables or waveguides are attached to the feed to connect the feed to the earth station equipment

Note that the 3 dB beamwidth in Eq. 3 is expressed in degrees. Further, a final formula connecting the 3 dB beamwidth with the wavelength and aperture diameter, this time with the 3 dB beamwidth in degrees, is shown in Eq. 4:

$$3\text{ dB beamwidth} = \left(\frac{75\lambda}{D}\right)\text{degrees} \tag{4}$$

Note that all dimensions are in meters, and Eqs. 2, 3, and 4 are dependent on the shape of the amplitude illumination over the aperture. The equations above assume an aperture distribution between cosθ and $\cos^2\theta$ (see Fig. 7 and Table 2, discussed in the next subsection). The parameter η, the efficiency of the antenna, relates the performance of a perfect antenna to that of a real antenna. The parameter η takes the value of 1 (perfect antenna) through 0 (antenna radiates nothing), but is usually expressed in percentages. Thus, a perfect antenna would have an efficiency of

100 %. For large, standard A, antennas in the fixed-satellite service (FSS) with $D =$ 18 m or greater, the efficiency can be as high as 70 %, since forward gain is maximized. With such large aperture diameters, the *sidelobes* generated by the antennas are very close to the main beam direction – the *boresight* – and so interference into adjacent satellites is virtually nonexistent for correctly pointed antennas. For smaller antennas, particularly very small aperture terminals (VSATs) and ultrasmall aperture terminals (USATs), and micro-terminals, cost is the main determinant, and the efficiency is consequently lower, on the order of 50 % or even less. The boundary between VSAT and USAT is not clear. Most designers generally consider any antenna aperture that is 100 wavelengths across, or less, to be a VSAT. A USAT and a micro-terminal are typically much less than this, and 20 wavelengths is a common size. A mobile handset is even smaller.

Mobile satellite service (MSS) has a significant challenge when it has to compete with terrestrial mobile service. The average distance between a terrestrial base station and a mobile user is less than 20 km. For MSS, the distance can range from about 700 km for LEO MSS satellites to 37,000 km for GEO MSS. The only way the MSS can compete is by offering service to regions where no terrestrial infrastructure exists. Even then, significant EIRP – equivalent isotropic radiated power – has to be generated at the satellite to compensate for the low gain of the user's handset and the huge distances involved. Nevertheless, a number of GEO MSS satellites have been launched and are providing service to large numbers of users. What may make MSS satellites more successful would be the development of so-called smart antennas for user handsets. Smart antennas would employ small phased array elements that would confer two valuable attributes: an ability to steer the beam toward a satellite and a forward gain of about 3–4 dB. A gain of 3–4 dB does not sound much, but 4 dB is a ratio of 2.5. Imagine if the mileage your car obtained increased from 30 to 75 miles per gallon: this is a 4 dB improvement in miles per gallon. Nevertheless, the greatest gain that can be achieved will be by using a parabolic antenna, and for small aperture diameters, this means that the beamwidth will be fairly broad. If the VSAT or USAT terminal is used in a receive-only mode – that is, it does not transmit – then interference into adjacent satellite systems is not a concern. However, if the earth terminal is used in both receive and transmit modes, the sidelobes can illuminate adjacent satellites in geostationary orbit. Suppression of antenna sidelobes is therefore a primary requirement for VSAT antennas that transmit to satellites.

Antenna Sidelobes

In Fig. 5b, it can be seen that energy radiated from the near-in focal point of a parabola by a feed element toward the reflector will be redirected along the main beam axis in a series of parallel lines to infinity. However, this is only true for an infinitely small feed aperture. A feed aperture is generally at least on the order of a wavelength across and usually much more. Clearly, an antenna feed is not infinitely small. Because of this, a *diffraction* pattern will be set up in the energy radiated from the parabolic reflector (Stutzman and Thiele 1998). A diffraction pattern causes a

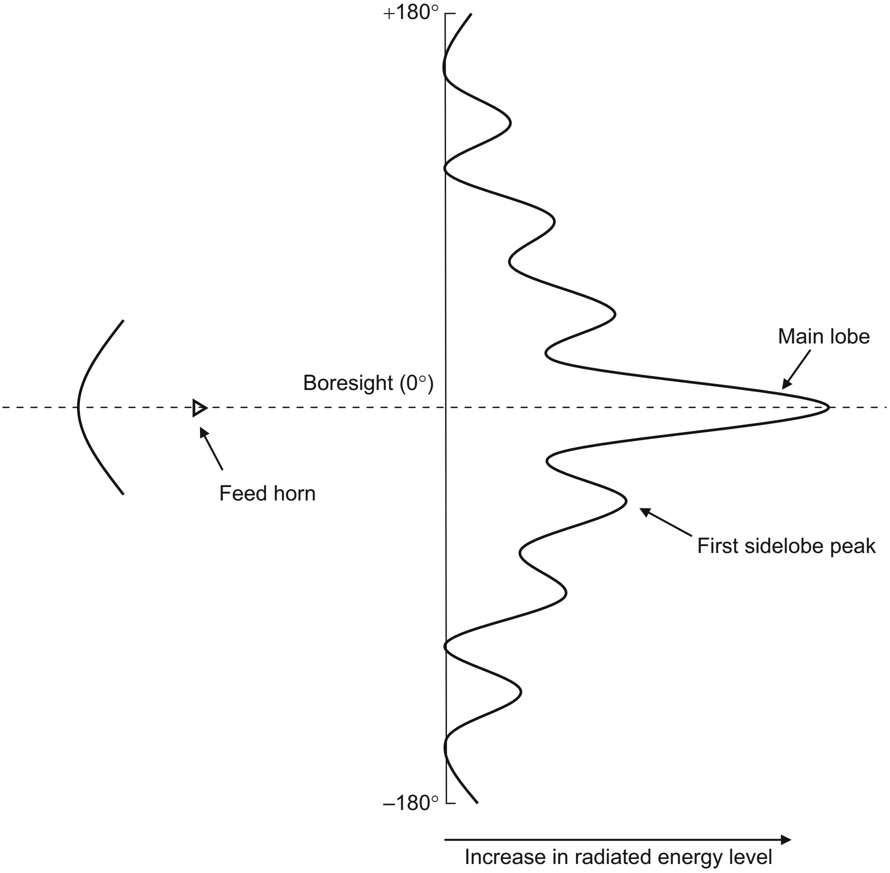


Fig. 6 Schematic of sidelobes from a parabolic reflector antenna in the far field. A diffraction pattern is set up when a source is not infinitely small. There are really three, somewhat ill-defined regions where different diffraction patterns exist. Very close to the reflector antenna, there is a *near-field* region where a constant beam from the antenna is not yet set up and energy levels can fluctuate significantly over small distances (laterally and in the direction of propagation). The next region from the antenna is called the *Frésnel region*. The energy levels still fluctuate significantly, but in a less irregular fashion than in the near field. And finally, there is the *far-field* region, also called the Fraunhofer region, where the diffraction pattern from the antenna is well defined. The far-field radiation levels are shown. It can be seen that there are many sidelobes either side of the main lobe peak. These sidelobes can cause interference into other satellite systems if not suppressed

maximum of energy to be directed along the main axis of the reflector, the boresight direction. Away from boresight, the energy density falls off in a series of ripples, and the peaks of these ripples are called the sidelobes. This is illustrated in Fig. 6. In most antenna systems, there are far more than just the three sidelobes seen in Fig. 6. Paradoxical as it may seem, there is also what is referred to as a *back lobe* where radiated energy exists almost diametrically opposite the peak forward energy – the *main lobe*. The gain (actually it should really be called a loss since it is negative) of a parabolic antenna is generally about −10 dB in the back lobe.

All high-gain antennas that use a parabolic reflector strive to keep the received (or transmitted) signal in phase so that the entire signal is received (or transmitted) coherently across the aperture. The same is not necessarily true of the amplitude distribution: by manipulating this distribution across the antenna aperture, the sidelobe amplitudes can be significantly reduced, thus lowering the potential for off-axis interference. This can be seen by reference to Fig. 7 and Table 2.

Figure 7 shows four common amplitude distributions, and Table 2 compares the beamwidth obtained with these four amplitude distributions with that obtained using an antenna with a uniform amplitude distribution across the aperture. Amplitude distributions across the aperture are not the only consideration when

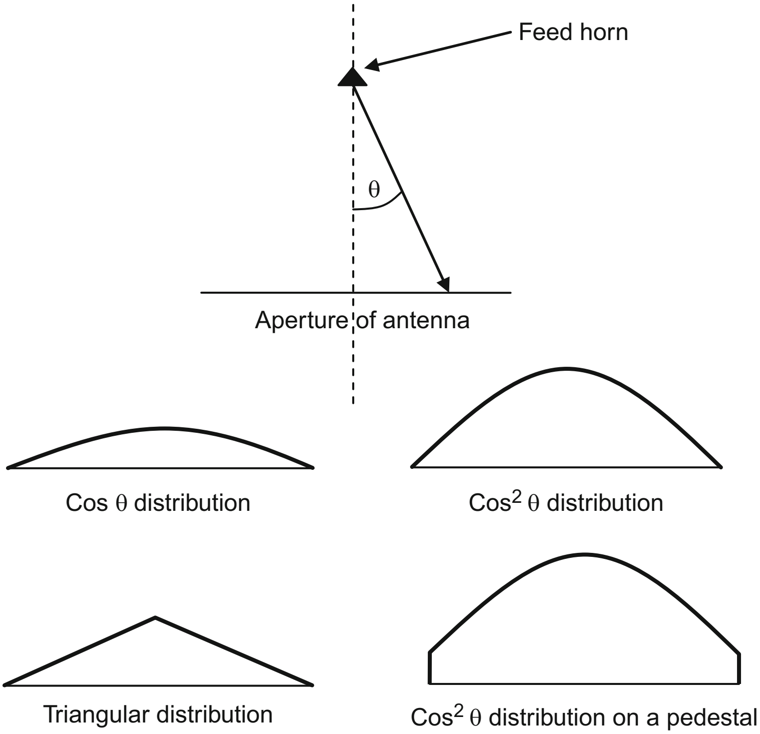


Fig. 7 Illustration of different aperture amplitude distributions. The thin horizontal lines are a cross section of the antenna aperture. The thick lines represent the amplitude distribution across each of the antenna apertures. An increase in energy is shown as an increase in the vertical direction (toward the feed element, not shown). As can be seen, in each of the examples, the maximum power is developed at the center of the antenna aperture. The more energy there is at the center of the antenna aperture compared with the edge of the aperture, the broader the beamwidth, the lower the sidelobe amplitudes, and the lower the peak forward gain. The angle θ refers to the angle between the feed horn and the antenna rim

Table 2 Relative peak sidelobe level to maximum forward gain for some typical aperture distributions (Tim Pratt, 2006, private communication)

Aperture distribution	Beamwidth in degrees	Peak gain relative to uniform distribution	Peak sidelobe level (in dB)
Uniform	51 λ/D	1.0	−13.2
Cosθ	69 λ/D	0.81	−23
Cos ^{2}θ	83 λ/D	0.67	−32
Triangular	73 λ/D	0.75	−26.4
Cos ^{2}θ distribution on a pedestal	63 λ/D	0.88	−26

building up an understanding of antenna systems and how to optimize their performance.

Antenna System Aspects

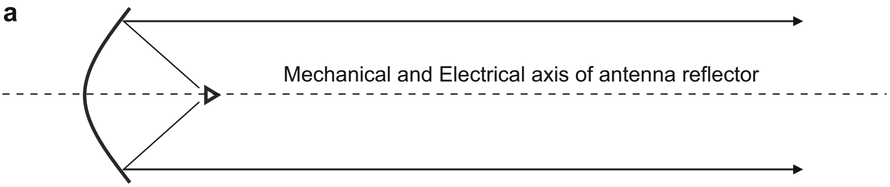
When the parameters of the satellite transponder are known and the link budget has been developed for a range of antennas operating to that satellite, there are some additional design aspects that need to be considered for the overall antenna system. These are considered below.

Blockage

There are two basic ways a parabolic antenna can be configured: on-axis and off-axis. In an on-axis configuration, the feed is located on the main axis of the antenna. In this way, the mechanical axis of the main reflector is the same as the electrical axis. The feed for an on-axis antenna will prevent some of the incoming energy from reaching the main reflector. This is referred to a *blockage*. Blockage will occur on both receive and transmit. To avoid blockage, the feed can be offset from the electrical axis. On-axis and offset-fed (or off-axis) parabolic, front-fed reflectors are illustrated in Fig. 8.

The designs shown in Fig. 8 can also be characterized as single-reflector antennas, that is, there is a single feed and a single parabolic reflector. The focal length, *F*, of such antennas is simply the distance between the feed horn and the parabolic reflector. The longer the focal length of a reflector antenna is, the better the off-axis performance. To create a longer focal length, dual-reflector configurations can be employed. The two principal dual-reflector configurations are Cassegrain and Gregorian, and these designs can be both on-axis (as shown in Fig. 9) and off-axis (as shown in Fig. 10).

Dual-reflector configurations provide easier siting of the feed and receiver, which can both be behind the antenna, thus easing maintenance requirements in terms of accessibility.



Front-Fed, on-axis, parabolic antenna

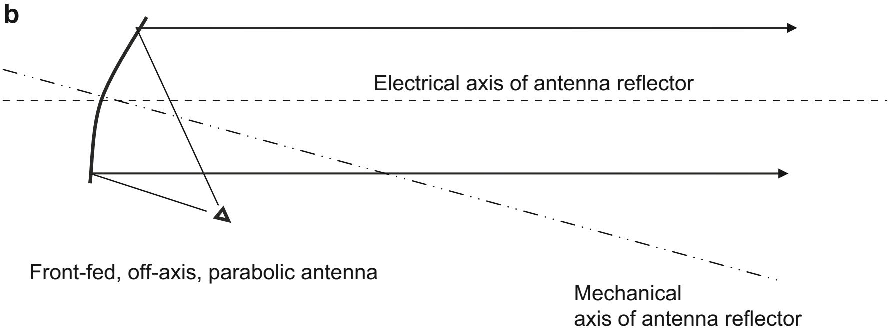
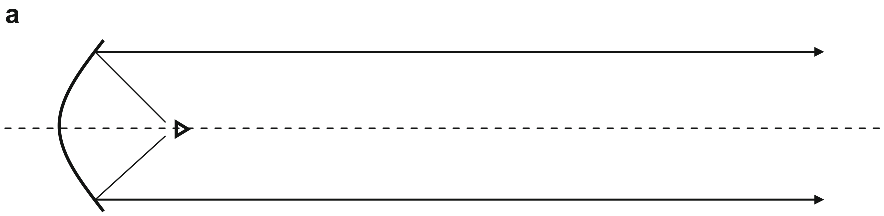


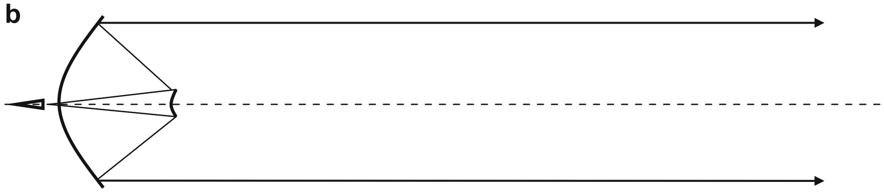
Fig. 8 Illustration of the two main types of parabolic antennas: front fed and offset fed. (**a**) The antenna is symmetrical, with the mechanical axis pointing in the same direction as the electrical axis, the boresight. The feed is placed in the center, in front of the antenna reflector, and so will block both incoming energy and transmitted energy that is on the electrical axis. (**b**) The parabolic shape is not symmetrical, being a section of the reflector to one side of the center of the parabola. The feed is displaced to one side and generates a main beam that is away from the mechanical axis. The feed in (**b**) is offset away from the electrical axis and so causes no blocking of the main beam direction. The reflector surface is still parabolic, but it is essentially a portion of the parabola away from the center of revolution of the parabola shown in (**a**)

System Noise Temperature

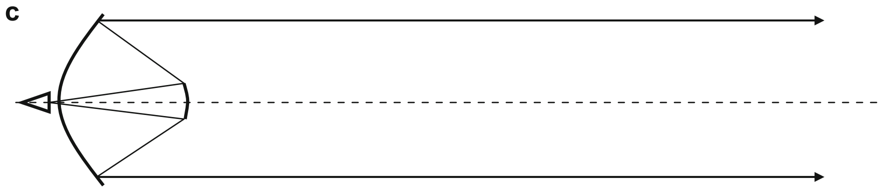
A key parameter in the design of a receiver in a communication link, and possibly *the* key parameter, is the carrier-to-noise ratio at the input to the demodulator (Pratt and Bostian 1986; Pratt et al. 2002). This is often written as C/N, where C is the carrier power and N is the system noise power, both expressed in watts. System noise power $N_{\mathrm{syst}} = kT_{\mathrm{syst}}B$, where $k =$ Boltzmann's constant $= 1.38 \times 10^{-23}\mathrm{J/K} \Rightarrow -228.6$ $\mathrm{dBW/K/Hz}$, T_{syst}, is the system noise temperature in degrees Kelvin, and B is the bandwidth in Hz. The calculation of system noise temperature for the receiving system is not the topic of this chapter, but system noise temperature is made up of two major components: the *internal* noise temperature of the receiving system and the *external* noise temperature. The internal system noise temperature of the receiver is generally constant, with the major components being the noise temperature of the front-end amplifier and the noise temperature induced by the feed run into the low noise amplifier. The external noise temperature is the additive sum of a number of noise



Front-Fed, on-axis, parabolic antenna



Cassegrain, on-axis, parabolic antenna

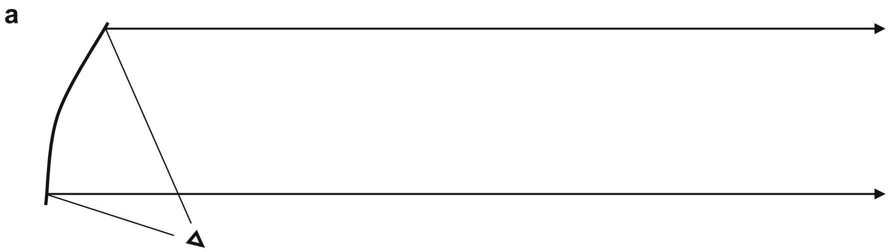


Gregorian, on-axis, parabolic antenna

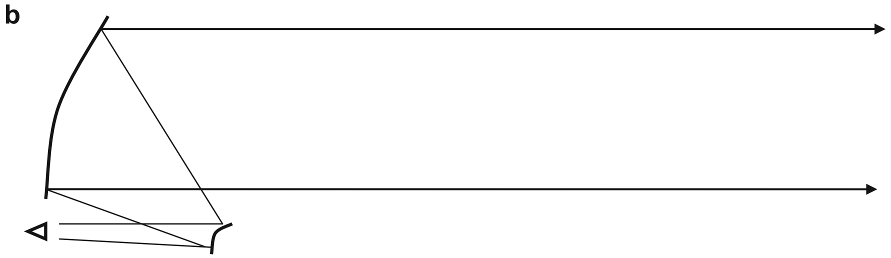
Fig. 9 Illustration of the three types of parabolic antennas that have their main feed system on-axis: front fed, Cassegrain, and Gregorian. All three of the parabolic antenna types shown are symmetrical, on-axis, antennas, with both the feed and the antenna reflectors being rotationally similar about the mechanical axis of the main reflector. In (**a**), the feed is in front of the antenna reflector, while in (**b**) and (**c**), the feed is behind the antenna reflector. Not shown is a hole in the main reflector that permits the energy radiated by the feed in (**b**) and (**c**) to pass through to the sub-reflector antenna and from thence onto the main reflector. Note that the sub-reflector for the Cassegrain antenna is hyperbolic in shape, while that of the Gregorian is elliptical in shape. Also note that the sub-reflector of the Cassegrain antenna has a curvature that is oriented in a similar direction to the parabolic main reflector, while the elliptical sub-reflector of the Gregorian antenna is oriented in an opposite sense to the parabolic main reflector

temperature components entering the antenna feed, and the sum is called the *antenna temperature*.

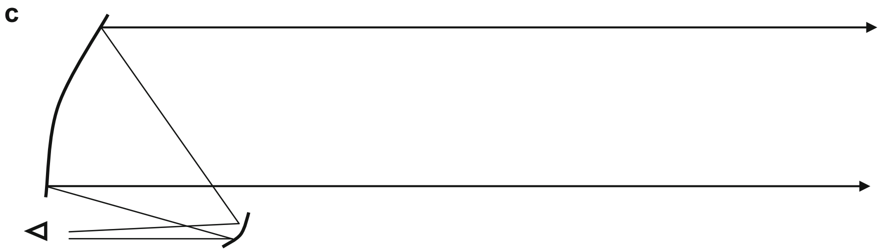
Noise temperature provides an incoherent source of unwanted energy into the receiving system. Since noise energy is not coherent, all components will sum arithmetically rather than as a vector addition. The noise temperature of an antenna will therefore consist of the sum of all noise components external to the antenna that enter via the feed from any direction. These components include the noise remnants from the *big bang* (around 2.8 K when the antenna is not pointed toward a radio star that emits significantly more noise than this or when the antenna is pointed toward the center of the home galaxy, which we know as the *Milky Way*), from gaseous



Front-Fed, off-axis, parabolic antenna



Cassegrain, off-axis, parabolic antenna



Gregorian, off-axis, parabolic antenna

Fig. 10 Illustration of the three types of parabolic antennas that have their main feed system off-axis: front fed, Cassegrain, and Gregorian. All three of the parabolic antenna types shown in are off-axis antennas. In (**a**), the feed is in front of the antenna reflector, while in (**b**) and (**c**), the feed is behind the antenna reflector. As in Fig. 9, the sub-reflector of the Cassegrain antenna has a curvature that is oriented in a similar direction to the parabolic main reflector, while the elliptical sub-reflector of the Gregorian antenna is oriented in an opposite sense to the parabolic main reflector

atmospheric constituents (mainly emission from oxygen and water vapor), and from the ground (the latter mainly through sidelobes that intercept the ground) (Recommendation ITU-R P.676-4 1999). This is illustrated schematically in Fig. 11.

The attenuation produced by oxygen and water vapor on a satellite-to-ground link has been carefully calculated and can be found in International Telecommunication Union (ITU) Recommendation 676 (Seema Sud 2008, private communication). The gaseous attenuation at microwave frequencies is largely absorptive, and the absorbed energy will lead to an enhanced noise temperature. The enhanced noise temperature is related to two parameters of the gaseous cloud: the physical temperature of the gaseous

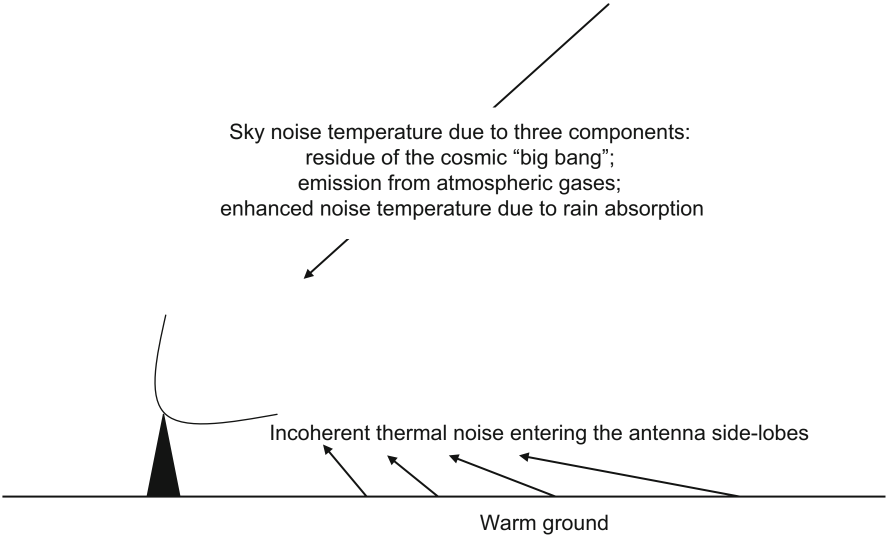


Fig. 11 Schematic of the external noise temperature components that can add to the system noise temperature of the receiver. The noise temperature components from directly in front of the antenna are due to three components: residue from the big bang (2.76 K), emission from oxygen and water vapor in the atmosphere, and enhanced emission from rain clouds. An additional component of noise temperature is that from the ground that is hot. The hot ground noise temperature enters through the antenna’s sidelobes

cloud and the specific transmissivity, σ, of the gaseous cloud. The parameter σ varies between zero (i.e., nothing is transmitted through the gaseous cloud) and one (i.e., everything is transmitted through the gaseous cloud). This is illustrated in Fig. 12.

The apparent sky-noise temperature of the gaseous cloud, T_{sky}, can be directly calculated from a knowledge of the physical temperature of the gaseous cloud, T_m, by:

$$T_{\text{sky}} = (1 - \sigma) \times T_m \tag{5}$$

For example, if the gaseous cloud causes 0.2 dB of attenuation, we can convert 0.2 dB to an analog value ⇒ 1.047. Inverting this analog value provides the value of $\sigma = 0.95$ and $(1 - \sigma)$ yields 0.05. If $T_m = 280$ K, then the value of $T_{\text{sky}} = 0.05 \times 280 = 14$ K. In this instance, a loss of 0.2 dB along the path due to atmospheric gaseous absorption gives rise to an increase in the antenna temperature contribution of 14 K.

Exactly the same increase in sky-noise temperature will occur in the presence of rain attenuation. If a rain cloud causes 3 dB attenuation ⇒ an analog loss of 2, which inverted gives a value of $\sigma = 0.5$, then $(1 - \sigma) = 0.5$. If the rain cloud is at a physical temperature of 280 K, then an additional 140 K will enter the feed of the antenna as an enhanced noise temperature contribution due to the presence of rain in the path. *This is a critical point to remember in antenna system design.*

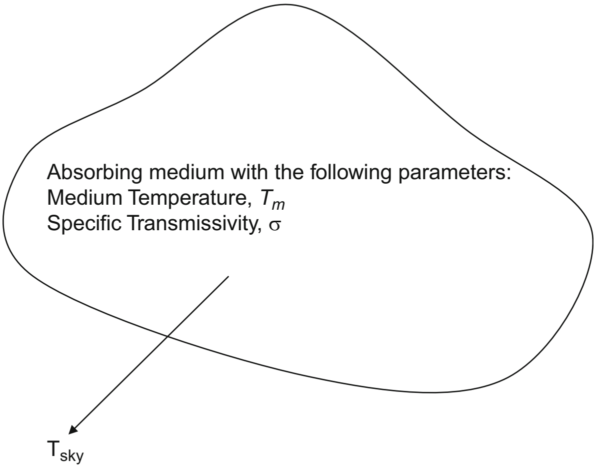


Fig. 12 Illustration of thermal emission due to absorption. An absorbing medium, which in this case is a rain cloud, causes a signal passing through it to lose some energy. If the signal entering the rain cloud has a power, S, then the signal leaving the rain cloud is $\sigma \times S$, where σ is the specific transmissivity of the rain cloud. Since energy is being absorbed by the rain cloud, it will emit at an enhanced temperature compared with clear sky, which is given by $T_{\text{sky}} = (1 - \sigma) \times T_m$

If the system noise temperature of the receiving system is, say, 140 K in clear sky, then under the rain attenuation in the example above, the total noise temperature will be 280 K, made up of 140 K from the receiver and 140 K from the noise temperature emitted from the rain cloud. The system noise temperature has doubled under rainy conditions (a 3 dB increase) at the same time that the received signal power went down by 3 dB. The *C/N* has therefore dropped by 6 dB under a rain attenuation of 3 dB. It should always be remembered that when calculating the system fade margin appropriate for the percentage time, the receiving system should provide adequate *performance* and *availability*. Performance and availability are not the same when considering communication systems. Performance is the level of service, often measured as the bit error rate (BER), delivered for a very high percentage of the time, generally 99 % or higher. Availability is the time the service is available above a usable threshold, again often measured in BER. For example, the performance level of a typical Ku-band digital satellite communication link will be set at a BER of 10^{-8} to 10^{-10} for at least 99 % of the time, while the availability would be a BER of 10^{-6} for 99.7 % of the time. For the early, and even some of the current, C-band (6/4 GHz) satellite communication systems, achieving a BER of better than 10^{-6} is very difficult, but the propagation impairments are not significant, and so the availability margin is easily met. Conversely, because Ku-band (14/11 GHz) and Ka-band (30/20 GHz) satellite communication systems have to provide fairly large fade margins to overcome rain attenuation, achieving a BER of 10^{-10} in clear-sky conditions is relatively straightforward.

Achieving the availability margins is a different matter altogether due to significant rain attenuation at Ku-band and above.

Rain Attenuation

Rain attenuation becomes significant at frequencies above about 10 GHz (Recommendation ITU-R P.618-9 2007). Rain attenuation is not closely correlated with the *total* amount of rainfall that accumulates on the ground over a relatively long period of time (e.g., 2 h), but it is directly correlated with the *rate* at which the rain falls. Rain fall rate "R" is measured in millimeters per hour (mm/h), and the rainfall rate is related to the *specific attenuation* "γ" by Eq. 6:

$$\gamma = kR^{\alpha}\mathrm{dB/km} \tag{6}$$

Specific attenuation is a term used to describe the amount of path attenuation, in decibels, experienced over a kilometer, hence the units dB/km. The parameters k and α are given in Table 3 for both vertical and horizontal polarization (Recommendation ITU-R P.838-1 1999).

The subscript "H" refers to horizontal polarization, and "V" refers to vertical polarization. The parameters γ is the specific attenuation in dB/km.

The rainfall rate, R, is usually measured on the ground by a rain gage. It is therefore the rainfall rate at a point. Attenuation on a satellite-to-ground link, however, takes place along the path through the rain. The rainfall rate along this path is not uniform, but Eq. 6 assumes a uniform rainfall rate. The key step, therefore, in moving from specific attenuation, given by Eq. 6, to path attenuation (i.e., the total attenuation experienced along the path due to rain) is to find the equivalent distance "L" over which the rainfall rate can be assumed to be constant.

Table 3 Regression coefficients for estimating the attenuation coefficients for specific attenuation, γ, where $\gamma = kR^{\alpha}$ (From Table 1 of reference (Recommendation ITU-R P.838-1 1999))

Frequency (GHz)	k_H	k_V	α_H	α_V
1	0.0000387	0.0000352	0.912	0.880
2	0.000154	0.000138	0.963	0.923
4	0.000650	0.000591	1.121	1.075
6	0.00175	0.00155	1.308	1.265
7	0.00301	0.00265	1.308	1.312
8	0.00454	0.00395	1.327	1.310
10	0.0101	0.00887	1.276	1.264
12	0.0188	0.0168	1.217	1.200
15	0.0367	0.0335	1.154	1.128
20	0.0751	0.0691	1.099	1.065
25	0.124	0.113	1.061	1.030
30	0.187	0.167	1.021	1.000

Table 4 Examples of path attenuation for three frequencies. In the examples given below, the rainfall rate, R, is 50 mm/h, the effective pathlength, L, is taken as 4 km, and vertical polarization is assumed

Frequency (GHz)	$\gamma = (kR^{\alpha}) \times L$	Path attenuation (dB)
1	$0.0000352\ (50)^{0.880} \times 4$	0.004
10	$0.00887\ (50)^{1.264} \times 4$	5.0
20	$0.0691\ (50)^{1.065} \times 4$	17.8

The ITU-R prediction method (Seema Sud 2008, private communication) describes a procedure to calculate the effective pathlength L. The total path attenuation will therefore be given by:

$$\gamma = (kR^{\alpha}) \times L \text{ dB} \tag{7}$$

ITU-R Recommendation 618 (ITU-R Recommendation P.618-9 2007) uses the point rainfall rate measured for 0.01 % of an average year to calculate the path attenuation experienced for an average year at the percentage time 0.01 %. This value is then extrapolated to lower, or higher, time percentages. It can be seen from Table 3 that the specific attenuation is very low for UHF frequencies but increases rapidly above a frequency of about 10 GHz. Between about 10 and 30 GHz, the path attenuation can be very approximately scaled as the square of the frequency, using dB to describe the attenuation. For example, if the path attenuation at 10 GHz is 3 dB, the scaled path attenuation along the same link for the same time percentage at 20 GHz is simply given by the ratio of the square of the frequencies. That is, the attenuation at 20 GHz $= (20/10)^2 \times$ (the attenuation at 10 GHz) $= 4 \times 3 = 12$ dB. In this example, the attenuation at 20 GHz was 12 dB when the attenuation at 10 GHz was 3 dB. Table 4 gives some other examples using the parameters k and α from Table 3.

Note that the more accurate formulation of path attenuation gives the attenuation at 20 GHz to be 3.56 times the attenuation at 10 GHz rather than a factor of 4 that would be given by the simple frequency squaring formula.

The effect of polarization is not critical for path attenuation, although attenuation for vertical polarization is generally less than or equal to that for horizontal polarization. However, the effect of rain and ice crystals on the polarization of a signal can be significant.

Depolarization

The polarization of an electromagnetic signal is given by the orientation of the electric vector, given the symbol E. There is always a magnetic field associated with any electric field, and this is commonly given the symbol M. Since both E and M propagate together, this is why the research area is called electromagnetics. The E and M fields are oriented at right angles to each other (i.e., *normal* or *orthogonal* to each other). They are also mutually orthogonal to the direction of propagation of the electromagnetic wave. This is illustrated in Fig. 13.

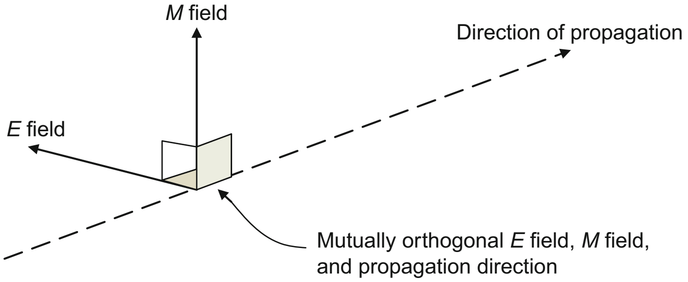


Fig. 13 Schematic of a linearly polarized electromagnetic signal, shown as *E* and *M* fields, propagating in the given direction. In the example above, the *E* and *M* fields are always oriented in the same direction, unless acted upon by an external medium (such as rain) or ionized media (such as the ionosphere). This form of polarized signal is referred to as a *linearly* polarized signal. If the electric (*E*) field is vertical to the local horizontal direction, then the signal is said to be *vertically* polarized. Similarly, if the *E* field is parallel to the local horizontal direction, the signal is said to be *horizontally* polarized

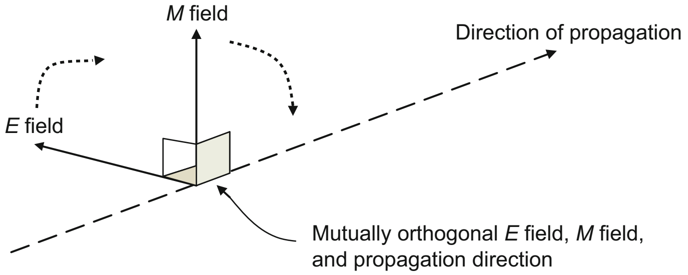


Fig. 14 Schematic of a circularly polarized electromagnetic signal, shown as *E* and *M* fields. In the example above, the *E* and *M* fields are rotating about the direction of propagation, while maintaining their mutually orthogonal orientation with respect to each other and the propagation direction. This form of polarized signal is referred to as a *circularly* polarized signal. In the example shown, the *E* and *M* fields are rotating in a clockwise direction with respect to the direction of propagation. This form of circular polarization is known as *right-hand circular polarization* (*RHCP*). If the *E* and *M* fields are rotating in the other direction, they form what is known as *left-hand circular polarization* (*LHCP*)

The example shown in Fig. 13 is for a *linearly* polarized signal. That is, the direction of the electric (and thus magnetic) field is always oriented in the same direction, unless acted on by an external medium (e.g., rain or ice crystals) or by an ionized medium, such as the ionosphere. An electromagnetic signal can also be launched in such a way that the electric (and magnetic) fields are not oriented in a constant direction, but are rotating about the direction of propagation. This is shown in Fig. 14.

The *E* and *M* fields can rotate about the propagation direction in one of two directions. If the propagation direction is away from you and the fields are rotating in a clockwise direction as viewed from your perspective, the polarization sense is called right-hand circular polarization (*RHCP*). If the fields are rotating in the opposite direction to this, then the polarization sense is called left-hand circular polarization (*LHCP*).

It is very rare that a polarization is "pure," that is, there is no residual energy in the opposite polarization sense. The general polarization state is elliptical polarization. There are two special cases for elliptical polarization: circular and linear. This is illustrated in Fig. 15.

Linearly polarized signals can be resolved into two orthogonal orientations (usually linear vertical and linear horizontal). For a purely polarized signal, there is no component in the orthogonal sense. If the signal is not a purely polarized signal or if the signal encounters a propagation medium that causes the signal to lose its purity of polarization (i.e., exhibit a depolarized component), then a signal component will be apparent in the orthogonal sense. This is illustrated in Fig. 16.

In Fig. 16a, the linearly polarized signal, shown as a vertically oriented linear signal, L_V, has no component in the orthogonal direction, L_H. The pure, linearly

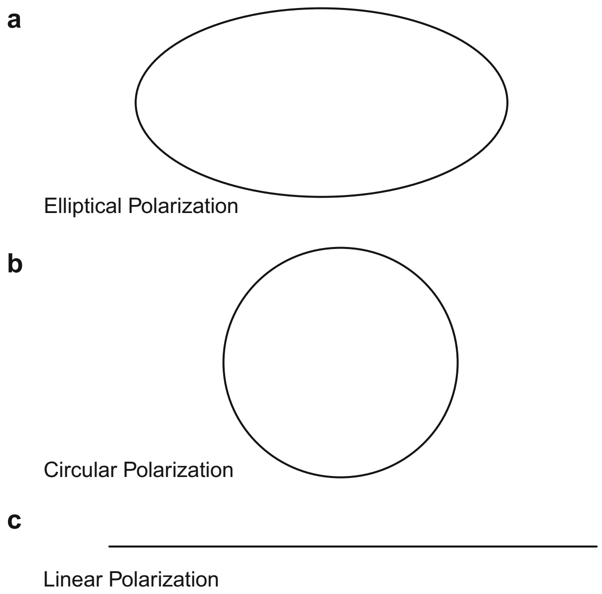


Fig. 15 Three examples of signal polarization. The general case for the polarization of a signal is elliptical polarization (**a**). The elliptical shape forms the locus of the electric (or magnetic) field vectors. There are two special cases for elliptical polarization: circular and linear. (**b**) The two axes of the ellipse (the semimajor and the semiminor) become the same and are equal to the radius of the circle. This is an example of pure circular polarization. By "pure" we mean that there is no component in the opposite sense of the polarization. The circle could represent a pure RHCP or pure LHCP signal, depending on the rotation direction of the signal with respect to the propagation direction. (**c**) Indicates the case when the semiminor axis of the ellipse becomes zero, and the entire signal is represented by the semimajor axis. Since there is no component of the signal in the orthogonal sense, the signal is said to be a pure, linearly polarized signal

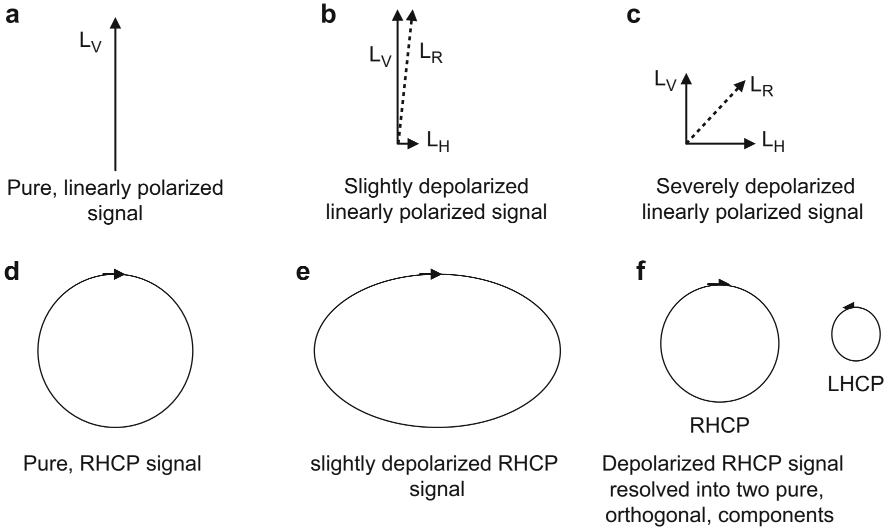


Fig. 16 Schematic representation of depolarization for linearly polarized and circularly polarized signals

polarized signal then encounters something in the propagation medium (e.g., rain), and some of the energy in the vertically oriented signal is *depolarized* into the horizontal direction, L_H. This is shown in Fig. 16b. The resultant signal that emerges from the propagation medium, L_R, is the vector addition of the two orthogonal components, L_V and L_H. A receiver that is set up to receive both polarization senses (linear vertical and linear horizontal) will therefore receive two components of the original signal: one in the originally polarized sense (L_V) and one in the oppositely polarized, or depolarized, sense (L_H). If another signal is supposed to be entering the receiver in the L_H sense, then it will encounter interference from the depolarized component of the original L_V signal. Figure 16c shows a case of severe depolarization where half of the received energy from the original L_V signal appears as an L_H signal in the receiver.

In Fig. 16d, the circularly polarized signal, shown as a right-hand circularly polarized (RHCP) signal, has no component in the opposite sense, left-hand circularly polarized (LHCP). The pure RHCP signal then encounters something in the propagation medium (e.g., rain or ice crystals), and some of the energy in the RHCP sense is depolarized into the LHCP sense, leading to an elliptically polarized signal, shown in Fig. 16e. If the receiver is set to receive signals in both RHCP and LHCP, it will receive an LHCP component that has been depolarized from the original RHCP signal. This will cause interference in the receiver. In Fig. 16f, the elliptical polarization shown in Fig. 16e has been resolved into two orthogonal circularly polarized senses, RHCP (the wanted signal) and LHCP (the unwanted signal).

Telecommunication systems need to be able to discriminate between the wanted polarization, which carries the information signal, versus the unwanted polarization, which carries the residual, depolarized signal. For a dual-polarized receiver that is set up to receive two different signals at the same frequency, but in oppositely polarized senses, it is essential that the energy in the wanted signal is well above that of the interfering signal, which has been depolarized into the wanted signal's channel from the other polarization sense. The term *cross-polarization discrimination* (XPD) is used to describe the power difference between the wanted polarization and the interfering signal. If the rotation of the wanted signal (we will assume it is L_V) in Fig. 16b to the resultant signal (L_R) is $\Delta\theta$, the cross-polarization discrimination, XPD, is given by the equation $\text{XPD} = -20 \log_{10} \tan(\Delta\theta)$. The minus sign is used by convention to produce a positive value of decibels. The multiplier 20 is used instead of 10 because the XPD is a power ratio and it is necessary to square the electric fields. For example, if the apparent rotation $\Delta\theta = 1°$, the $\text{XPD} = 35$ dB. An XPD of 35 dB means that more than 1,000 times more power is in the wanted orientation than in the unwanted orientation. This is an excellent value of XPD for a dual-polarized system. Typically, rain and ice crystals can cause significant depolarization along a path. A minimum operating XPD for digital communication systems is about 12 dB.

The apparent rotation of the linear vector in Fig. 16 is caused by two effects in the propagation medium, differential phase, and differential attenuation. *Differential* in this case means the difference in the level of the phenomenon between the two polarizations. In Fig. 16b, if the propagation medium has a different attenuating effect in the linear vertical polarization to that in the linear horizontal polarization, this is referred to as *differential attenuation*. Similarly, if there is a phase difference between the two linearly polarized vectors in Fig. 16b, this is referred to as a *differential phase*. Differential phase effects are dominant at C-band (6/4 GHz), since the attenuation in rain is very small to begin with. As the frequency increases to Ku-band (14/11 GHz) and to Ka-band (30/20 GHz), rain attenuation starts to dominate and differential attenuation is the primary depolarization mechanism. The change from a differential phase-dominated depolarization mechanism to a differential attenuation-dominated depolarization mechanism as the carrier frequency used increases from 4 to 30 GHz has an interesting system effect. For each decibel of attenuation at C-band, there is a much higher resultant depolarization effect than at Ku-band and especially at Ka-band. The result is that depolarization is the dominant performance and availability limiting parameter at C-band, while depolarization can be largely ignored as a limiting phenomenon at Ka-band: attenuation effects dominate the performance and availability margins at Ka-band. The margin provided to account for signal loss and depolarization in adverse propagation conditions is also significantly affected by the choice of modulation.

The choice of whether to use linear or circular polarization for an operational system involves a number of parameters to be considered. Linearly polarized antenna feeds are much simpler to design and build and are thus cheaper. They also have generally much better on-axis XPD properties. A good dual-polarized linearly polarized antenna system can generally achieve a clear-sky XPD of 30 dB without much difficulty, and more than 40 dB can be achieved. A comparable circularly polarized antenna system is normally limited to 27 dB, unless extraordinary care is taken in the design and construction. Circularly polarized antennas, particularly small earth station antennas, do not need to have their feed system aligned to the orientation of the satellite signal. The receivers are also largely unaffected by rotation of the electric vector, particularly due to Faraday rotation at C-band. Geostationary satellite antennas appear to have better off-axis properties when circular polarization is employed.

Another factor to consider is the *rms* surface tolerance of the antenna being used for dual-polarization operation. To achieve adequate to good performance in single-polarization operation, an rms tolerance of about a quarter of a wavelength is required for the antenna reflector surface. If dual-polarization operation is contemplated, the rms surface tolerance must be on the order of a tenth of a wavelength. The better the rms surface tolerance, the higher the manufacturing costs will be.

Modulation

Modulation is the technique used to modify one or more parameters of the transmitted signal so that information can be placed on the carrier. More importantly, it will permit the information to be retrieved at the receiver using a process called demodulation. There are a number of techniques used to modulate a digital carrier, the principal three being amplitude-shift keying (ASK), frequency-shift keying (FSK), and phase-shift keying (PSK). There are also combinations of the modulation techniques where, for example, both amplitude and phase are used to define a symbol, and these forms of modulation have the generic name *quadrature amplitude modulation* (QAM).

A key point to realize in digital communications is that the information is *not* being sent as bits (ones and zeros) between the transmitter and the receiver: it is being sent as *symbols*. A symbol is a particular state that is impressed onto the carrier signal: it can be a change in level for ASK, a change in frequency for FSK, or a change in phase for PSK. A simple formula that connects the number of symbol states M in a digital modulation scheme with the number of bits that are used in each symbol, n, is:

$$M = 2^n \tag{8}$$

For example, if you are using binary phase-shift keying (BPSK), there is 1 bit for every symbol (i.e., $n = 1$), and so the total number of symbol states is $2^1 = 2$. For quadrature phase-shift keying (QPSK), $n = 2$ and so $M = 4$, that is, there are four symbol states. These states can be expressed in bit form as 00, 01, 10, and 11. The more bits there are per symbol, the smaller the occupied bandwidth becomes. However, there is a downside to having a smaller occupied bandwidth: the more bits there are per symbol, the more carrier-to-noise power the systems need to develop to provide the same BER. Table 5 illustrates the increase in raw power needed to provide the same BER for a given modulation.

Table 5 Carrier to noise required for m-QAM (Tim Pratt, 2006, private communication)

Modulation	Bits/symbol	C/N for BER $= 10^{-6}$ (dB)	Relative bandwidth[a]
BPSK	1	10.6	1.0
QPSK	2	13.6	0.5
16-QAM	4	20.5	0.25
32-QAM	5	24.4	0.20
64-QAM	6	26.6	0.17
256-QAM	8	32.5	0.125
1024-QAM	10	38.5	0.10

[a] Relative bandwidth means the bandwidth occupied relative to a modulation of BPSK

Clearly, there is a trade-off between occupied bandwidth and the *C/N* required to develop the required BER. In addition, as the frequency increases, so does the level of path attenuation in a given rain event, and so achieving the same BER for the same percentage of the time at Ka-band as was achieved at C-band will require a much larger fade margin. And, as we have seen, as the frequency increases, there will be a concomitant increase in the perceived antenna temperature under rain conditions. In many system designs, a useful combination parameter is used to characterize an earth station designs: *G/T*. The parameter *G/T* is the ratio of the gain of the antenna divided by the system noise temperature of the receiver.

G/T

The link budget for a satellite-to-ground link can be expressed as shown in Eq. 9:

$$\frac{C}{N} = \frac{P_t G_t G_r}{k T_s B} \left[\frac{\lambda}{4\pi R} \right]^2 \tag{9}$$

where (C/N) is the carrier-to-noise ratio, P_t is the power transmitted in watts, G_t is the gain of the transmitting antenna, G_r is the gain of the receiving antenna, λ is the wavelength of the signal in meters, k is Boltzmann's constant, T_s is the system noise temperature in Kelvin, B is the bandwidth of the receiver in Hz, and R is the distance between the transmitting and receiving antennas in meters. It is useful to note in Eq. 9 that (C/N) is proportional to (G_r/T_s). Increasing the receiving antenna gain, G_r, will increase C/N, and reducing the system noise temperature, T_s, will also increase C/N, and vice versa, of course.

The external noise temperature – the antenna noise temperature – of the earth station system will vary with the perceived noise temperature emitted by constituents along the path to the earth station. The antenna noise temperature will therefore vary with the total pathlength through the atmosphere. If the elevation angle of the earth station is reduced so that it can operate with a different satellite, the total pathlength through the atmosphere will increase, as shown in Fig. 17. If the pathlength increases through the atmosphere, the total absorption will increase due to additional

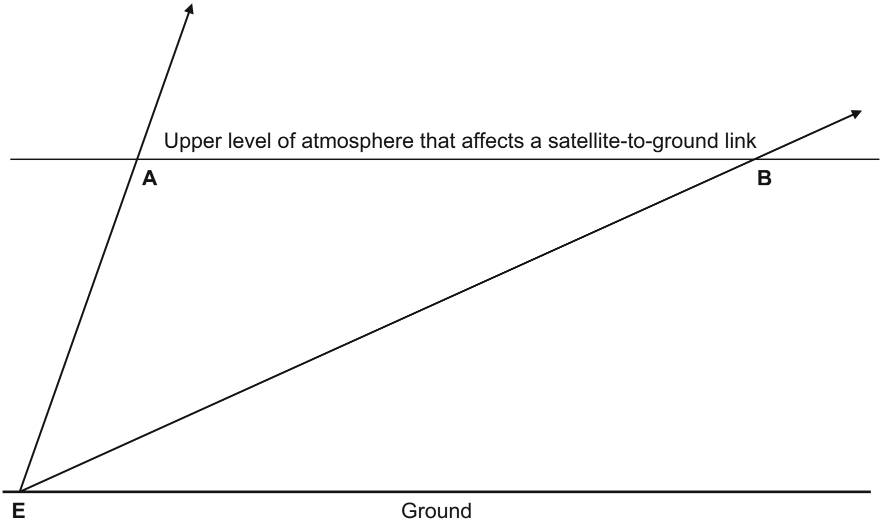


Fig. 17 Schematic showing the change in pathlength through the atmosphere. A high elevation angle link from the earth station, E, exits the atmosphere at A, and a relatively low elevation angle link from an earth station, E, exits the atmosphere at B. Path EA is shorter than path EB. Since the specific transmissivity, σ, of the atmosphere will decrease as the elevation angle is reduced, the resultant noise temperature emitted by the sky will increase as the elevation angle reduces. This can be seen in Fig. 14

constituents in the path. As a result, the perceived sky noise will also increase. This is shown schematically for a typical standard A earth station in Fig. 18.

As the elevation angle of a large earth station antenna is reduced, additional problems occur that cause tracking of a satellite to become more difficult.

Tracking

A communication satellite is considered to be in geostationary orbit if it is at geostationary altitude with an eccentricity of ≤ 0.001 and an inclination of $\leq 0.05°$. For a geostationary satellite, orbital height + earth radius = 35,786.03 km + 6,378.137 km (average) = 42,164.17 km. The *station-keeping box* for a geostationary satellite can therefore be seen to be ±0.05° east-west and north-south. Using Pythagoras' equation, the largest movement of the satellite in this box is 0.14°, from one corner of the box to the other, diagonally opposite, corner. If the earth station has a 1 dB beamwidth that is smaller than this, then the earth station will have to use tracking. (The 1 dB beamwidth is approximately half that of the 3 dB beamwidth.)

Tracking can be active (i.e., the use is made of the incoming signal from the satellite to update the pointing of the antenna) or passive (i.e., the use is made of the satellite ephemeris data to predict the position of the satellite and software code used to passively point the antenna toward the predicted position of the satellite). The cheapest form of

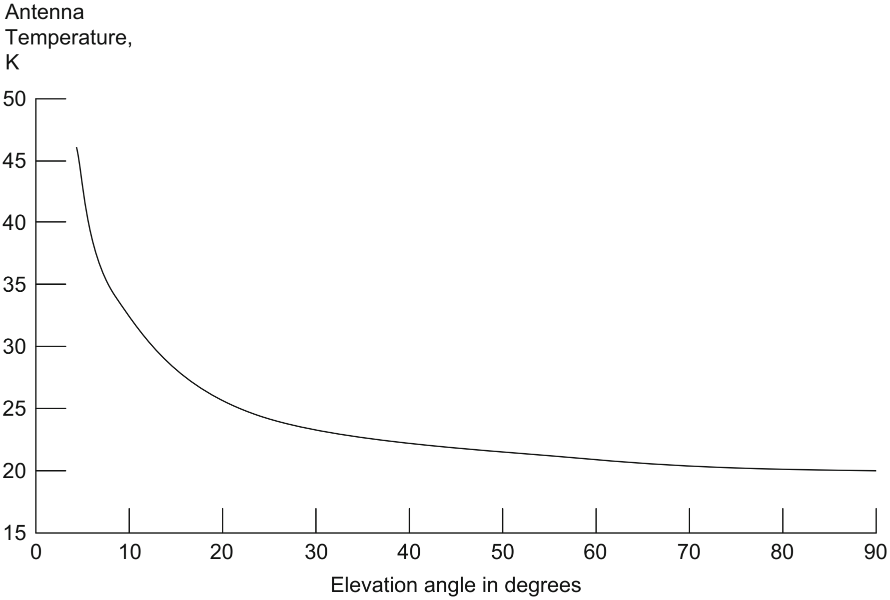


Fig. 18 Plot of a typical standard A antenna noise temperature versus elevation angle. The increase in the antenna noise temperature as the elevation angle reduces is due to two principal effects: (**a**) an increasing number of the antenna sidelobes intercept the ground, which is often at a temperature well above freezing, and (**b**) the path through the atmosphere becomes longer, and so the absorption of the gaseous constituents leads to a concomitant increase in the noise temperature of the sky that is picked up by the antenna (see Fig. 11)

tracking used initially for earth stations operating to geostationary satellites was *step tracking*, sometimes called *sequential lobing* or *hill climbing*. In this form of tracking, the satellite is initially acquired under manual control. The earth station tracking is then put under automatic control. The automatic controller then waits a given interval (15 min, sometimes longer), and the antenna is steered a given amount east and west about the nominal position of the satellite and then north and south of that same position. These movements do not lose the satellite signal, since the angular movement is small. The antenna is then steered back to the point where the signal appeared to be a maximum. While this form of tracking works for targets that are moving very slowly (like a geostationary satellite), problems start to occur when the elevation angle becomes relatively low, especially below 15°. Below 15°, and especially as the elevation angle gets close to 5°, clear air propagation effects become increasingly significant. These effects can be summarized as ray bending, defocusing, angle of arrival, atmospheric multipath, antenna gain reduction, tropospheric scintillation, and low-angle fading. The cumulative effect of these propagation problems is to prevent step-tracking antennas from operating effectively. Low-cost program tracking can overcome much of these propagation problems that affect step tracking, although they will not reduce the effect of the propagation impairments.

Active tracking that is used for some earlier radar systems is *conical scan*, but for large earth stations operating to geostationary satellites, the best form of tracking is *monopulse* tracking. Conical scanning requires the main beam of the antenna to be spun about its mechanical axis, forming a "cone" around the target, and if the target moves, the energy difference between the sides of the cone allows corrective action to be taken. This type of tracking always receives a lower signal than that which would be received on-axis and so is not employed in satellite systems, where received power has to be maximized. Monopulse tracking, so called because it derives the pointing commands from one pulse (if it is radar) or one set of input signals received at the same time (such as in satellite antenna systems), uses four sets of input signals to develop sum and difference channels. Figure 19 illustrates the principle.

Monopulse tracking is the most accurate form of tracking available to earth station antenna systems. Whether the tracking is absolutely precise or somewhat relaxed, it is very likely that the earth station will be located within a region where

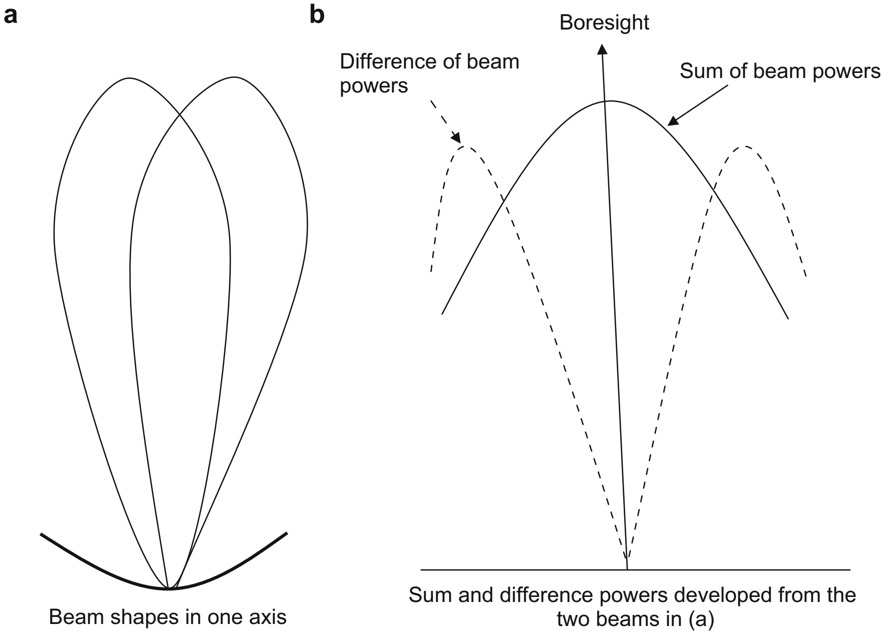


Fig. 19 Schematic representation of one axis of monopulse tracking. A monopulse antenna, in its simplest form, consists of four feed horns close to the focus of the parabolic main reflector. Two of the feed horns are orthogonal to the axis of the other feed horns. One of the axes is depicted. (**a**) The shape of the two main lobes created by the two feeds is shown, with the angular separation of the two beams exaggerated for clarity. (**b**) The power of the sum beam (adding the two beam powers together) and the difference beam (subtracting the two beam powers from each other) is shown. It can be seen that the antenna has only to move off-track by a very small amount for the power in the difference beam to increase significantly. The feedback tracking loop seeks to minimize the difference beam at all times

other systems operate on similar frequency bands, and so it may be necessary to protect the earth station antenna with what is known as *site shielding*.

Shielding

Earth stations often have to be sited in areas where there already exists a significant interference potential, not just from other satellite systems but from terrestrial systems. An example of interference into a satellite earth station from a terrestrial source is shown in Fig. 20a.

Sitting of earth stations is closely controlled by the national organization of the country which the earth station is to be sited in. For the USA, this is the Federal Communication Commission (FCC). The control process is referred to as *coordination*. New antenna or satellite systems are required to *coordinate* with all preexisting

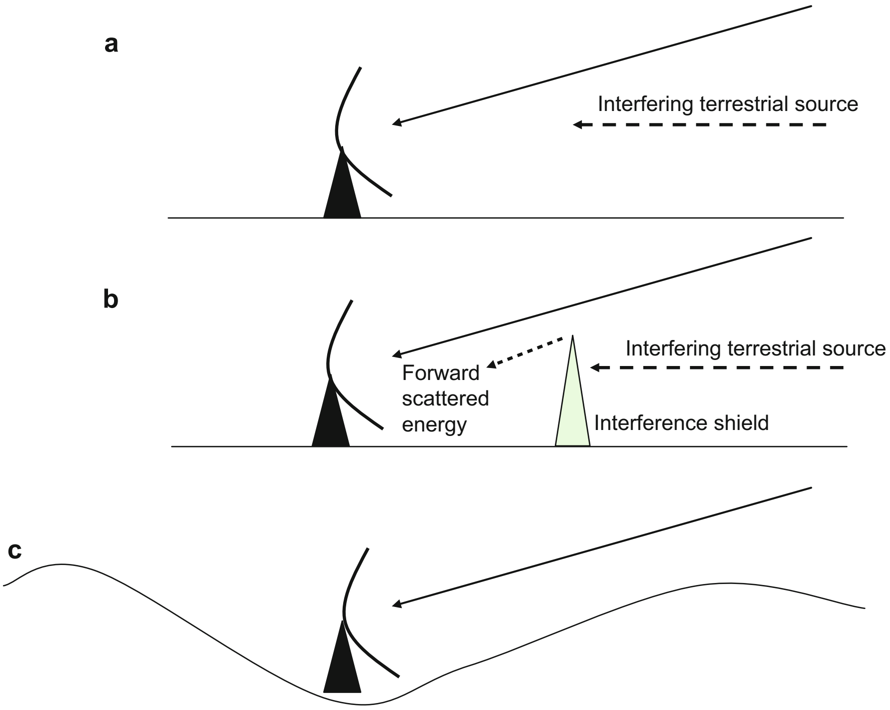


Fig. 20 Illustration of the use of site shielding. (**a**) A terrestrial interference source is shown entering the antenna main beam. In some situations, erecting a metal barrier in the form of an interference shield will provide adequate protection (**b**), but if the shielding fence is incorrectly designed, forward scattered energy can still disrupt communications in the earth station antenna. Possibly the best solution to use if there is sufficient space available is to dig a shallow hole and, with the soil removed from the hole, build a berm around the hole, as depicted in (**c**). The antenna would be located inside the shallow hole. Forward scattered interference is significantly reduced if the shield has a rounded top: the bigger the radius of curvature, the better

systems to ensure that interference potential is at a minimum. In cases where potentially interfering signals can exist into, or from, a system operating in another country, *international coordination* is required between the affected countries, and this is generally administered through the offices of the International Telecommunication Union (the ITU). The ITU is based in Geneva, Switzerland.

The process of coordination, whether national or international, requires the calculation of the likely interference levels. In some cases, the distance between the two interfering systems is not sufficient to provide the required level of protection, and operators need to resort to ways in which they can artificially protect their antennas from interference. One of the most popular methods is *site shielding*.

Site shielding can be natural or artificial. One of the best natural shields is rolling terrain, or even better, a mountain range. In the absence of natural shielding, artificial shielding is resorted to. The simplest artificial shield is a metal fence, as depicted in Fig. 20b. Diffraction of energy is highest over an obstacle when the obstacle has a sharply defined edge, so site shields should be rounded, if possible. If sufficient space is available at the proposed earth station site, a shallow hole should be dug with the soil removed from the hole placed around the hole to form a raised rim, called a *burm*. This is depicted in Fig. 20c. Earth station operators often make the mistake of thinking the best location for an earth station is on the top of a hill, but this exposes the earth station to the maximum potential for interference. Locating the earth station in a shallow valley would be better than on a hill top.

Whether in a valley, a hill top, or the center of a city, the earth station antenna will be exposed to the elements, and so consideration must be given to providing adequate protection to the antenna and feed from the weather.

Weather Protection

There are essentially three meteorological elements that the antenna systems will possibly need protection from, depending on the climate it is operating in: water, snow/ice, and the sun.

Water: Liquid water can cause significant attenuation at frequencies of 10 GHz and above. It can also cause oxidation on metallic surfaces. Feed covers are often used over feed horns to prevent water entering the feed. Care must be taken to ensure the feed covers are cleaned regularly.

Snow/ice: Snow and ice buildup on feed covers and antenna reflector surfaces can cause two effects. The first really is not in evidence until the temperature rises above 0 °C, at which point the frozen particulates melt, and a layer of water will cover the feed cover or antenna reflector surface. The second is if there is a significant fall of snow onto the surface of a large parabolic antenna. A heavy layer of snow in one part of a large reflector can cause the reflector to distort out of a parabolic shape, thus lowering the gain of the antenna. If the earth station is sited in a climate where the temperature regularly falls below freezing point for many weeks, consideration should be given to heating the feed cover and the reflector surface.

Sun: Many medium-sized earth stations and VSATs are located in hot regions of the world. To maintain stable operation of the receiving equipment, it is not unusual to have the equipment box heated to a temperature above the maximum expected outside temperature. However, if the equipment box is located where it is possible to receive direct heating from the sun for several hours in the day, the temperature inside the equipment box can go well above the anticipated temperature. In a VSAT located in Hong Kong, the equipment box was heated to 45 °C, but direct heating from the sun caused the equipment box to reach an internal temperature of 70 °C. To solve the problem, a sun shade was erected over the equipment box.

Many operators who have earth stations located in regions of the world where snow and freezing temperatures persist for several months house the complete antenna system inside a radome shelter that protects the entire antenna system from the elements. In such cases, care should be taken to ensure that the radome, and any particulates that adhere to the outside surface, do not degrade the performance of the link. In particular, if dual-polarized operation is contemplated with an antenna inside a radome, the depolarizing effects of the radome should be characterized.

Feed Systems

A simple rule of thumb to decide whether to employ a single-reflector or a dual-reflector configuration is as follows: if the aperture diameter is $\geq 100\ \lambda$, then a dual reflector is preferred; if the aperture diameter is $< 100\ \lambda$, then a single-reflector configuration should be used. In the section on system noise temperature, we saw that one of the major contributions to system noise temperature is the feed run that connects the feed horn to the receiver. In VSAT systems, where the antenna reflector is fairly small, the feed horn is usually offset from the mechanical axis to minimize blockage. For a dual-reflector antenna, offset configurations are employed in some situations, but for really large antenna systems, with the aperture diameter of 18 m or larger, the dual-reflector configuration is almost always on-axis. The aperture blockage from the sub-reflector is not significant enough to warrant an offset-fed design. However, for large antennas, the feed run from behind the main reflector to the receiving system below can be long – perhaps more than 100 ft – and so the waveguide loss can create a significant noise temperature contribution (see Fig. 21).

To reduce the feed loss and hence the noise temperature contribution of the feed, a *beam waveguide* configuration can be used. This is shown in Fig. 21b. A beam waveguide consists of a series of concave reflectors that conduct the received signal through the antenna support structure, down to the receiver. The system of reflectors

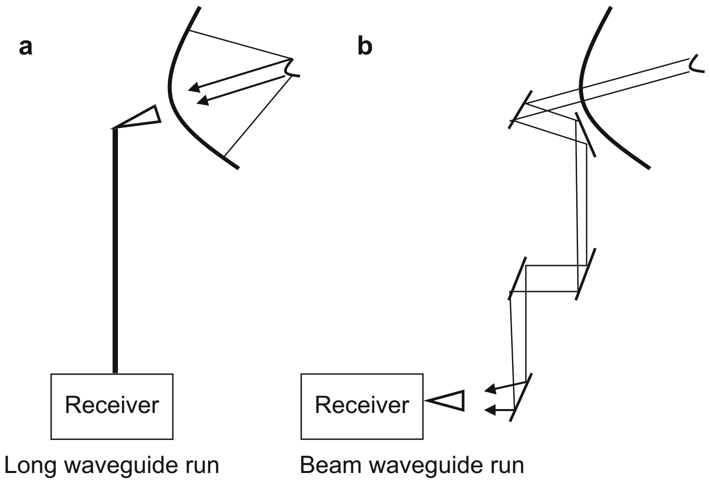
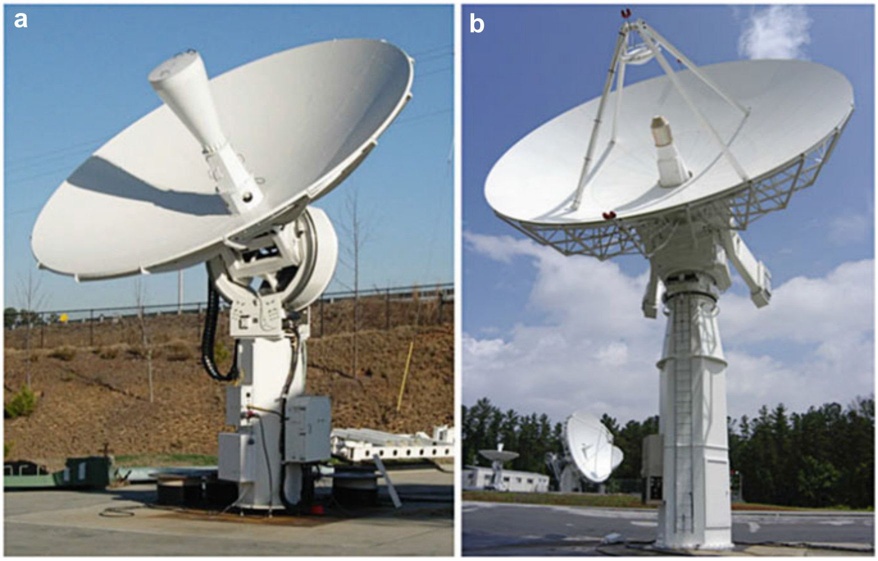


Fig. 21 Illustration of the difference between a standard waveguide run and a beam waveguide. Traditionally, a long waveguide run will connect the feed horn of the Cassegrain antenna to the receiver (**a**). The loss of the waveguide feed can be as high as 0.3 dB, leading to an increase in the noise temperature at the input of the receiver of about 18 K. A beam waveguide reduces the loss considerably, and a beam waveguide configuration is shown in (**b**). (Note: the reflector surfaces in the beam waveguide are not flat, but slightly concave to focus the beam within the narrow confines of the earth station physical structure. The earth station structure is not shown for clarity)

looks like a type of periscope, and so antennas that use beam waveguides are sometimes called *periscope antennas*.

Conclusion

Antenna systems form a key link in the transmission and reception of signals from satellites. The antenna is designed to maximize the efficiency of receiving, or sending, electromagnetic signals. The design changes with the operating frequency, and the need to provide high-gain and good sidelobe characteristics. While omnidirectional antennas permit mobile systems to operate without the user needing to know where the satellite is located, they considerably reduce the communications throughput due to their low gain. A parabolic main reflector is seen to provide the highest gain, with a number of designs available for specific systems (two examples are shown in Fig. 22). While offset-fed designs are preferred for small aperture sizes, as the size of the main reflector increases, so symmetrical, on-axis designs become optimum in terms of gain, design, and ease of construction. The location of the antenna is also important when considering interference issues and whether to adopt a natural or artificial site shield. The choice of modulation is a critical system design parameter as it will impact the occupied bandwidth, the fade margin required, and



5.4 m antenna
© ViaSat Inc., Satellite Ground System Division
Reproduced with permission

11 m antenna
© ViaSat Inc., Satellite Ground System Division
Reproduced with permission

Both of the above earth station antenna systems were developed for operation to low earth orbiting satellites, although they could equally well be used for MEO and GEO satellites. In the example on the left, a Cassegrain configuration is used, with the sub-reflector attached to a cone that extends from the primary feed. In the example on the right, the much larger antenna employs four struts to hold the sub-reflector in a Cassegrain configuration. Both antennas are "Az-over-El", that is the elevation axis is supported above the azimuth axis. The smaller antenna on the left does not require counter-balancing loads to offset the weight of the main reflector, while the 11 m antenna on the right has two counter-balances to reduce the wind, and other, loading forces on the steering mechanism.

Fig. 22 Examples of X-band (approximately 12 GHz) earth station antenna systems developed by ViaSat

the ability to resist interfering signals. Frequency reuse systems that employ dual polarization confer a significant increase in operational bandwidth but at some cost. Weather characteristics in the operating region should also be part of the overall design of the antenna and its feed system. The emphasis in this chapter has been to provide a basis understanding of satellite earth station technology, particularly for GEO systems. Satellite Communications Antenna Concepts and Engineering that follows discusses some of the newest concepts to support earth antenna systems for LEO constellations providing broadband services.

Cross-References

- ▶ Satellite Communications Antenna Concepts and Engineering
- ▶ Satellite Antenna Systems Design and Implementation Around the World

References

J.E. Allnutt, *Satellite-to-Ground Radiowave Propagation* (Peregrinus, London, 1989). ISBN 0 86341 157 6

J.E. Allnutt, *Satellite-to-Ground Radiowave Propagation*, 2nd edn. (IET, London, 2011). The 10 digit ISBN is 1849191506 and the 13 digit ISBN is 9781849191500

T. Pratt, C. Bostian, *Satellite Communications* (Wiley, New York, 1986)

T. Pratt, C. Bostian, J. Allnutt, *Satellite Communications*, 2nd edn. (Wiley, New York, 2002)

Recommendation ITU-R P.618-9, Propagation data and prediction methods required for the design of earth-space telecommunications systems (2007)

Recommendation ITU-R P.676-4, Attenuation by atmospheric gases (1999)

Recommendation ITU-R P.838-1, Specific attenuation model for rain for use in prediction methods (1999)

W.L. Stutzman, G.A. Thiele, *Antenna Theory and Design*, 2nd edn. (Wiley, New York, 1998). ISBN 0-471-02590-9

Technical Challenges of Integration of Space and Terrestrial Systems

John L. Walker and Chris Hoeber

Contents

J.L. Walker (✉)
Lockheed Martin Space Systems Company, Littleton, CO, USA
e-mail: johnnie.l.walker@lmco.com

C. Hoeber
CFH Engineering, Palo Alto, CA, USA
e-mail: chris@cfhengineering.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_22

Abstract

This chapter discusses the challenges of integrating space and terrestrial systems as well as some of the unique solutions and approaches to solving those challenges. While the first satellite systems were stand-alone and akin to a private network in today's terminology, virtually all current satellite systems are interconnected through some component of the terrestrial infrastructure, e.g., the Internet, PSTN, or private fiber. This chapter presents examples of the current challenges in integrating space and terrestrial systems by considering two satellite system classes which have unique requirements for interconnection and interoperability: the Mobile Satellite Systems (MSS) and the Broadband Satellite Systems for Internet Access.

Keywords

Ancillary terrestrial component (ATC) • DBSD • Doppler • Ground-based beam forming (GBBF) • Inclined orbit • Internet protocol over satellite (IpoS) • Internet protocol (IP) • Internet router in space (IRIS) • ISO 7-layer • Lightsquared • Mobile satellite systems (MSS) • Multiple input/multiple output (MIMO) • O3B • Performance-enhancing proxies (PEP) • Protocols • Satellite • Satellite phone • TCP/IP • Terrestar • Web acceleration

Introduction

This chapter discusses the challenges of integrating space and terrestrial systems as well as some of the unique solutions and approaches to solving those challenges. The History of Satellite Communications was presented in an earlier chapter, and began in the 1950s. The first satellites celebrated the ability for man to place a satellite in orbit and were used to gather information to evolve the art of space communications. The first active communication satellite was not launched until 1960, and the next few years' satellites were launched to demonstrate and provide a relay capability for teletype, voice, and even television. Since then, satellites and satellite systems have evolved to provide a variety of services.

In today's world, more and more things are interconnected. A person can surf the Web on their cell phone, track the location of their pet on their computer, and monitor and adjust the temperature of their home while they are away. They can remotely program their satellite television receiver to record a show their friend at the office recommends and watch it when they get home or even watch the show remotely on their laptop or cell phone.

While the first satellite systems were essentially stand-alone and akin to a private network in today's terminology, virtually all current satellite systems are interconnected through some component of the terrestrial infrastructure, e.g., the Internet, Public Switched Telephone Network (PSTN), or private fiber. The complexities of satellites have evolved since their beginning and so has the need and depth for satellite systems to connect and operate with terrestrial systems and infrastructure.

This chapter presents examples of the current challenges in increasing this interconnection and interoperability by considering two satellite system classes which have unique and intricate requirements for terrestrial integration: the Mobile Satellite Systems (MSS) and the Broadband Satellite Systems for Internet Access.

In the first section, the challenge of market economics for mobile satellite operators is discussed along with the advent of the Ancillary Terrestrial Component (ATC). Key challenges for MSS operators are summarized, followed with a discussion into current and future solutions to solving those challenges. A synopsis of the recent MSS deployments is provided along with the observation of the importance of ground-based beam forming (GBBF) as an enabler for this industry. The GBBF approach is described, including an overview of the architecture and the challenges in developing the concepts and implementing and deploying the system. All of the recent geosynchronous MSS satellites are in an inclined orbit, and the reasons are discussed.

The MSS integration challenges are not limited to connecting the Gateway with the PSTN and Internet but include the additional challenge of integrating the satellite user terminal (satellite phone) with the ATC. Finally, Multiple Input/Multiple Output (MIMO) systems are described in the context of addressing challenges in satellite link performance.

The chapter moves next to address the challenges in broadband satellite systems for Internet access. The challenges are categorized into three areas, followed by discussions of current and future approaches to their resolution. A brief discussion on the market challenges is provided as this leads to the innovations required by the technical solutions.

Protocols and reference models are discussed with the goals: to indicate the development frameworks in which protocols have been and are being developed; to indicate many of the protocols currently in use in both terrestrial and satellite-terrestrial integrated systems; and to discuss some of the specific challenges in current protocol development.

One of the ongoing challenges facing the integration of satellites and terrestrial networks concerns the protocol implementation of the Transport Layer Protocol, TCP. Since a large amount of Internet traffic is directly related to TCP, the issues and solutions envisioned are discussed including some examples of current implementations.

Finally, a section is provided to present the challenges, potential advantages, and disadvantages of placing an Internet router in space.

Mobile Satellite Systems and Ancillary Terrestrial Component

Mobile Satellite Systems (MSS) present unique challenges to both the space and terrestrial cellular industries. The MSS category refers to satellite systems where the user communication terminals are mobile and portable. The user frequency band is smaller than the broadband category to be discussed later, although the term "broadband" is sometimes used in the context of particular MSS data services such as Web browsing. The user frequency bands typically span 20–40 MHz one-way. For example, the FCC licenses MSS services in the 2 GHz (1990–2025 MHz and 2165–2200 MHz), the L-band (1525–1544 MHz/1545–1559 MHz and 1626.5–1645.5 MHz/1646.5–1660.5 MHz), and the "Big LEO" (1610–1626.5 MHz and 2483.5–2500 MHz) band.

Market economics are extremely challenging for mobile satellite operators, as the build-out of terrestrial cell phone services has been rapid. In addition, the terrestrial cell phone technologies and services have evolved rapidly through the progression of 2G, 3G, and 4G digital mobile communications.

In the 1990s it was recognized that MSS operators were challenged with the market economics to provide mobile services with commercially successful systems. The build-out of terrestrial cell phone coverage was expanding, as well as the march to improve cell phone technologies, increase the services offered, and enhance the quality of service provided. Nevertheless, there were areas unserved by terrestrial towers that a satellite-based solution could address.

The FCC recognized the value of MSS to provide advanced communications to areas not readily or economically served by terrestrial systems and released a notice of proposed rulemaking in 2001 (FCC 2001). The consideration was to allow the MSS operator to use terrestrial base stations to augment the satellite coverage. In this rulemaking, the satellite provides the primary communications. However, when the link performance is poor due to foliage and terrain, or when the users are in buildings, the terrestrial fill-in base stations provide the communications link. In 2003, the first rulemaking was released authorizing MSS operators to add these base stations, referred to as the Ancillary Terrestrial Component (ATC), to their networks.

The approach has continued to evolve with the latest rulemaking in 2005 (FCC 2005). The FCC set a gating criterion that all MSS/ATC equipment must be able to communicate via both the satellite and the ATC and that the services offered must be available through both the MSS and the ATC. While this provided economic opportunity for MSS operators, key technical challenges needed to be solved, some of which are still ongoing as of the time of this writing.

Interestingly, the challenges presented in developing and integrating Mobile Satellite Systems are not limited to integrating the satellite gateway with the terrestrial Internet and telephony networks. These challenges extend to providing adequate quality of service to handheld users, coordination and noninterference between satellite systems, and efficient sharing of the user spectrum between those users linked through the ATC and those linked through the satellite. In addition, challenges exist in the development of user terminals which must consider dual mode (satellite and terrestrial) and the rapid advance of terrestrial cellular phone applications and services.

To summarize, the key challenges for MSS are:

1. User terminal handset compatibility with terrestrial and satellite communications
2. Uplink and downlink interference management within and between MSS/ATC operators
3. Commercially successful viability in offering services via MSS/ATC with the continuing rapid advance of terrestrial networks and service evolution: 2G, 3G, 4G
4. Flexibility of system architecture to adapt to market evolution

For challenge 1, the FCC rulemaking is limited to addressing the need for the user terminal to have compatibility and common services within the MSS/ATC operator network. However, when challenge 3 is folded in, consideration must be given to multimode user terminals which can communicate not only with the MSS satellite and corresponding ATC but also have roaming and compatibility with a standard terrestrial network provider.

Multimode phones eliminate the need for the consumer to have both an MSS handset for rural and remote areas and a terrestrial cell phone for more economic connection as well as the likely advanced services which have evolved more rapidly in the terrestrial networks. TerreStar networks pursued this approach and developed the GENUS phone which is a multimode smart phone as depicted in Fig. 1. In

Fig. 1 TerreStar GENUS phone

addition, they developed business relationships with AT&T to offer both satellite and cellular communications on one device with a single phone number on a single bill. Smartphone features such as text, email, contacts, and calendar were included in both satellite and cellular mode (TerreStar).

In considering challenge 3 and challenge 4, along with meeting aggressive milestones to maintain FCC licensing, the need was prevalent for a very flexible satellite system architecture. MSS operators needed spot beams and frequency reuse plans that could be modified as the ATC roll-out was implemented and as the traffic and service demands evolved. Combining this requirement for flexibility with the need for shorter satellite schedules and the necessity of keeping the total satellite costs as low as practical led to the development of the Ground-Based Beam Forming (GBBF) solution. Although all the MSS-ATC systems employ similar GBBF technology, these US-based networks have all suffered financial reversals for a variety of different reasons as discussed in chapter "▶ Space Telecommunications Services and Applications."

At the time of this writing, all of the recent MSS programs (DBSD, TerreStar, LightSquared) employ Ground-Based Beam Forming (GBBF). With GBBF, beams can be added, removed, or reconfigured to enable a satellite to operate from different orbital locations or to adapt to changes in traffic patterns and evolve to service new applications. Further, with beam forming performed on the ground, the cost and time to deliver a highly flexible satellite is significantly reduced.

The worlds' first satellite to utilize a two-way Ground-Based Beam Forming (GBBF) system, DBSD G1 (previously named ICO G1), was launched from Cape Canaveral, Florida on April 14, 2008. Designed and built by Space Systems/Loral (SS/L), G1 has a 12 m unfurlable reflector and provides 250 fully configurable transmit and receive beams covering the continental US (CONUS) including Alaska, Hawaii, and Puerto Rico. SS/L was first in developing and deploying a two-way GBBF and the on-orbit performance met or exceeded all the original requirements. SS/L selected and subcontracted to Hughes Network Systems LLC (Hughes) to develop and implement the high-speed signal conditioning and processing ground equipment that enables GBBF.

The Terrestar-1 satellite, also built by SS/L, was launched a little over 1 year later in July, 2009 on an Ariane-5 in French Guiana. This was the largest, most advanced commercial communications satellite launched at the time. It hosts an 18 m unfurlable reflector and extends the coverage over the DBSD G1 to include Canada.

LightSquared, formerly MSV Skyterra, launched SkyTerra 1 from the Baikonur Cosmodrome, Kazakhstan, on November 14, 2010. This satellite incorporates a 22 m unfurlable antenna and an onboard digital channelizer to ease with interoperability of other L-Band systems within the fragmented L-Band spectrum. The GBBF ground equipment is provided by Viasat. The coverage area is extended even further to include Mexico and Central America.

Central to these MSS/ATC systems is the need to share the spectrum between the satellite and the ATC base stations. Commercial viability requires a large numbers of users. The frequency reuse schemes, along with the flexibility of the satellite network

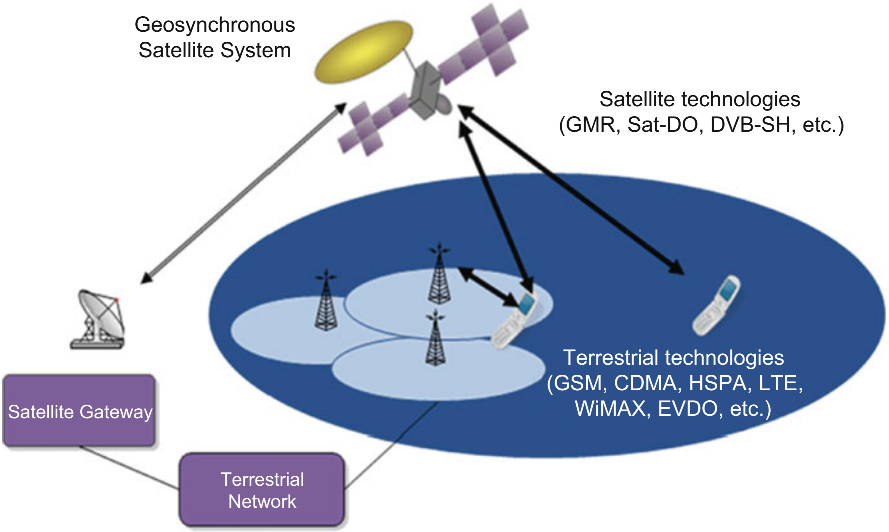


Fig. 2 Hybrid Mobile Satellite System (MSS) with ancillary terrestrial component (ATC)

to place beams, adjust beam power and to perform interference suppression is key to maximizing the overall ATC and satellite capacity.

In an optimized architecture, the ATC cell size is much smaller than the satellite beam cells. Figure 2 illustrates the hybrid architecture of a MSS with ATC. The GBBF provides flexibility of assigning user frequencies to beams, adjusting the size and shape of the formed beam and varying the amount of power applied to each beam. The only limitation on the flexibility is the geometric constraints of the antenna optics on the satellite. Large unfurlable reflectors are used to achieve high G/T values and provide the link performance required to communicate reliably with cell phone style satellite user terminals.

Interestingly, the three MSS systems mentioned deployed increasingly larger reflector size and coverage area. Going from 12 to 18 to 22 m and increasing the coverage area drives additional payload equipment to handle the increased feed element count in the feed array that illuminates the reflector. This trend is likely to continue with further miniaturization of payload hardware and the advantages in link performance to mobile users.

Perhaps an even more important factor to drive an increased reflector size is related to challenge 2: uplink and downlink interference management within and between MSS/ATC operators. The size of the smallest spot beam that can be formed is not limited by the GBBF, but rather is limited by the aperture size of the satellite antenna. For example, a 12 m reflector would have a.47 deg 1-dB beamwidth or 1.2° null-to-null and span 775 km on the Earth. Whereas, a 22 m reflector can form a spot beam at S-Band with a 0.26° 1-dB beamwidth or 0.66° beamwidth null-to-null. The null-to-null angle would span approximately 430 km in diameter on the Earth from the geosynchronous orbit.

The larger reflector size and correspondingly smaller spot beam results in increased capacity for the satellite system. The smaller spot beam directivity rolls off much quicker over the geography than a larger beam, so the frequency can be reused more times over the coverage area. In addition, the smaller beam provides increased ability for interference cancelation to enhance the capacity of the satellite and ATC total network and increased directivity for the RF links.

In summary, the GBBF offers several advantages to conventional satellite systems. First, it simplifies the satellite design when compared to onboard processing solutions. Second, the GBBF architecture is very robust to changing markets and business plans. The GBBF equipment on the ground uses flexible digital signal processing hardware and software which allows beam patterns to be easily changed, allowing the system to provide EIRP and G/T to match traffic demands. Also, the ground algorithms and equipment can be readily upgraded to enhance system capability as needed and is readily accessible for repair and replacement.

As an example to understand the architecture and terminology of these GEO MSS GBBF systems, consider the DBSD G1 Satellite and GBBF architecture shown in Fig. 3.

The DBSD Space Segment and GBBF consist of the G1 Satellite, the Gateway, and four Pointing Beacon Stations (PBS). The Gateway includes the Radio Frequency Subsystem (RFS) and the Ground-Based Beam former (GBBF) subsystem. The base station complement to the user terminals is also included at the Gateway. The G1 Satellite services S-Band user terminals and employs a Ka-Band feeder link. The GBBF is integral to the workings of the G1 Satellite and is therefore included with the satellite in what is defined as the Space Segment. Consequently, specifications were developed to include the GBBF and RFS in what is normally specified for the satellite alone, e.g., G/T, EIRP, pointing error. The GBBF assemblies are

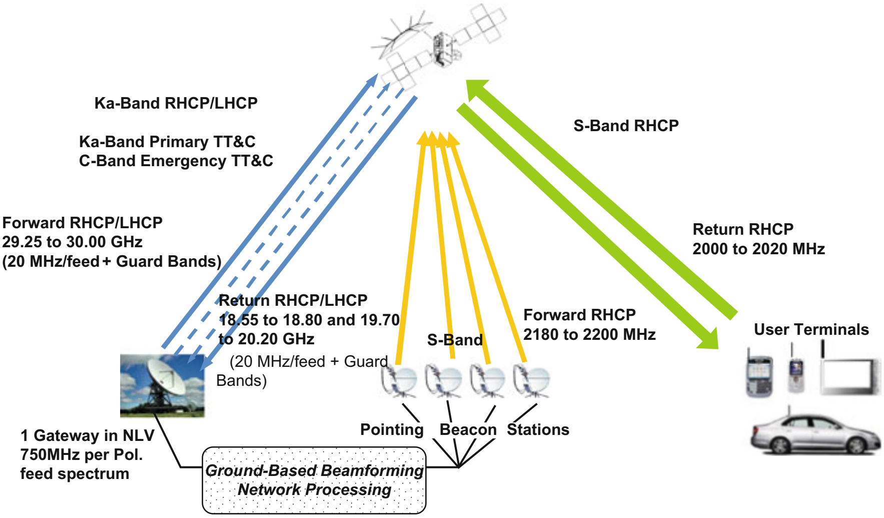


Fig. 3 DBSD space segment architecture with GBBF

physically located within the Gateway in north Las Vegas and the PBS Stations are located at four locations within the Continental United States (CONUS). The communication signal flow places the GBBF assemblies in between the user terminal head-end, named the Mobile Satellite Systems (MSS) Communications Processing equipment (MCP), and the Gateway RF Subsystem.

The return link is defined as the signal path transmitted from the user terminal on the ground up to the Satellite and down to the Gateway, through the GBBF and to the MCP. Therefore, the return link consists of both an uplink from the user terminal to the Satellite and a downlink from the Satellite to the Gateway. The forward link is defined as the signal path transmitted from the MCP source through the Gateway up to the Satellite and down from the Satellite to the user terminals on the ground. Likewise, the forward link consists of both an uplink from the Gateway to the Satellite and a downlink from the Satellite to the user terminal receiver.

The Space Segment provides service to the user terminals via the forward (transmit) and return (receive) beams in the coverage area. The number of beams and their shapes and sizes are controlled by the GBBF.

The communication paths between the Satellite and Gateway are referred to as the feeder links. The forward feeder link corresponds to the Ka-Band transmission from the Gateway to the Satellite (uplink). Similarly, the return feeder link corresponds to the Ka-Band transmission from the Satellite to the Gateway. The Ka-Band feeder link requires both polarizations simultaneously in order to accommodate the forward and return signals to and from each feed element of the satellite S-Band antenna used to form the beams to communicate with the S-Band user terminals.

The Gateway consists of a Radio Frequency Subsystem (RFS) which includes a 13.2 m Ka-Band dish, High Power Amplifiers (HPAs), and up and down conversion equipment which connect to the GBBF at an intermediate frequency of 140 MHz.

The GBBF subsystem consists of two GBBF assemblies with four racks of equipment and four Pointing Beacon Station (PBS) transmitters. The GBBF assemblies are located in the Gateway and the PBS transmitters are positioned at four locations over CONUS to aid with spot beam pointing control.

The Satellite S-Band antenna consists of a 12 m reflector and a 46 element feed array for CONUS coverage, with two additional feed clusters which employ a traditional onboard beam forming network and provide coverage for Hawaii and Puerto Rico.

As will be described, the Satellite in the GBBF Space Segment architecture is essentially a bent-pipe transponder converting each feed element in the Satellite feed array between S-Band and Ka-Band. The GBBF ground equipment processes the received signals from each feed element to form the beam outputs in the return direction. Likewise, in the forward direction, the user traffic signal is applied to the desired beam input to the GBBF, where the GBBF decomposes the signal into the individual signals to be applied to each of the Satellite S-Band feed array elements. These signals are then sent up to the Satellite on the Ka-Band feeder link where they are transponded, amplified, and sent to the S-Band feed array. The individual signals from each feed element radiate and combine in the far field to form the desired transmitted user traffic signal level over the coverage area (beam).

In summary, the DBSD G1 Satellite and GBBF architecture and implementation offers the following features: (Walker et al. 2010)

- GBBF can form up to 250 beams independently in each direction.
- GBBF can form beams of various sizes and shapes.
- GBBF can use between 0 and 20 MHz in each beam.
- GBBF is flexible to assign spectrum to any beam.
- GBBF is flexible to assign power to beams.
- GBBF is flexible to use any modulation scheme.
- GBBF is flexible to accommodate Ancillary Terrestrial Component (ATC).

GBBF Development and Implementation

To understand GBBF operations, we will begin with a discussion of beam forming independent of where it is accomplished. Beam forming is commonly used in radar and communication systems by combining the signals from a multiple element array (Johnson and Jasik 1984). Figure 4 illustrates beam forming operations in the return direction.

In the return direction a signal transmitted from the ground is received at the feed element array depicted on the left of the figure. Each feed element of the array receives the signal at a different amplitude and phase due to the spatial geometry between each feed element of the satellite antenna and the point on the ground where

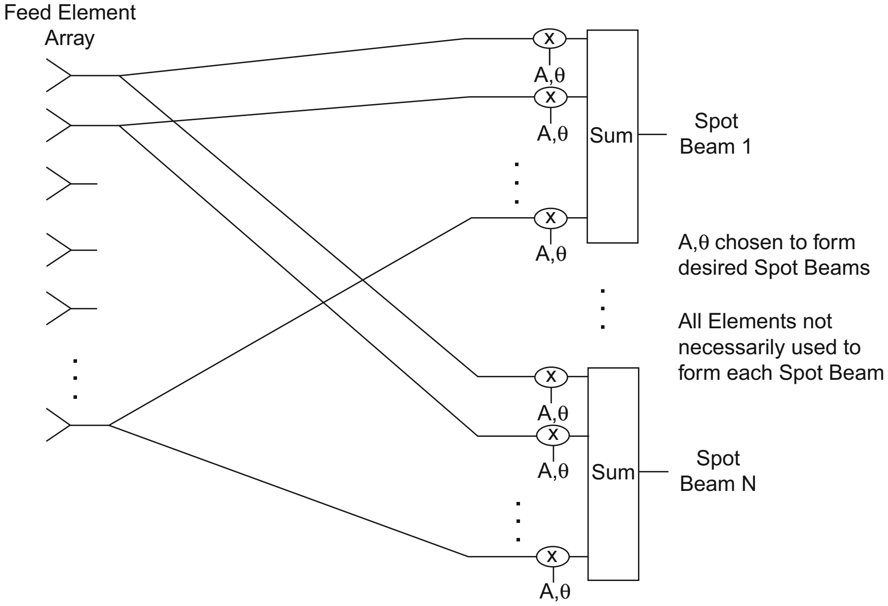


Fig. 4 Return basic beam forming operations

the transmitter is located. The beam forming operation is shown on the right of the figure and consists of taking the signal from each feed element array and applying a beam forming coefficient weight to shift the amplitude and phase of the received signal by A_c and Θ_c. A spot beam is formed by combining the beam coefficient weighted feed element signals as shown by the Sum block. By choosing different coefficient weightings (A_c, Θ_c), different beams may be formed.

Each A_c and Θ_c value is required to be a specific number to form any particular beam contour. By changing the values of A_c, and Θ_c the formed beam is changed. For example, by changing the values of the weighting, a beam may be steered in pointing, narrowed to a spot, broadened to a regional beam, or designed to have additional sidelobe suppression or even nulls at specific geographical points. Further, virtually any number of beams may be formed by replicating the weighting and sum operations with various combinations of the feed element signals.

Beam forming by weighted combination of feed elements is graphically illustrated in Figs. 5 and 6. Figure 5 shows the raw feed element antenna pattern on the ground. By amplitude and phase weighting these feed elements with the proper coefficient weights, the raw feed patterns are converted into spot beams as shown in Fig. 6.

The beam forming operation, as described so far, is well understood and has been used extensively in communication systems of all types. In Ground-Based Beam Forming, the beam forming operations described are moved to the ground as illustrated in Fig. 7.

The signals from each feed element of the array are transponded to the Gateway through the feeder link and sent to the GBBF assembly. The GBBF assembly has an input corresponding to each feed element in the array. Within the GBBF, the beam

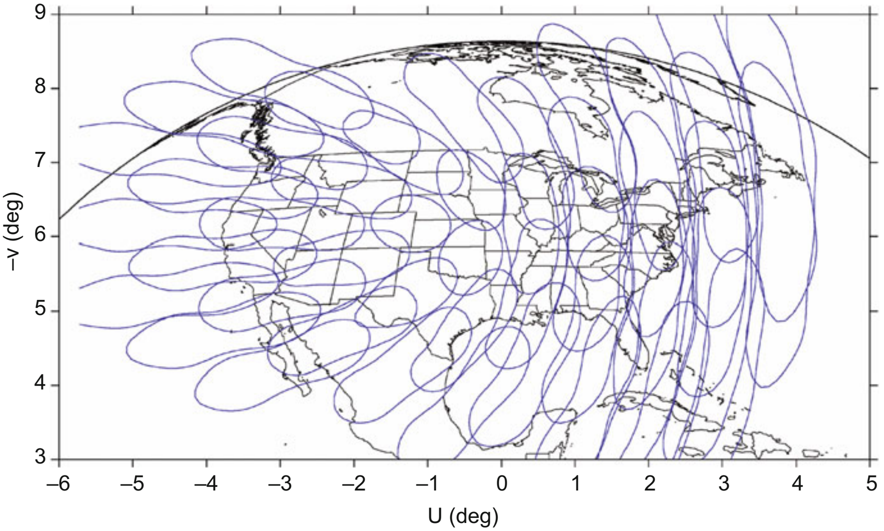


Fig. 5 Forty-six return feed element 3 dB contours before beam forming

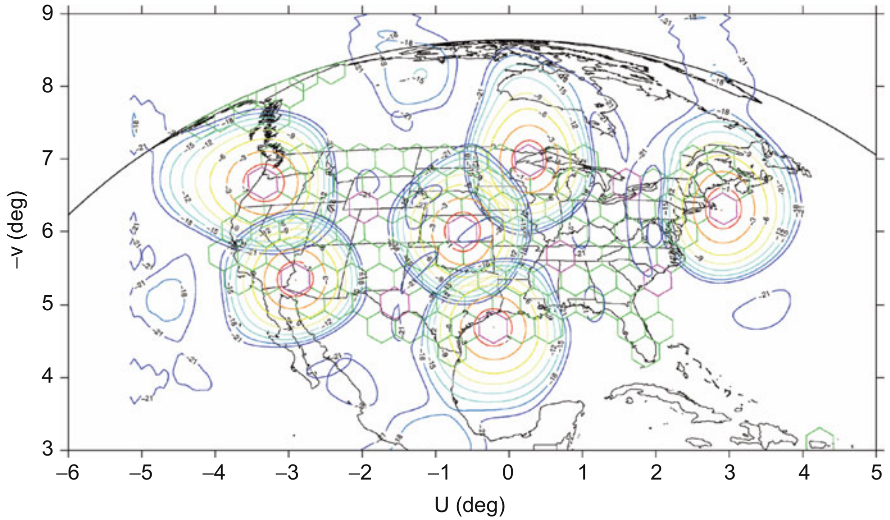


Fig. 6 Six return spot beam contours after beam forming

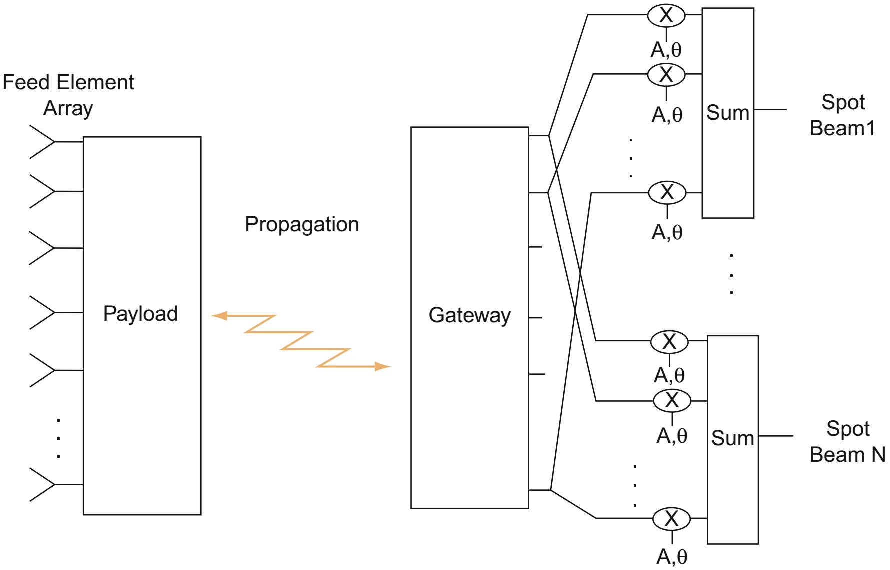


Fig. 7 Ground-Based Beam Forming

forming operations occur as shown in the diagram and the GBBF provides output ports for each of the formed beams.

By moving the beam forming operations to the ground, any amplitude and phase mismatch between the transponded pathways must be determined and compensated for to ensure the desired beams are formed correctly. These pathways include the Satellite payload, the Ka-band feeder link propagation, and the Gateway RF subsystem. Any amplitude and phase fluctuations in any of these three components will alter the formed beam pattern if left uncompensated.

The Ka feeder link contains each feed element frequency division multiplexed along with special calibration channels. Thus, the satellite payload may be thought of as a bent-pipe transponder for each feed element as depicted in Fig. 8. This approach not only simplifies the payload design but allows the payload manufacturing and test operations and procedures to be performed in the usual manner for transponder satellites.

With such flexibility as provided by the architecture to form any number, size, or shape of beams, it is necessary to define a reference set of beams at the beginning of the program to drive the design and to quantify the performance of the Satellite and GBBF. DBSD and SS/L chose the reference case depicted in Fig. 9 and developed the primary specifications for performance for this scenario.

This reference beam set consists of 135 beams total, with 124 beams over CONUS, 9 beams over Alaska, plus 1 beam each over Hawaii and Puerto Rico. There are 133 hexagonal cells over CONUS and Alaska which correspond to geographic areas. The Space Segment forms spot beams, using the GBBF and the Satellite S-Band antenna, which are approximately centered over these geographic cells. The communications performance of the Space Segment is defined in terms of the beam performance achieved in each of these geographic cells.

It is important to understand that the cells represent the geographic definition and the actual formed beam contours extend beyond the cell as shown by the beam contour lines in Fig. 6. The reference case cell edges correspond to approximately 1 dB down from the peaks of the formed beam contours. The system provides full flexibility to redefine the cell layout and form different sized beams to provide the best quality of service to the users. Nevertheless, a reference case must be defined

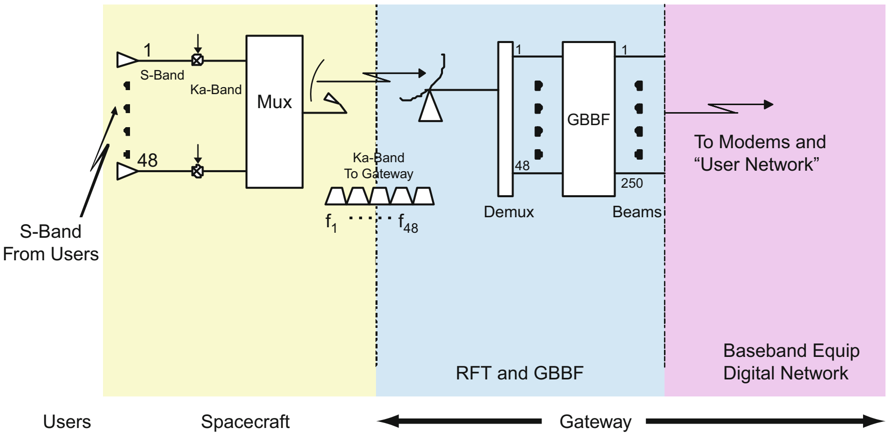


Fig. 8 Return payload is a bent-pipe transponder for each feed element

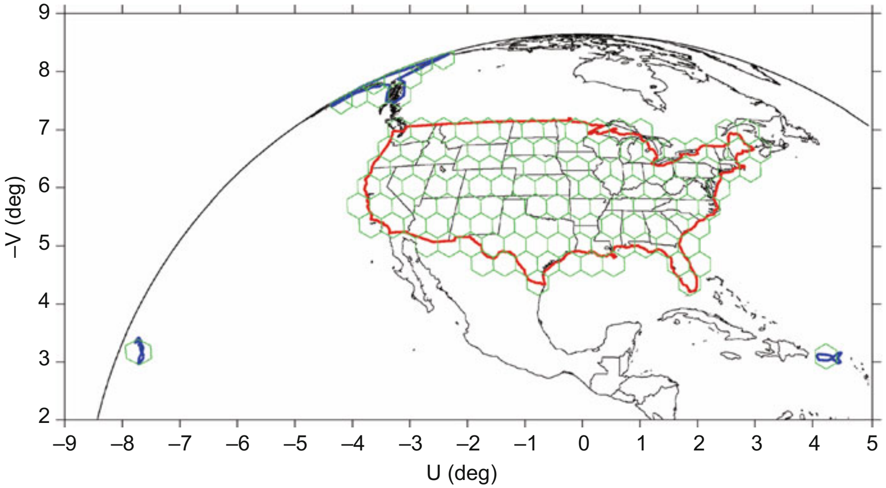


Fig. 9 Example of a reference cell layout over the coverage area

such that the standard metrics of communication performance may be specified, e.g., G/T and EIRP.

The reference set is used for design budgets and verification tests to quantify the performance of the system and to provide the allocations to the subsystems and elements. However, the architecture easily allows different beams to be formed by simply changing the beam coefficient weights. The only limits being the amount of hardware processing deployed and the satellite antenna optics. Additional hardware processing may be included in the initial roll-out to accommodate expected growth, or added at a later date.

GBBF Calibration Scheme

To form a specific beam, the amplitude and phase weighting must be set to the appropriate numbers for each element. Further, these weightings must effectively be applied at the feed element aperture. For example, take two elements and assume the desired beam is formed with an amplitude of 1 and a phase of 50° for element 1 and an amplitude of ½ and a phase of 100° for element 2. In typical onboard beam forming, the beam forming operation is very close to the feed aperture, so it is much simpler to set these values correctly. However, with GBBF, the signals traverse down independent payload conversion paths, through the propagation media at different frequencies which may have differing amplitude and phase channels, and then through independent conversion paths in the RFS. With this example, let us say that the element 1 path experiences 3 dB additional attenuation and 50° of additional phase shift as it traverses through the payload, RFS, and propagation path. Without knowledge and compensation of this difference, the beam forming weights at the aperture will not be the desired values but instead element 1 would be equivalent to ½° and 100°, the same as element 2. Depending on the error experienced, the desired beam may be mispointed, misshaped, or even dispersed so grossly as to not be recognizable as a spot beam. However, if the value of the amplitude and phase difference between the element paths between the feed element aperture and the GBBF beam forming operation is known, it may simply be compensated for by adjusting the feed coefficient weights or compensating for the shift before applying the feed weights. Consequently, in order for the GBBF to function as required, a calibration scheme must be implemented to determine and compensate for the amplitude and phase variations between the feed element paths.

This distributed calibration architecture provides many benefits. First, the satellite design consists of traditional bent-pipe transponders. Consequently, the payload may be designed, built, and verified to traditional specifications without any specific knowledge of GBBF. This approach minimizes the cost, schedule, and complexity of the satellite payload. Second, much flexibility is built into the architecture to allow changes to the calibration signaling and algorithms as needed to solve any unforeseen challenges and to evolve the system as markets and traffic profiles change. Third, the testing of the satellite and GBBF after launch is greatly simplified as the satellite and GBBF operate together in a transparent fashion allowing them to be tested together using traditional satellite In Orbit Test (IOT) methods.

Next, let us discuss the details of the calibration architecture. The calibration scheme contains five processes: feeder link Doppler correction, return feed element path gain and phase imbalance correction, forward feed element path gain and phase imbalance correction, pointing error estimation and correction, and uplink power control.

Feeder Link Doppler Correction

The diagram in Fig. 10 is used to describe the Doppler correction scheme. The upper right of the diagram illustrates the satellite-based components of the distributed calibration architecture, whereas, the lower left of the diagram illustrates the ground-based components. The upper left and lower right images show the Ka-band feeder link frequency plan: the uplink and downlink.

The Doppler effect is well known and refers to a frequency shift that occurs between a transmitter and receiver when they are in relative motion. The relationship between the received and transmitted frequency is given by:

$$f_r = \left(1 - \frac{v_r}{c}\right) f_{tx}$$

The satellite is launched in an inclined orbit of 6° which results in the Doppler shift varying between plus and minus 3.2 KHz at the Ka-Band feeder link frequencies over the 24 h period. As discussed earlier, in order to form the desired beam, the amplitude and phase weights must be set at the S-band feed aperture to the correct

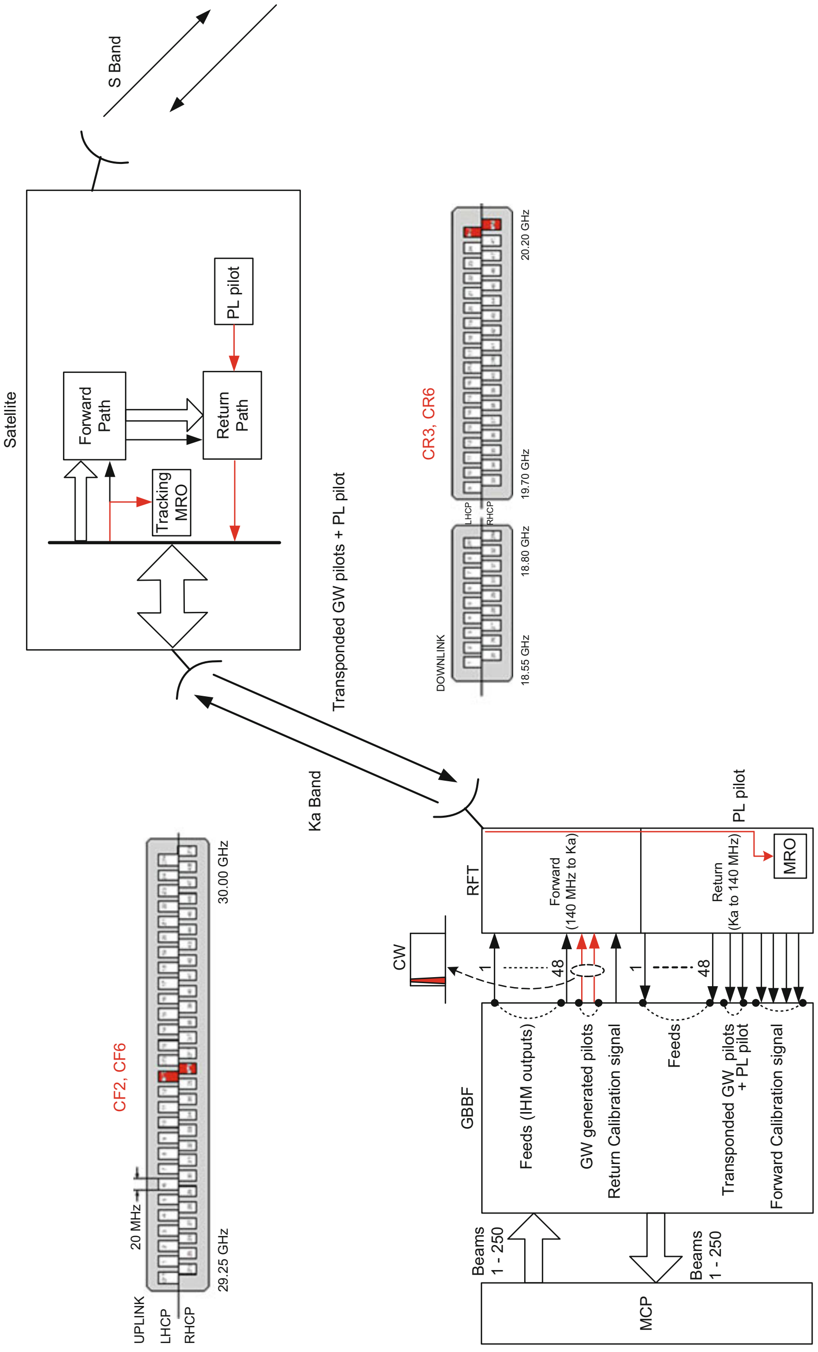


Fig. 10 Doppler correction scheme

values. With a differential Doppler shift of even 1 Hz between the signals in the feed element channels, this would mean that the phase is changing 360°/s. In order to form the beam, the differential phase between elements must be kept small, e.g., 5°. Furthermore, in the forward direction, the signals leave the GBBF with their set amplitude and phase coefficients in advance of their arrival at the feed aperture by the propagation delay from the GBBF output to the feed aperture of approximately 250 ms. Thus the Doppler correction must be precise.

The method of Doppler correction is as follows. First, the GBBF includes a very precise low phase noise Master Reference Oscillator (MRO) which is the master frequency reference for the entire system. A continuous wave (CW) signal derived from this reference is sent up to the satellite in the special calibration channels highlighted in red on the uplink frequency plan of Fig. 10.

The satellite receives this CW signal at Ka-band and routes it to a Tracking MRO (TMRO). The TMRO follows the Doppler frequency shift of the CW reference signal as the satellite moves through the orbit. The satellite payload uses the TMRO as the master reference for all frequency conversions on the satellite. Thus, as the frequency of the calibration signal moves higher, the local oscillators used to translate the Ka-Band signal to S-Band also move higher. The Doppler effect causes the 20 MHz channels to move higher in frequency and also to be spaced further apart. However, the local oscillators on the satellite move in a likewise fashion such that the signals in each 20 MHz channel arrive at each S-Band feed at the same frequency.

In the return link the process is similar. A payload pilot CW signal is generated on the satellite and sent down to the Gateway RF Terminal subsystem in the special calibration channels highlighted in red in the downlink frequency plan. The RFT also has a TMRO which receives the payload pilot and tracks the feeder downlink Doppler shift. The return frequency conversions in the RFT are locked to the Gateway TMRO so that all the return channels arrive at the GBBF lined up in frequency.

Return Feed Element Path Gain and Phase Imbalance Correction

The return calibration scheme is illustrated in Fig. 11. Notice the two red X's on the diagram. One is at the S-band feed aperture on the satellite in the upper right. The second is at the GBBF feed element input where the RFT translates the transponders to provide the 48 feed element channels.

In order to form the desired beam, the amplitude and phase weight for each beam must be set at the feed aperture: the first red X. The S-band signals from users on the ground arrive at each feed element and travel through the return conversion path where they are upconverted to Ka-band and frequency division multiplexed as highlighted in red on the Ka-band downlink frequency plan. The signals from each feed element are then received by the RFT and translated in frequency from Ka-band to 140 MHz and applied to the GBBF. The signal from the feed elements experiences the amplitude and phase shifts that occur along the path from the

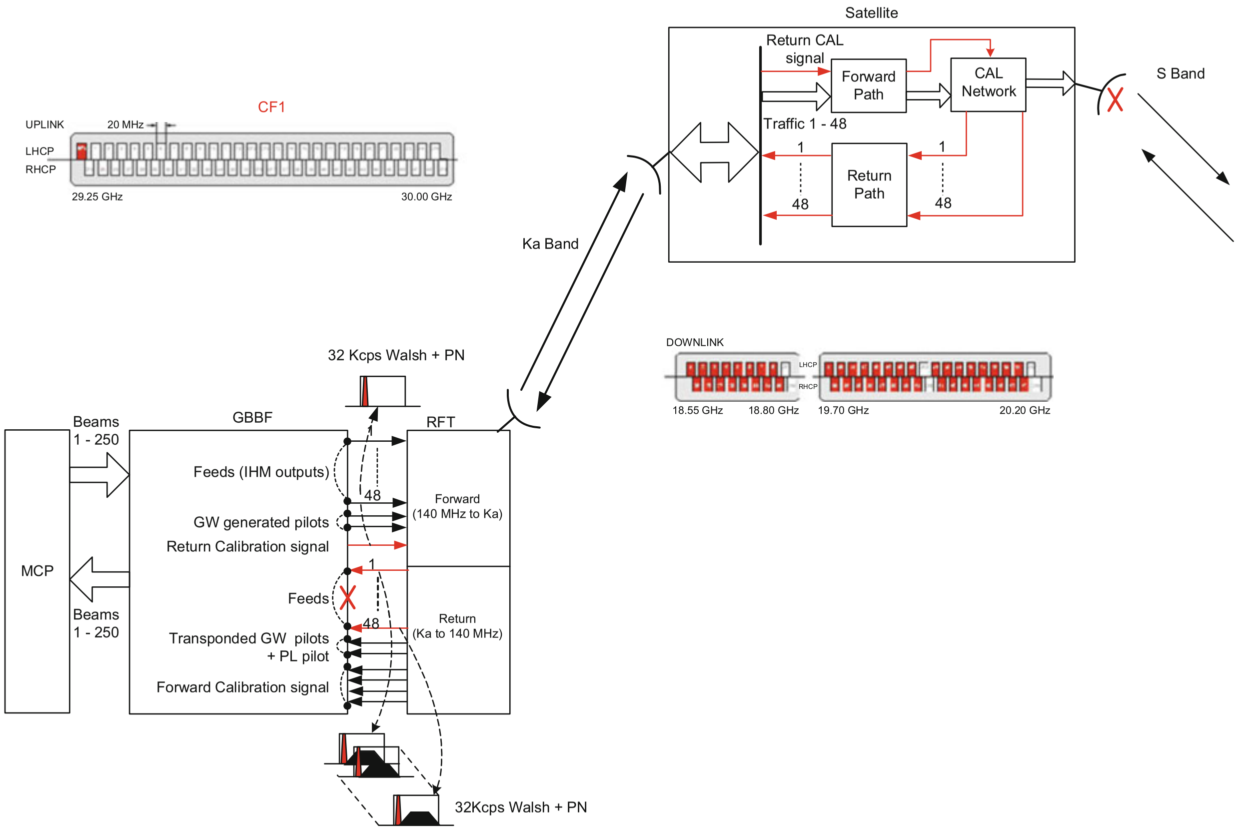


Fig. 11 Return feed element path gain and phase imbalance correction scheme

satellite to the GBBF: the payload, the Ka-band propagation, and the RFT conversions. The calibration scheme's goal is to determine the amplitude and phase for each element path from the aperture, the first red X, to the GBBF beam former, the second red X. Once this value is known for each path, it is easily compensated for in the GBBF.

The value is determined by measuring the amplitude and phase of a calibration signal which traverses the complete path. Specifically, a single calibration signal is coupled into the feed aperture and travels down each return feed element path along with the S-band return traffic. This signal is then received in each of the GBBF feed element inputs. The GBBF processes the calibration signal using a correlation receiver and determines the relative amplitude and phase offset between each of the elements.

An important goal during the calibration architecture design was to have flexibility to solve any unforeseen challenges that might occur during the deployment. This goal, along with the complementary goal of simplifying the satellite and putting complexity on the ground, drove the decision to generate the return calibration signal on the ground where changes could easily be made. In Fig. 11 this is shown as the return calibration signal, 32 Kcps Walsh + PN waveform, which travels up to the satellite on a special calibration channel. The satellite receives the signal, converts it to the S-band receive band, and couples it into the feed aperture of each element.

A calibration network is used on the satellite to couple the calibration signal into each of the feed apertures. Any amplitude and phase variations that the calibration signal experiences on the uplink path before coupling into each feed element will be common between each of the return feed element paths. Therefore the uplink variations will not contribute to the differential amplitude and phase measurements of each feed element path in the GBBF and the desired beam will be formed correctly.

GBBF Ground Equipment

The GBBF Subsystem implementation is shown in Fig. 12. The GBBF ground equipment consists of four elements: the Beam forming Element (BFE), Management Element (ME), Diagnostics and Test Element (DTE), and Pointing Beacon Station (PBS) Element.

The BFE is the computation engine of the GBBF Subsystem and performs the high-speed real-time computations. On the return path, the BFE converts and shapes the analog element signals, received from the satellite through the Radio Frequency Subsystem (RFS). The BFE sends the signals for each beam to the MSS Communications Processing (MCP) Subsystem, where the user beam signals are processed and provided to the end user. In the forward path, the BFE processes user beam signals from the MCP and generates the signal data to apply to each feed element. The BFE develops the channel signals from the feed element signal and converts each to an analog Intermediate Frequency (IF). The channel signals are then transmitted to the satellite by the RFS. The BFE includes the generation

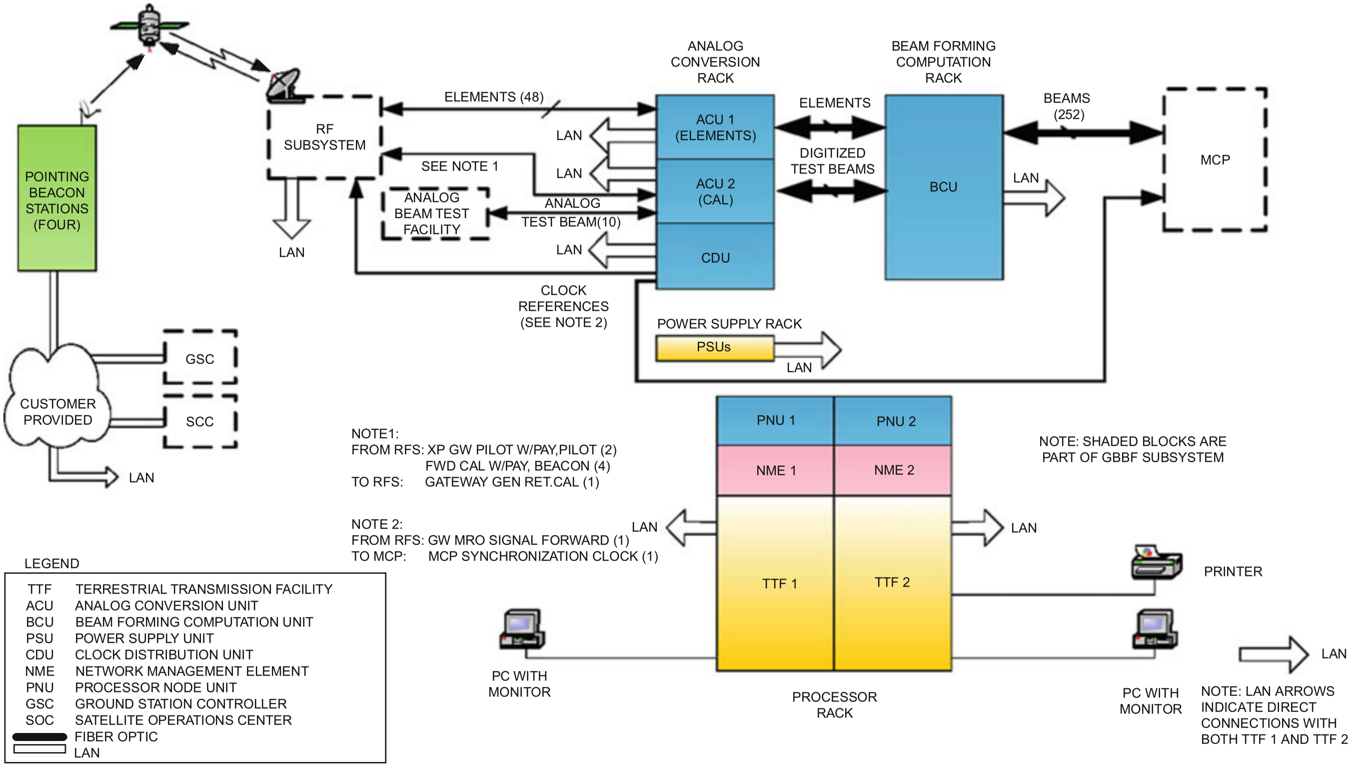


Fig. 12 GBBF subsystem implementation

and processing of the calibration signals as described in the architecture description.

The ME is responsible for the monitoring, control, and management of the GBBF Subsystem and the PBS element. The ME communicates with the Gateway System Controller (GSC) for the purpose of system monitoring and control and with Satellite Control Center (SCC) in order to obtain ephemeris data update.

The DTE performs the role of GBBF Subsystem diagnosis and testing. It has both online and offline diagnostic capabilities. The ME exercises the diagnostic tests and records the outcome.

The PBS Element is comprised of four Pointing Beacon Stations (PBSs) that are positioned in four designated geographic locations within the Continental US (CONUS). The PBSs transmit pointing beacon signals which are received by the satellite in each feed element and passed through to the GBBF along with the return user signals. The BFE forms special monopulse beams at the PBS locations to track and electronically steer the user formed beams to compensate for satellite motion and beam pointing errors.

Inclined Operation of MSS Geo Satellites

One of the challenges in system design is balancing the cost and complexity of the satellite including launch with the cost and complexity of the ground segment. An interesting solution in the MSS to this problem is to allow the Geo satellite to drift in north–south inclination. This reduces launch costs and provides more options of launch vehicles by eliminating the need to launch expensive station keeping fuel. Then, the GBBF is employed to compensate for the orbital inclination by reshaping and pointing the spot beams as needed.

This problem arises due to a combination of a need for a large satellite and the need to operate the satellite served user terminals with those using the same frequency band in the ATC. Although the GBBF simplifies the satellite when compared to onboard processing approaches, the satellites are still very large due to the multielement beam array and the corresponding payload hardware. In order to operate with the ATC and ease the frequency reuse planning, it is important to keep the satellite spot beams aligned with the geographic cells.

When a Geo satellite is used to provide Fixed Satellite Services (FSS), such as the broadband satellites, the orbital position is controlled and held within tight limits referred to as the station keeping box. For example, the box dimensions might be +/− 0.05° horizontally and vertically from the orbital center. This is needed for FSS systems, so the user terminals can have inexpensive fixed pointed dish antennas which only require alignment to point toward the satellite during initial installation.

Since the MSS Systems are designed to communicate with user terminals with near omnidirectional antennas, the satellite north–south station keeping can be removed. North–south station keeping is expensive in fuel and is often the limiter on the operational life of a satellite. Without station keeping, the inclination of the satellite orbit increases continually at about 0.8°/year. Thus, for a 15-year operational

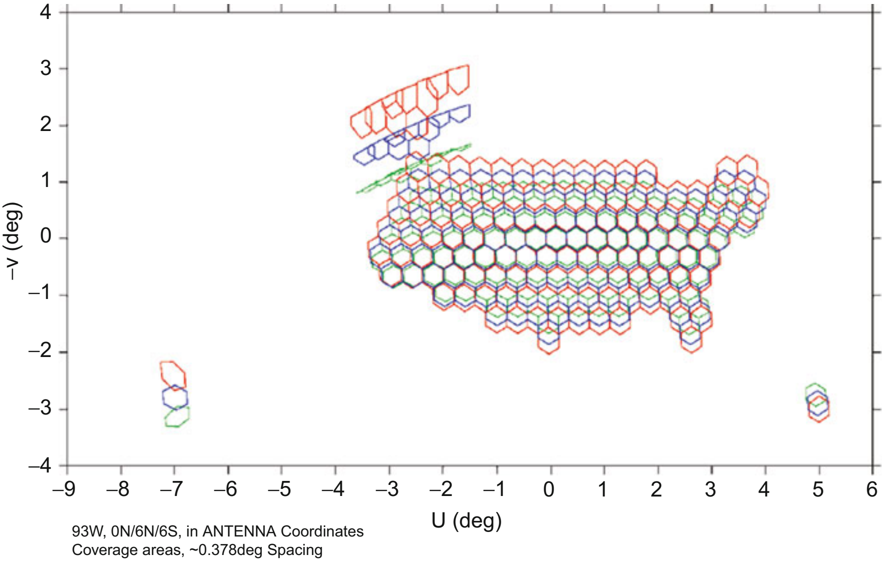


Fig. 13 Projection of beam cells at −6, 0, +6 inclinations

life, the satellite may be initially placed in a (−6°) inclination, and allowed to drift up to (+6) degrees.

While the satellite drift in inclination does not cause a problem for the mobile user terminal pointing to the satellite, it does give rise to an issue with the spot beams from the satellite to the user. As the satellite moves through the inclination, a fixed spot beam from the satellite will project on the earth with a different size and location. This is illustrated in Fig. 13 which shows that in the center of CONUS, the effect is small; however, for beams near the edge, the difference is significant.

The GBBF has the ability to compensate for this effect by using different beam forming coefficient weights to adjust for the satellite's position in the inclined orbit. Thus, the spot beams can remain in alignment with the geographic cells and use the same frequency reuse tables.

The combination of eliminating the need to fly fuel for north–south station-keeping and using the GBBF flexibility to keep the beam shape and pointing aligned with the cells addresses this challenge.

MSS User Terminal Links and MIMO

Even with ATC, there is a need for the satellite phone to provide a similar performance and user experience as the terrestrial network for commercial viability. Thus a key challenge is to improve the link performance between the satellite and the satellite phone. In addition to larger antenna apertures on the satellite and adaptive

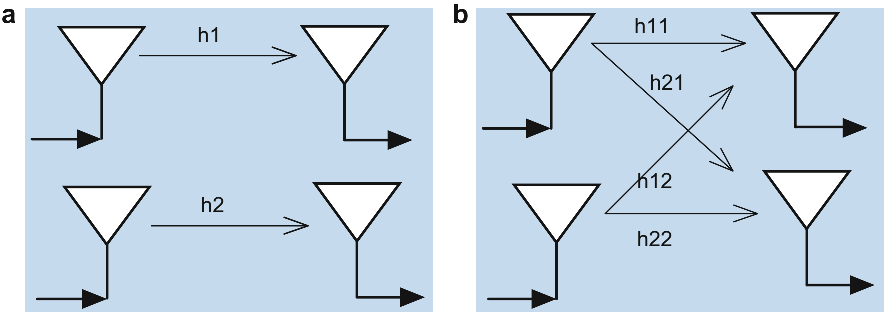


Fig. 14 Doubling of capacity by doubling the number of paths: (**a**) SISO system and (**b**) same design that allows reception of the signal from each transmitter by both receivers, making it a MIMO system

coding and modulation on the links, consideration of advanced satellite phone antenna technologies and diversity is underway.

When Multiple Input/Multiple Output (MIMO) systems were described in the mid-to-late 1990s by Gerard Foschini and others, the astonishing bandwidth efficiency of such techniques seemed to be in violation of the Shannon limit. But there is actually no violation because the diversity and signal processing employed with MIMO can, in effect, transform a point-to-point single channel into multiple parallel channels as depicted in Fig. 14. The gains, however, come not from an increase in power but from an opportunistic mitigation of channel fading. The required SNR of a fading channel is often more than 20 dB higher than an Additive White Gaussian Noise (AWGN) channel. MIMO designs allow us to operate in a fading environment but reduce the required SNR significantly through introduction of space and time diversity.

Pieces of what we now call a MIMO system have existed for the last 30 years in the form of phased array antennas. Adding additional capabilities to phased array antennas, such as adaptive change of weights, leads to the evolution of smart antennas. The first application of these smart antennas was with digital TV set-top boxes where the antenna automatically adjusts array weights to maximize received signal gain. On the transmit side, phased arrays or Multiple Element Antennas (MEA) are used on transmit towers to transmit multiple channels over multiple antennas as well for beam forming. For satellite application, beam forming can be an enabler for forms of MIMO.

In fading channels, the multiple transmit and receive antennas create both Tx and Rx diversity. This is a way to increase data rates, by transmitting data through N different channels. And even if higher data rates are not the goal, the use of different paths provides greater robustness. MIMO systems use *space-time signal processing*, whereby time is complemented with the spatial dimension inherent when using several spatially distributed antennas (at the transmitter and receiver). Such systems can improve bit error rate (BER) or they can increase capacity, or both, without expending any additional power or bandwidth.

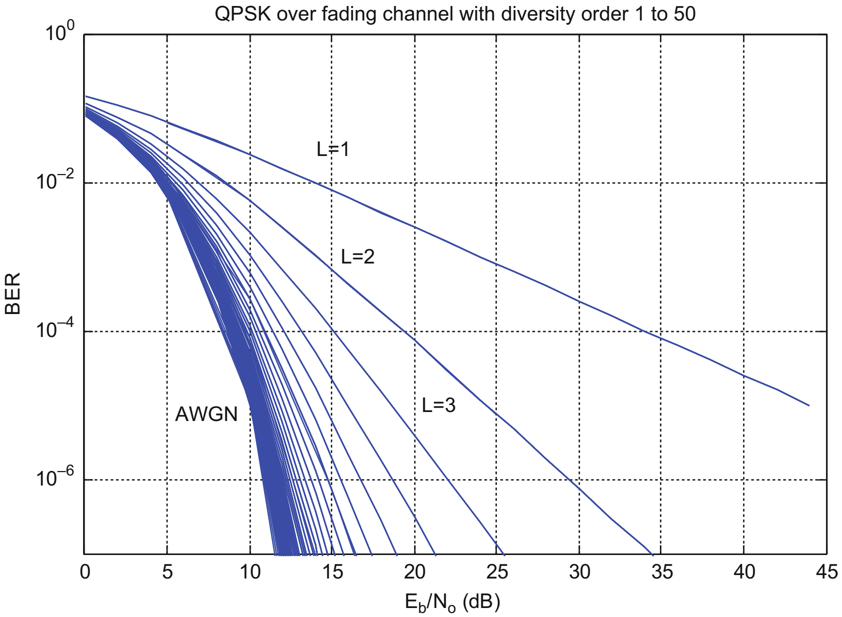


Fig. 15 Gains possible as more antennas are added to a link, at BER of 10–2, A SISO channel requires 14 dB, with two antennas, this goes down to 8 dB and with one more antenna, the total gain is (14–6.5)–7.5 dB

For wireless communications, the new paradigm introduced with MIMO is: Instead of combating multipath, we exploit it by transforming a point-to-point channel into multiple parallel channels (a matrix channel), thereby achieving greater capacity (and/or robustness). Special version of trellis coding that operates over the multiple antennas is an integral part of the capacity enhancement usually described as a sum of increased array gain, higher SNR, and multiplexing gain (higher data rates). The capacity calculation of MIMO indicates that adding a small amount of diversity (more antennas) is sufficient to provide large gains, with diminishing returns as more antennas are added. In Fig. 15, the performance is seen to approach that of an AWGN channel as the number of antennas keeps increasing.

MIMO systems are employed in channels that experience several types of fading, multipath and Doppler being two main areas. A typical satellite link is a line-of-sight link. However, as many new systems have direct-to-user components, the channels have to cope with Doppler and fading. Due to the large delays, the channel knowledge that is required for MIMO is a challenge to both the transmitter and the receiver in a satellite link. However, it is not an insurmountable problem, and some satellite systems such as DBSD have explored technology to overcome it.

For a pure line-of-sight (LOS) link, MIMO does not offer a huge benefit, but when a link suffers from fading such as in an urban environment, mountainous terrain with shadowing and during rain, MIMO can increase link performance by a significant amount. A normal SISO link can be turned into a SIMO link when receiver is equipped with multiple antennas located at just half a wavelength apart. The multipaths created by this environment can then be advantageously combined to reduce the losses usually allocated for these channels. As we see in Fig. 15, using three antennas instead of one can reduce the required Eb/N0, under a fading condition, by approximately 8 dB, enough to make up for rain loss at Ku-band. Future work in developing MIMO receivers for satellite applications is expected to help address the discussed challenges in MSS systems.

Broadband Satellite Systems and Internet Access

Broadband Satellite Systems are used to provide Internet access to users via a satellite terminal. The physical integration of the satellite system with the terrestrial network occurs at the gateways. The satellite user terminal side, commonly called Customer Premise Equipment (CPE), communicates via two-way links through the satellite to the Gateway which is connected to the Internet. Since the user terminals only communicate with the satellite, the MSS system challenge of dual handsets and frequency sharing with terrestrial cell towers using the same frequencies is avoided.

Nevertheless, significant challenges remain. Some were encountered and addressed during the development of the current systems. More are being addressed to enhance the existing systems and in the development of the next generation systems. These challenges stem primarily from the need to provide each user of the system an acceptable and desirable experience, while providing the same level of performance to enough user customers for the system to be a commercial business success. The challenges can be categorized into three areas:

1. Bandwidth and Capacity: Individual User data rates and total capacity of the system
2. Latency over the Satellite links and the users' Internet experience
3. Interfaces and Standards

Figure 16 illustrates a typical Broadband Satellite Internet Access System. These systems consist of Gateways, Satellites, and User Terminals. Multiple user spot beams are provided to maximize the total capacity of the system through frequency reuse. The architecture presented in Fig. 16 is referred to as a hub spoke architecture. Each user terminal in a spot beam corresponds to spokes connected to a central gateway hub. The Gateways are typically interconnected via the Internet in a private network as well as serving to access the public Internet for the users. The user links and Gateway links consist of forward links from the gateway through the satellite to the user terminal and return links from the user terminals through the satellite and to the gateway.

In addressing challenge 1, broadband satellites must take maximum advantage of their frequency assignments to maximize the total capacity. This is accomplished by using polarization diversity and frequency reuse. As the user beam count is increased, the need for multiple gateways arises as each user signal must pass

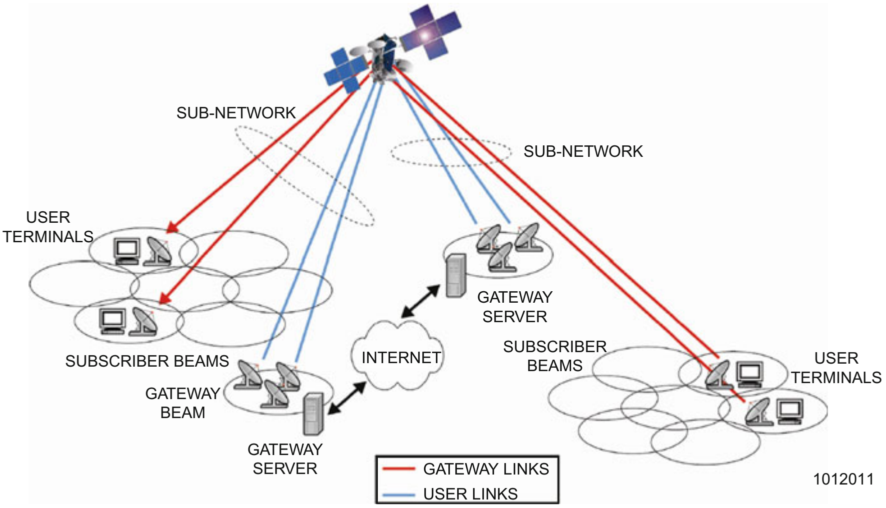


Fig. 16 Broadband satellite for internet access

through a gateway in the hub spoke architecture. Other architectures provide for user beam to user beam connectivity as will be discussed later when considering an Internet router in space.

The frequency bands used for broadband access are currently C, Ku, and Ka. While C and Ku are typically limited to 500 MHz of spectrum, Ka can provide 1,000–1,500 MHz of licensed bandwidth. The industry is also looking forward to V-band for increased capacity.

In 1998, the FCC took another step to address the increasing needs for satellite bandwidth and to ensure technological development with preliminary allocations at Q-/V-band. The current allocations provide at least 4 GHz of bandwidth which is considerably greater than the satellite systems in use today. RF components for both ground and space are already in development, and its likely broadband access satellite systems will make use of this increased capacity in the next 5–10 years.

The total capacity of the satellite is limited by the number of beams, the frequency reuse scheme, the bandwidth available to each beam, and the information data rate of the selected modulation and coding for each beam. Simply, 60 spot beams with 500 MHz of bandwidth and 2 bits/Hz signaling provides 60 Gbps of available capacity. The total capacity of a given satellite network is typically defined by adding the forward and return capacities, with the current systems targeting hundreds of Gbps.

Nevertheless, the maximum capacity of a satellite Internet access system is much less than that provided by terrestrial fiber infrastructure due to the RF bandwidth constraints. Commercial business success for the satellite system, then, must carefully assess the infrastructure build-out plans and market needs for services and applications.

O3b Networks Limited founder, Greg Wyler, is not trying to compete against fiber infrastructure. Instead, he has defined the company's mission to make the Internet accessible and affordable to those who remain cut off from the information highway (3B Networks). O3b is short for the Other Three Billion, referring to the nearly half of the world's population that has little or no access to the World Wide Web. Many investors are on board, including: SES S.A., Google Inc, HSBC Principal Investment, Liberty Global Inc, and North Bridge Venture Partners.

O3b considers their mission a social responsibility. "Every member of the O3b team is passionate about bringing affordable, state-of-the-art broadband services to the three billion people who have been denied them for reasons of geography, political instability and economics" (O3B Web).

O3b Networks plans to launch a constellation of satellites in an orbit only 8,000 km from earth. This is four times closer than geosynchronous satellites. Consequently, the O3b satellites will continuously move relative to the user as they circle the Earth. The Internet traffic will be provided by the satellite over the region. Before the satellite passes out of range, the traffic will be routed over the next satellite coming through.

The MEO orbit and architecture reduces the round trip latency of a geosynchronous orbit from over 500 ms to approximately 100 ms. O3b believes this is important and will allow their system to provide the user with a Web experience significantly closer to terrestrial systems such as DSL or Optical Fiber. The satellites are also equipped with steerable antennas and support a reconfigurable RF payload to adapt to changing needs and market evolution.

Nevertheless, they recognize the demand for connectivity goes beyond the bandwidth they can provide with the satellite system. Consequently, their plan includes partnering with PC and device manufacturers to eliminate bottlenecks in the system to provide the most capacity possible.

Hughesnet, WildBlue, and Viasat are continuing with their Geo Satellite orbit approach to providing Internet access systems. These systems eliminate the need to develop efficient satellite hand-off methods, and the developers concentrate on improving the network efficiency with advances in modulation and coding, and reduced overhead and latency in the network protocols.

Protocols and Network Performance

A significant challenge in the integration of satellites and terrestrial networks involves the development and acceptance of the ways digital messages are formatted and the rules governing how these messages are interchanged – in short, the protocols. Protocols describe the syntax, semantics, and synchronization of communications between the wide range of entities that are part of an integrated satellite-terrestrial network. Protocol developers must ensure effective communication performance and guarantee that entities that develop products that follow the protocol standards are able to communicate.

This section will present some general development frameworks in which protocols have been and are being developed. These frameworks, called Protocol Reference Models, map a process by which protocols can be developed and products implementing these protocols can be commercialized.

Some of the protocols currently in use in both terrestrial and satellite-terrestrial integrated systems will be discussed along with the challenges in protocol development to improve network performance in integrated satellite-terrestrial architectures.

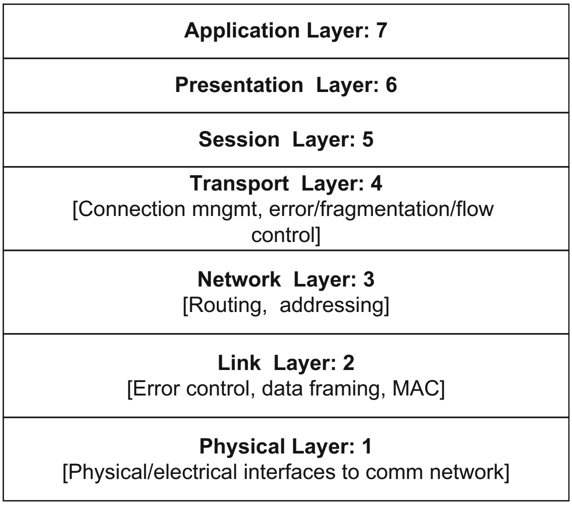
Reference Models

A protocol reference model provides a generally agreed-upon framework in which communication protocol development can take place. To do this, a model categorizes the role for protocols in terms of functionalities in the overall execution of the end-to-end communication of message data.

The types of reference models discussed here are called layered reference models. Figure 17 illustrates the general structure of a layered reference model, the International Standards Organization (ISO) Seven Layer Reference Model (RM) for Open Systems Interconnect (OSI) (Gibson 1997, page 570). Each layer of the RM defines a set of well-defined functions in the context of overall communication. The layer operates by sending messages to the corresponding layer (called the peer layer) of the remote entities with which it is communicating. Each layer also provides services to the layer above and uses the services of the layer below. Peer layer messages can consist of both user and control data.

Figure 18 shows the total communication process. Note that conceptually, each layer communicates with a similar peer layer; however, in practice, the resulting protocol message units of a particular layer are passed by means of the services of the

Fig. 17 ISO reference model



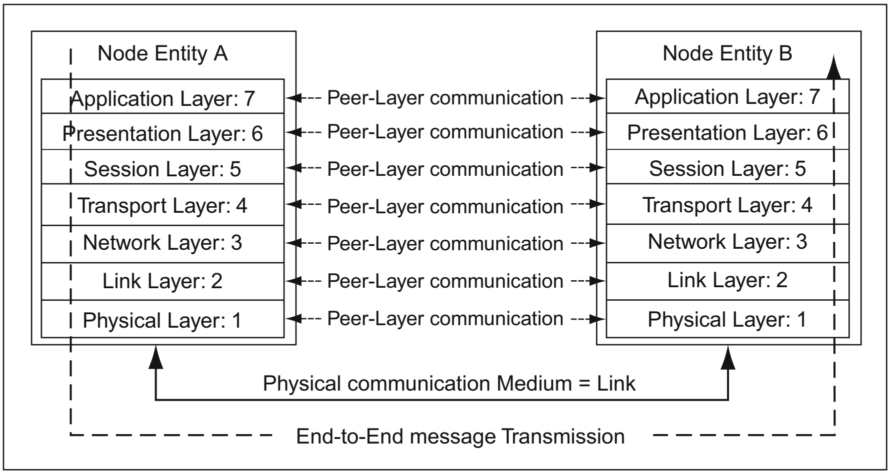


Fig. 18 ISO reference model communications process

lower layer. As illustrated in the figure, node entity A sends a message to node entity B, with all bits flowing over the physical link. The digital message from node A employs the protocol specified services of the descending layers (7 to 1) until generating message bits with the peer layer protocol message unit overhead included from each layer. Through this process, protocols at each layer encapsulate the inputs from the layers above. Data received at entity B is processed in reverse. Each peer layer message is extracted from the encapsulation and the resultant is passed to the next higher layer until the message is received by the final destination application at Layer 7.

Two examples illustrate some of the different ways the peer layer messaging can work. For example, consider layer functions that perform data translation type of operations, like data compression, data encryption, and forward error correction (FEC) encoding; the transformed data may be simply messaged to their corresponding remote receiving peer layers without any additional control or instruction exchange. Upon receipt by the remote peer layer, the message data are decoded per the definition of that layer's function, e.g., decompression.

At another extreme are the many variations of the Transport layer protocol called Transport Control Protocol (TCP), which is discussed in more detail below. The TCP protocol sends a variety of messages to the remote receiving layer including set-up signaling messages, return receipt acknowledgments, and segments of the entity-to-entity digital message that is being transmitted.

Actual protocols that conform to a RM must be defined by specification standards following guidelines of the RM. There are two types of standards for each layer: the protocol standards, defining how peer entities at this level communicate (the rules, conventions, and message formats for peer layer interactions) and the service standards with interface specifications for the services that are provided to the higher layer immediately above. In order of an RM-conforming set of protocols to support community of interacting communication entities, it is necessary to have at least one protocol specified at each layer of the RM. The resulting set of protocols is called a *protocol profile*, a name that is often prefixed by the name of the RM (e.g., a OSI protocol profile). Each level of a protocol profile can have several protocols, each providing a different service to the higher layer.

There are several communication protocol RMs in use that have relevance to the challenges related to the integration of satellites into terrestrial networks. The main three RMs that are discussed below are the ISO seven layer Protocol RM (also known as the OSI RM), the TCP/IP Protocol RM (also called the Internet Protocol RM or Suite) ((Gibson 1997, p. 575) and (Stevens 1994)), and the Broadband Satellite Multimedia (BSM) RM (architecture) (ETSI TR 101 984 V1.2.1 2007).

Role of Reference Models

A principal factor in the successful integration of satellites with terrestrial networks is the development and wide range acceptance of protocols that not only enable an effective and transparent (seamless) working system combining these two elements but also enable this relationship to continually evolve as these two elements advance technologically. The reference models and the development methodology they require are essential for the ongoing process.

The RM methodology provides for the development of layered protocol profiles where each layer's functionality is well-defined and its interfaces with the two adjacent layers rigorously specified. This open systems approach allows for the independent development and implementation of proposals independently for those of the other layers and enables the interchangeability of protocols meeting the layer's specifications. Thus, continued innovation is made possible as is the wide range acceptance of products that implement these innovations. Potential producers can focus on manageable parts of the overall system with reasonable assurance that products resulting from this focused work will be compatible with existing implementations and thus, if superior, have the potential of gaining wide range support. And, of course, wide range acceptance both makes possible and is made possible by increased affordability of protocol implementations.

ISO 7-Layer Reference Model for Open System Interconnect

The ISO 7-Layer RM is illustrated in Fig. 17. The development and issuing of standards for protocols complying with the ISO RM are the work of the International Standards Organization. The ISO RM is also widely called the OSI RM.

The ISO RM yields protocol profiles that are to be implemented in communication networked entities that are linked by some physical medium (radio links, fiber optic links, cable links, IR links, and multiple networks of such links). The highest ISO layer, Layer 7, is closest to the user interface (or application programs); the lowest ISO layer, Layer 1, provides the connection to the physical link. Below is a brief description of the general functionality of each of the ISO layers (Note: For the purposes of satellite/terrestrial network integration, the focus is on the lower four ISO layers):

Layer 1 – Physical Layer: defines the means of transmitting/receiving raw *bits* rather than logical data packets over a physical transmission medium connecting network nodes – the bit stream may be grouped into symbols that are converted to signals for transmission (and conversely, for reception) and provides an electrical, mechanical, and procedural interface to the transmission medium, including (as appropriate to the medium) the specifications for: the shapes and properties of the electrical connectors, the frequencies, the modulation scheme to use, and similar low-level parameters.

Link Layer 2: provides the functional and procedural means to transfer data, in *frames*, between peer Link Layers, using hardware addresses and focusing on local delivery, addressing, and media arbitration; and may also provide for the detection and correction of physical layer errors. In the computing environment, the Link Layer is often implemented in software as a "network card driver."

Often the Link Layer is subdivided into the two following sublayers:

Link Control Sublayer (data multiplexing/decoding, error control – rebroadcast of damaged frames, acknowledgments, error notifications, flow control – what to do when a frame is received)

Medium Access Control (MAC) Sublayer (multiple access control, framing. synchronization, data queuing/scheduling, hardware addressing, QoS control, level 2 switching)

Network Layer 3: This layer provides the functional and procedural means of transferring variable length data sequences (called *packets*) from a source to a destination host via one or more networks while maintaining the quality of service functions; includes addressing, packet queuing, and packet forwarding.

Transport Layer 4: This layer provides transparent transfer of data between end users, providing reliable data transfer services to the upper layers. The Transport Layer controls the reliability of a given link through flow control (of *segments*), segmentation/desegmentation, and error control. There are a range of connection and reliability service options available to the upper layer.

OSI Layers 5–7 – Collectively, Application Layers (Session Layer, Presentation Layer, Application Layer): Collectively (and almost universally) referred to as the Application Layers, – for complete descriptions of OSI layers 5–7, refer to Gibson (1997): establishes, manages, and terminates the connections between the local and remote application, manages the quality of the connection (reliability, security, full-duplex, half-duplex, or simplex operation), and connection setup – checkpointing, adjournment, termination, and restart procedures. These layers contain the functionality that supports user software applications.

Note the vocabulary lesson presented in these above layer descriptions relative to the naming of the *protocol data units* that are used for peer layer-to-peer layer communications: at Physical Layer, *bits*; at Link Layer, *frames*; at Network Layer, *packets*; and at Transport Layer, *segments*.

TCP/IP Protocol Suite

Figure 19 illustrates the TCP/IP or Internet Protocol Suite or Reference Model. The development of protocol and standards based on this RM is under the aegis of the Internet Engineering Task Force (IETF), which is now a standards organization for the protocols for use on the Internet. The TCP/IP Protocol Suite is not quite the top/down comprehensive design reference for networks as is the ISO RM. It was formulated for the purpose of illustrating the logical groups and scopes of functions needed in the design of the suite of Internet working protocols of TCP/IP, needed for the operation of the Internet. In general, direct or strict comparisons of the OSI and TCP/IP models should be avoided. TCP/IP Protocol Suite does not have specific protocols for each ISO RM layer; however, loose comparisons are often made between the two. The Internet Protocol Suite is a layered model widely in use for computing systems networked on the Internet and the IETF has established standard development procedures (Brandner 1996), which lead to open system, vendor-independent standards based on the TCP/IP Protocol Suite.

Figure 19 provides the names of each layer of the TCP/IP RM and provides a general description of each layer's functionality by reference to the ISO RM (again, these are not strict comparisons). More details of the layer functionality of the Internet RM are found at CH-CRC 1997, p. 575.

Broadband Satellite Multimedia Protocol Architecture

Figure 20 introduces a RM based on the needs of the integration of satellites with terrestrial networks, the BSM Protocol RM (ETSI TR 101 984 V1.2.1 2007). The

Fig. 19 TCP/IP protocol suite

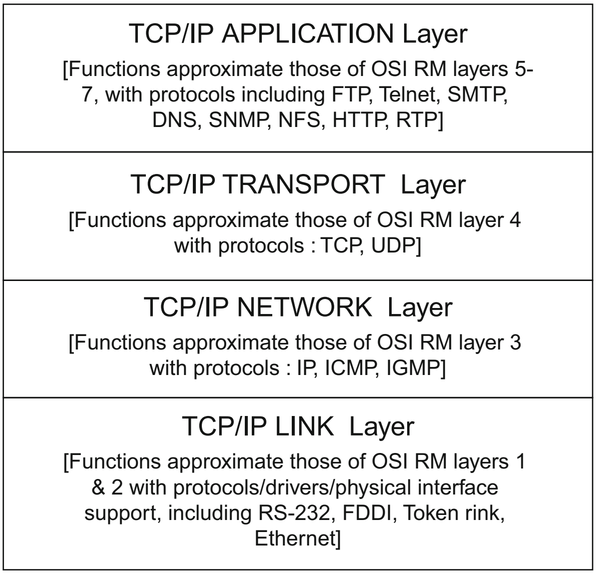
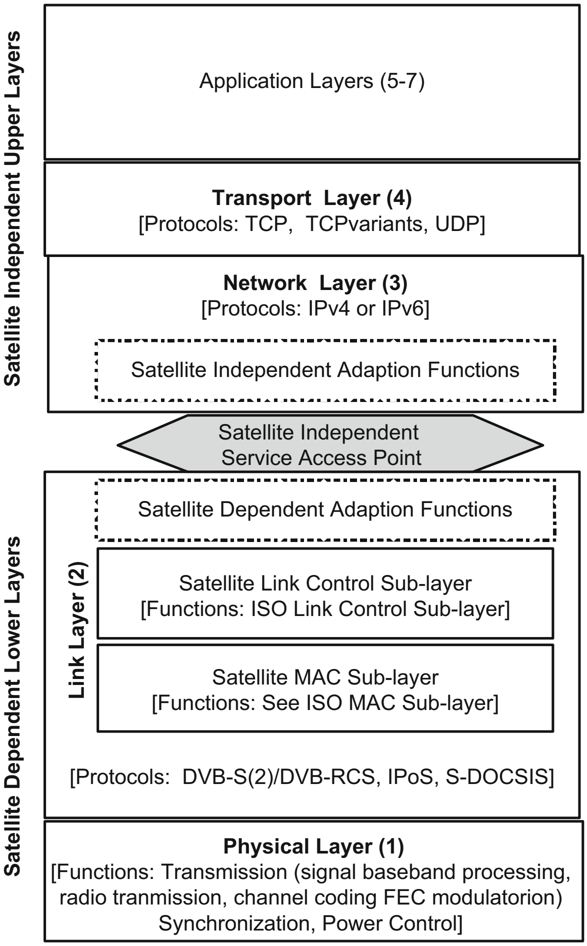


Fig. 20 ETSI-BSM protocol architecture



BSM RM is developed and maintained by the third international standards group, the European Telecommunications Standards Institute (ETSI). ETSI has issued a number of standards applicable to the integration of satellite links into terrestrial networks. The BSM RM is a layered model that has been partitioned into a satellite-dependent set of layers (the lowest two layers) and the satellite-independent layers. This RM has an explicitly defined interface, the Satellite-Independent Service Access Point (SI-SAP), between the satellite-dependent and satellite-independent layers. The intention is to focus the ETSI's satellite link to work on the specification of protocols for the lowest two satellite-dependent layers and to work on the standardization of the SI-SAP. Thus, once the Network layer has been enhanced to interface with the satellite-dependent layers (the Satellite Independent Adaptation Functions), layer 3 and above development can proceed independently of satellite concerns. In practice, this approach has some difficulties that are discussed in the section below TCP Challenges.

BSM RM in its lower two layers can be compared rather strictly to the corresponding layers of ISO RM (again loosely, as in the case of the TCP/IP RM). A number of protocol profiles conforming to the BSM RM at these lower layers have been developed, specified, and are in use in satellite links carrying Internet traffic. Three will be mentioned here:

- DVB Standards: Digital Video Broadcast over Satellite (DVB-S2) (ETSI-DVB-S2 2009) for the forward link and Digital Video Broadcast-Return Channel over Satellite (DVB-RCS) (ETSI DVB-RCS 2009) for the return link. These standards have video broadcasting as their base and are ETSI's principal standards for BSM RM layers 1 and 2.
- IP over Satellite (IPoS) (ETSI TS 102 354 V1.2.1 2006) – mostly a pointer to TIA Document TIA-1008-A, and (Hughes Network Systems 2007). These protocols are used by Hughes Network Systems in its offering of Internet access via satellite.
- Satellite Data Cable Service Interface Specifications (S-DOCSIS) – the satellite version of DOCSIS (2010) and (DOCSIS 3.0 2010) the Physical and Link Layer protocols that are used for Internet accesses via cable modem.

The Satellite Independent layers of the BSM RM above layer 2 can be compared largely with the TCP-IP RM, where the Network Layer (3) uses Internet Protocol version 4 (IPv4) or version 6 (IPv6). And at the Transport Layer (4), User Datagram Protocol (UDP) and a version of TCP are used. In the following section, it is shown that challenges arise to keeping the development of BSM-layer 4 protocol independent from the implementation of the lowest two DSM layers, the Satellite-Dependent layers.

TCP Challenges

One of the ongoing challenges facing the integration of satellites and terrestrial network concerns the protocol implementation of the Transport Layer protocol, TCP (Stevens 1994). The importance of this protocol layer is emphasized by the amount of traffic currently on the Internet directly related to TCP. Analysis of the traffic on the Internet backbone between 2006 and 2009 has shown more than 80 % of the bytes and the IP packets at the Network layer level carrying TCP traffic (Dusi and John 2010). Thus, satellite links and systems cannot be well integrated with Internet and other terrestrial system unless traffic resulting from TCP protocols can be effectively and transparently transported by the satellite link.

Basic TCP

The original TCP protocol was specified in September 1981 (Transmission Control Protocol 1981) to provide a reliable transmission of data over the Internet, where "reliable" means that the service provides a guarantee that all the user data will be received without error. To accomplish this, TCP must establish a dialogue with the receiving peer layer and then retransmit segments of data that have been corrupted or have been lost. This process must continue until the receiving TCP has accurately received all sent data. The mechanism to determine if a data segment was successfully received by the receiving TCP is for the receiving entity to acknowledge with a message, an ACK, that tells the sender that the segment was successfully received.

Thus, the TCP at the sending node must retain a copy of a transmitted segment until an ACK for this segment is received. Consequently, the TCP of the sending node must buffer all sent segments that remain unacknowledged. The size of this buffer is called the congestion window. The receiving TCP must also buffer received segments that are missing at least one segment that should have already been received. TCP partitions the total data to be transmitted into an ordered set of segments, which correspond to the order originally sent. The receiving TCP passes the full, accurately received data message to the higher layer; it cannot send up the data of segments out of order.

Segments may be missing for several reasons. They may have been lost to some collision or due to queue overflow at some intermediate router – both of these losses are considered to be traffic congestion related. There may also have been loss of a segment due to an uncorrectable data corruption due to a bit error in the network links. Another network phenomenon that may give the appearance of a segment loss is the reordering of segment arrivals from the original transmission order; this is due to the possibility that the transmitted segments may follow different network routes to the receiver, each of different transit times and of different congestion levels.

When the receiving TCP experiences an out-of-order segment arrival (when a segment arrives before one of its predecessors), it cannot immediately know whether this is because of a segment loss or due to a delayed arrival of a prior segment. The arrival of an out-of-order segment is ACK'ed (called a *duplicate ACK*) not with the segment number of the arrived segment, but with the segment number of the last segment received in order (all preceding segments already having arrived). Of course, the sender upon receiving an ACK or a duplicate ACK knows that all segments preceding the one whose number is in the ACK have been successfully received and, with the receipt of a duplicate ACK, that the next segment has not yet been received.

The simple description of the original and basic elements of the TCP protocol algorithm is:

- *Initiation*: There is a starting exchange of messages to initiate the TCP session, initiated by the sender.
- *Slow Start*: The sending TCP starts with one segment and increases the *rate* of segment sending based on the number of (success) ACKs received (until a parameterized threshold is reached, at which point the rate increases linearly until a limiting rate is reached).
- *Retransmission*: If an ACK for a segment has not been received within a given time frame based on a computation of latency for a round-trip, a *timeout* occurs and the segment is retransmitted.

- *Recovery*:
  - If a timeout occurs, resume retransmission with slow start.
  - If a duplicate ACK is received, halve the current transmission rate.
- *Terminatio*n: The TCP session ends with a message handshake, initiated by sender.

A full description of the basic TCP protocol is given in Stevens 1994.

Network Environment and TCP Optimization

The original TCP provided a basic mechanism that was discovered to not be very efficient in terms of data throughput. Its effectiveness is very dependent on the network environment of the TCP session. Among the important factors affecting TCP efficiency include:

- Network latency (delay time to receive the ACK, the round-trip-time, (RTT))
- Size and symmetry of the bandwidth available to the sender/receiver
- Rate and the causes of segment loss
- Topology of the network (single path or multiple path possible for segment transmission –affects the variability and ordering of segment arrival)
- Network traffic conditions in which the TCP-based communication takes place

From the basic form of TCP originated in 1981, a number of varieties have been and continue to be developed. Each variety is usually developed to enhance the effectiveness of the protocol in a specific range of network environments: The performance of each is optimized for this specific environment. At least three major environmental network types can be distinguished:

- Terrestrial Internet – characterized by small latency (short RTT) links, symmetric links, segment loss due to congestion not to corruption (small BERs), and a topology allowing for multiple routes; the TCPs for Internet are discussed in the following section
- Satellite links – characterized by large latency (long RTT), generally single links (to/from), often unequal bandwidths for user in the forward and return directions, links with high BERs, and often single link communications
- Long, very high-speed/capacity links – characterized by large latency (large RTT), symmetric with high bandwidth links which are of low-loss (low congestion and BER) (it is noted that these “long, fast” links are of growing importance for Internet and there are several TCP variants developed to optimize performance in networks with these type of links: TCP CUBIC, TCP HighSpeed, HTCP (Hamilton TCP), TCP Illinois, and TCP Scalable – for all these TCP variants, refer to Caini and Firrincieli (2009)

Given the complexity of the TCP algorithm, it has not been feasible to provide analytic verification of the superiority of one TCP protocol over another. Judgments of this type are normally based on end-to-end network simulation results where a valid representation of the relevant factors modeled.

In fact, modeling, simulation, and emulation are extremely important tools in all phases of a networked satellite-terrestrial system: architecture, design, deployment, and operation. Due to the complexities of the interaction between the protocol layers, including the resource allocation algorithms, these tools are required if the system developer wants to understand the true achievable network performance, efficiency, and capacity with heavily loaded traffic. The tools may also be used to test and beta-license new applications without consuming the actual satellite network resources (Lindberg-Walker 2003).

Classical TCP

The basic TCP protocol was soon found to be very efficient in terms of segment throughput. This realization led to a sequence of improvements for the Internet environment that culminated in the early 1990s in what is called here the Classical TCP, which is best represented by a TCP protocol called New Reno. The sequence of protocols leading to New Reno were: TCP Tahoe, which speeded up the retransmission process, an improvement called Rapid Retransmit (Stevens 1994); TCP Reno, which speeded up the recovery process in a process called Fast Recovery (Stevens 1994 and Allman and Paxson 1999); and finally, TCP New Reno, which improves on TCP Reno's Fast Recovery (and still called Fast Recovery) (Floyd and Henderson 1999). New Reno is probably the most widely adopted TCP variant (Caini and Firrincieli 2006a); it has been one of the TCP available in the Linux kernel, where it is called Reno (Caini and Firrincieli 2009).

Satellite Link TCP

During the 1990s, as satellite links use was growing, it became evident that the classical TCP variants would not function well in this new environment. The satellite link environment is characterized as: High latency (with RTTs often more than 550 ms), high segment loss due to channel error and not congestion, usually a single pipe topology (between the Gateway (GW) and the Satellite Terminal (ST)), and often asymmetric in that the forward link (GW to ST) is allocated more bandwidth for user traffic than the return link (ST to GW).

By the end of the 1990s, numerous possible solutions had been proposed (Allman et al. 1999). Many of these proposed approaches were developed and simulated over the subsequent decade and this development process is continuing (Joing et al. 2009; Roseti et al. 2010). Below is a list of some of the recent or often cited TCP variants or options that have been developed to improve TCP performance on satellite links (Note those TCP variants that accelerate New Reno's fast recovery rely on the usually valid assumption that in satellite links segment loss is not an indication of congestion but to channel error – so there is no reason to slow down transmission to avoid congestion):

- SACK (Mathis et al. 1996), an option that can be used with other TCP protocols, allows TCP receivers to inform TCP senders exactly which packets have arrived. SACKs allow TCP to recover more quickly from lost segments, as well as avoid needless retransmissions. SACK is a TCP option available in recent Linux kernels (Caini and Firrincieli 2009).
- *TCP Hybla* (Caini and Firrincieli 2009), a TCP protocol that augments the classical TCP by the use of the SACK option and by increasing the New Reno's fast recovery rate by making the rate of increase proportional to RTT (7-Caini, C; Firrincieli, R. – Hybla, West). TCP Hybla is a TCP variant found in the Linux kernel.
- *TCP Westwood* (Caini and Firrincieli 2009) is a TCP variant that augments the classical TCP by increasing New Reno's fast recovery rate by, instead of halving the segment transmission on segment loss detection, setting it to the available bandwidth, thus avoiding harsh slowdown (Westwood's fast recovery algorithm is called *Faster Recovery*). This TCP variant is found in the Linux kernel.
- *TCP Peach* (Akyidldiz et al. 2001;Joing et al. 2009; Luglio et al. 2004) is a TCP variant that used dummy segment transmission to access available bandwidth. TCP Peach speeds up the classical slow start algorithm, calling it *sudden start*, and speeds up the classical fast recovery algorithm, calling it *rapid recovery.*
- To improve TCP over satellite links many other transport layer protocols have been proposed, including recently, TP-Satellite (Joing et al. 2009); SCPS-TP (CCSDS 2006) introduced as part of CCSDS's Space Communications Protocol Specifications protocol suite; and TCP Noordwijk (Roseti et al. 2010).

TCP Performance-Enhancing Proxies (T-PEP)

As mentioned previously, a very major portion of Internet traffic is TCP traffic and the satellite network must integrate with the terrestrial network and handle TCP efficiently for commercial viability. TCP, however, is a connection-oriented protocol and defines the protocol dialogue between the Transport layer peers at the two end points of the communication, whereas end-to-end communications in general traverse the links of several different network environments (classical Internet links, satellite links, long-fast links, and, perhaps mobile links). Different TCPs have been developed to optimize the data throughput performance of each of these link types separately. But, how can the overall end-to-end performance of TCP sessions be optimized?

One solution is to divide the end-to-end TCP session into several shorter ones, each of which crosses a single link type where each dedicated TCP session can be optimized. An approach for implementing this is through the use of Performance-Enhancing Proxies (PEPs) which when applied to the TCP layer are called TCP PEPs (T-PEPs).

Figure 21 provides an example of the use of T-PEPs and is based on an ETSI standard for PEPs (ETSI EN 302 307 V1.2.1, DVB-S2 2009). The protocol architecture includes the Internet (TCP/IP) Protocol Suite and the protocol profile based on the BSM RM. The general model is that of a satellite access system for Internet.

In Fig. 21, an Internet-connected node on the Internet is delivering data content to a Satellite-connected user; this needs to be via a reliable communication service. Without TCP session splitting, there would be one end-to-end TCP session; however, in this BSM PEP architecture (38-ETSI PEP), the data content delivery is via three TCP sessions, each occurring in a different network environment:

- TCP Session A (Internet node TCP to the GW PEP Standard TCP): This TCP session is in the Internet environment.
- PEP-to-PEP TCP Session (GW PEP Satellite TCP to the ST PEP Satellite TCP): This TCP session is principally in a satellite link environment (with perhaps two Ethernet LAN at each end).
- TCP Session B (ST Standard TCP to the Sat-connected node TCP): This TCP session is via Ethernet LAN or perhaps some private local terrestrial Internet work.

In this example, TCP sessions A and B would most likely be a classical TCP variety such as TCP New Reno. However, the PEP-to-PEP TCP session would be a TCP variant that has been optimized to the satellite link – and this would depend on the network environment of this satellite link (GEO/LEO for RTT, bandwidth size and symmetry, link BER, satellite link topology – one or a network; this would also depend on the protocols in use in the Link Layer). Note that in both the GW T-PEP and the ST T-PEP there would be PEP applications that receive and direct TCP traffic from/to the appropriate TCP protocol.

Figure 21 illustrates one type of satellite T-PEP architecture, called distributive. The Transport Layer TCP communications between two communication systems are split into three separate TCP sessions: A – Internet-connected communication system-to-GW PEP; PEP-to-PEP TCP session; B – Sat Terminal PEP-to-Sat-connected communication system.

Alternatively, an integrated architecture is obtained by eliminating the Sat Terminal PEP and using the Sat TCP protocol in the Sat-connected communication system. Other TCP splitting architectures have been studied, including that of splitting the end-to-end session onboard the satellite (Luglio et al. 2004). Additional T-PEP models and techniques are described in (38-ETSI PEP).

TCP PEP architectures offer the opportunity to split the end-to-end TCP session and enable the selection of optimal TCP protocols based on the network environment of each section. However, this approach has some problems and development challenges, three of which are discussed below:

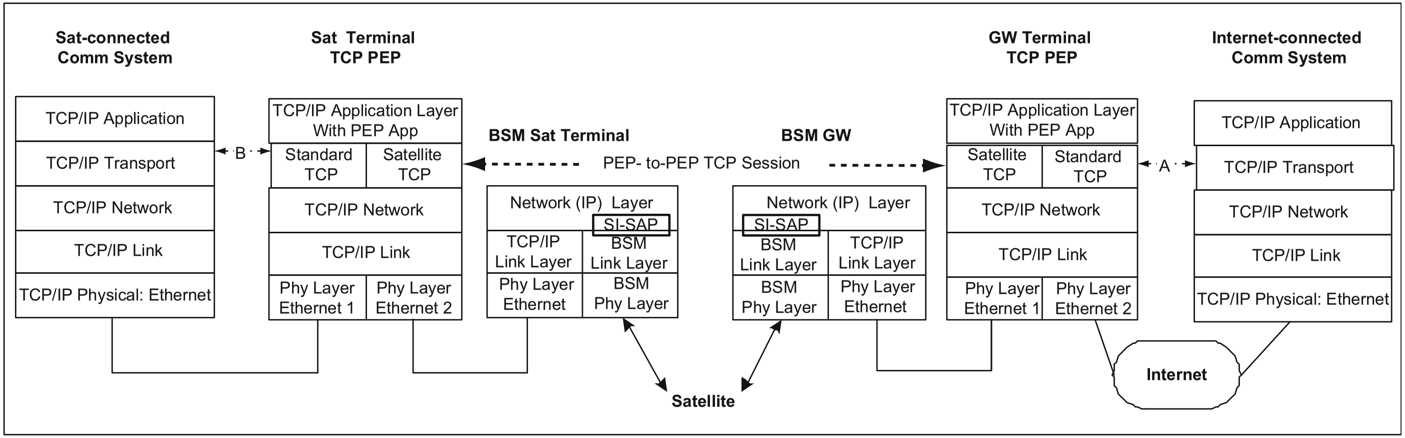


Fig. 21 Distributed TCP-PEP architecture for BSM internet access BSM satellite communication system

- System Complexity: Again referring to Fig. 21, the T-PEP architecture necessitates the addition of one (Integrated architectural case) or two (Distributed case) PEP, each of which must provide the TCP processing and the related PEP application layer processing. The PEP on the GW (Internet) side must provide for each user full TCP session processing and buffering for the two TCP sessions: Session A and the PEP-to-PEP TCP session. This can require a large amount of GW processing and memory capacity, which becomes even more difficult for onboard PEPs. For Integrated PEP architectures, the receiving ST, user terminals, would be required to maintain duel protocol stacks.
- Network Security: The T-PEP is not compatible with the growing use of IP encryption, IPSec. End-to-end IP encryption is not possible because the TCP headers, which are encapsulated as data in the IP packets, are encrypted and thus not accessible for the intermediate processing needed to support TCP splitting using T-PEPs. Thus end-to-end secure communications would necessitate an application level encryption scheme.
- Mobility: The main challenges with mobility and the T-PEP architecture are those related to the GW PEP or, if onboard, the onboard Satellite PEP when a satellite user needs to handoff to either a different Gateway or a different satellite. This can occur when the satellite user moves from one beam to another beam, that is, in a different subnet (see Fig. 16) or with non-Geo Satellite constellations where the satellite passes through the coverage area. Changes in the middle of an end-to-end communication could introduce unacceptable disruption to the user communications as the entire complex TCP state of the GW PEP might need to be managed by another GW. Issues and solutions to T-PEPs and mobility are discussed in 18 – Dublis PEPissues.

Cross-Layer Signaling

The importance of the challenge of TCP optimization has led to some solutions that violate the rules of the layered protocol models. One set of such solutions are those employing means to exchange information between nonadjacent layers, called cross-layer signaling. For example, a cross-layer protocol could provide a means for the Transport Layer to receive information from the Link Layer via message exchanges that are either in-band (using normally passed headers) or out-of-band (passing the information by other means, e.g., SNMP messages). The goal is to better understand the behavior of the satellite link in order to accurately control the Slow Start and Fast Recovery mechanisms and improve the overall network efficiency. Refer to IET-Comm for more on cross-layer signaling methods.

TCP and Web Acceleration: TurboPage®

As discussed, when using a geostationary satellite network for Internet access one concern is the latency due to the delay caused by satellite path propagation time.

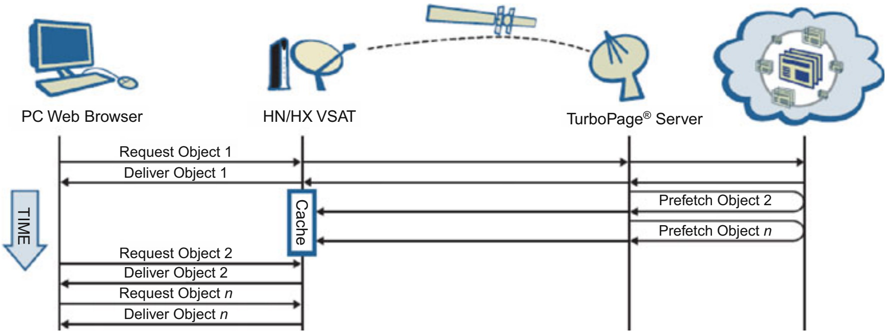


Fig. 22 HTTP TurboPage pre-fetch

Typical Web page layouts require multiple hops to various servers in order to fetch the required data and graphical objects. HTTP Pre-fetch, which is marketed by Hughes as TurboPage, is an application acceleration feature that enhances the performance of the HTTP (Web browser) protocol by prefetching the HTML (Hypertext Markup Language) objects embedded in a Web page into the remote terminal before the application requests them. This feature uses a hub traffic server to request HTTP objects in advance and on behalf of browsers at the remote terminal. The proxy server pushes these objects to the remote VSAT where it is cached and, therefore, immediately available when the client browser requests the object. This concept is illustrated in Fig. 22. This feature provides significantly better user experience as Web pages are "painted" more quickly.

HTTP Pre-fetch is primarily aimed at improving Web page response time rather than at improving bandwidth efficiency. However, it does contribute to the latter in two ways. First, because objects are prefetched, GET requests for the objects that do not need to be sent across the return channel, thus conserving bandwidth. Second, HTTP Pre-fetch applies V.44 compression to HTTP objects as they are forwarded, reducing forward channel bandwidth requirements.

IP Routers in Space

The challenge of reducing the latency and maximizing the transport efficiency has led to the consideration of placing an Internet router in space. The question remains if the IP router functionality supporting full IP satellite communication networks is best implemented in a hub on the ground or onboard the satellite.

Claimed benefits of a spaceborne router include increased security, increased network manageability, increased capacity, and decreased delay. The actual benefit of implementing an IP router on a satellite depends on the service and comparable satellite architecture being implemented. These benefits, which in most cases are marginal, do not come without a cost as the modems and routers consume satellite resources that have direct relationship to satellite capacity.

The potential benefits of spaceborne processing for Internet routing derive from two functional areas: resource allocation algorithms and beam-to-beam switching. These two functions may be separated (one in Space, one at the Gateway) or combined.

The use of onboard routing for beam-to-beam switching permits a user in one beam to communicate directly with a user in another beam in “one-hop” through the satellite. The hub spoke architectures discussed previously are “two-hop” links. A user in one beam, say beam A, sends a message to the satellite which relays that message to a “gateway” Earth station. The gateway station recognizes the message is intended for a user in beam B, retransmits the message up to the satellite, and the satellite downlinks the message to beam B. Providing beam-to-beam switching addresses two architectural points. It reduces the user-to-user delay and reduces the Gateway bandwidth requirements.

Although Performance-Enhancing Proxies (PEPs) have and continue to minimize the impact of delay on the network efficiency, onboard routing has a real and significant advantage to applications requiring low delay. In situations where satellite users communicate directly with other satellite users through beam-to-beam routing, the delay is reduced by 250 ms since the RTT to the Gateway is eliminated. Applications that are real-time interactive, such as fast-paced gaming, are most affected by the added 250 ms delay if beam-to-beam routing is not available.

Interactive voice communication also receives much discussion regarding delay. Current MSS satellite networks have accepted the two-hop for mobile voice communications; however, there is a perceptible difference from terrestrial service, since users generally start recognizing delay when the round trip delay exceeds 480 ms.

As an aside, there are methods of achieving beam-to-beam connectivity in one-hop through a satellite that does not require an Internet Router on the spacecraft. For example, one or more transponders on a C-band uplink beam can be connected on a satellite to a Ku-band downlink beam, with the corresponding Ku-band uplink beam transponders connected to the C-band downlink beam. Then, a user in beam A who wants to send a message to beam B simply transmits the message in the appropriate uplink frequency which is routed through the satellite transponder directly to beam B. Such a connection, although effective in applications such as a VSAT network, is fixed and represents a loss of usable capacity if traffic patterns do not match the fixed assignments.

However, a full onboard processing architecture makes the satellite literally a “switchboard in the sky.” The satellite traffic is formatted into packets and each uplink packet includes the customized satellite system “address” of the intended downlink beam. This system has the advantage that, even with many beams, any uplink beam can be connected to any downlink beam, and the system can easily respond to changing traffic patterns. However, the onboard digital processing necessary to correctly route every packet requires a large fraction of the satellite resources (mass and power), leaving less resources for satellite RF transmit power which results in less overall capacity for the satellite network.

Since capacity is directly related to the satellites radiated power, satellite designers tend to optimize their architectures to maximize the power dedicated to

EIRP. Any resources directed at processing remove resources from transmit power, thus having a negative effect on the total capacity.

In addition to the beam-to-beam interconnectivity advantage of an IP router, there is also a potential advantage to placing the resource allocation mechanism in space. A dynamic multiple access allocation algorithm reduces wasted bandwidth while maintaining service quality by allocating bandwidth to users only when data is available for transport and by leveling traffic by prioritizing access based on the application or user authenticated priority.

This process is generally performed in the Gateway Network Operations Center (NOC) today. However, two-hops are required for a user to establish a connection and receive their frequency and time slot assignment. Further, any changes to that allocation require another two-hops. If this function is placed in space, there will be less delay before the resource assignments can be made or changed. Consequently, more efficient use of the capacity is possible, by packing the available time, frequency, and/or code slots more fully. The smaller the data to transmit, the greater efficiency onboard allocation algorithms provide since, with large amounts of data, the time to perform the allocation is a small percentage of the overall connection time.

In November 2009, Intelsat 14 (I-14) was launched with a payload that includes a "demonstration" Internet router in space (IRIS) as shown in Fig. 23 (Cisco_Website), with C-band and Ku-band beam coverage. This payload permits IP data over a portion of the bandwidth in any of three beams. A single transponder of 36 MHz bandwidth for each of the three beams is connected to the router.

Experiments will continue with the I-14 payload to help assess the true benefits of implementing a full payload with Internet routing. Nevertheless, two major hurdles must be overcome: justifying the reduction in the total satellite capacity by directing

Fig. 23 Internet router in space launched on intelsat-14

limited satellite resources from transmit power to IP processing and having sufficient flexibility and robustness in the spaceborne router to keep up with technological advances and the evolution of traffic profiles.

Conclusion

There are many technical challenges in the integration of satellite and terrestrial networks. Both MSS and Broadband Satellites for Internet access provide service today through innovative solutions to many of these challenges. Nevertheless, enhancements to the current solutions, as well as new solutions will be needed to keep up with the rapid advances in the terrestrial marketplace.

The long life of satellites (in excess of 15 years), when compared to terrestrial products, will continue to demand integrated solutions which design the satellite node of the architecture to be as flexible, robust, and transparent as possible. While some limitations are unavoidable in the satellite node, such as the frequency band, orbit, and round trip time for Geo satellites, clever implementations in the satellite payload design and in the ground node such as PEPs and GBBF can provide the flexibility needed for satellite networks to continue to play an important role in the rapidly advancing network services and applications throughout the world. The design of ground systems for new large scale constellations in LEO orbits that are optimized for Internet services are addressed in chapter __. Innovations in electronic beam formation and steering for low cost ground terminals will be key to providing cost effective Internet services for new large scale constellations in LEO orbits.

Cross-References

- ▶ Broadband High-Throughput Satellites
- ▶ Distributed Internet-Optimized Services via Satellite Constellations
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Satellite Communications Video Markets: Dynamics and Trends
- ▶ Satellite Earth Station Antenna Systems and System Design
- ▶ Satellite Transmission, Reception, and Onboard Processing, Signaling, and Switching

References

I.F. Akyidldiz, G. Morabito, S. Palazzo, TCP-peach: a new congestion control scheme for satellite IP networks. IEEE/ACM Trans. Netw. **9**(3), (2001)

M. Allman, D. Glover, L. Sanchez, Enhancing TCP over satellites channels using standard mechanisms. RFC 2488 (1999)

M. Allman, W. Paxson, TCP congestion control. RFC 2581 (1999)

S. Brandner, The internet standards process. RFC 2026, Version 3 (1996)

C. Caini, R. Firrincieli, End-toEnd TCP enhancements performance on satellite links, in *Proceedings of the 11th IEEE Symposium on Computers and Communications* (2006a), pp. 1031–1036

C. Caini, R. Firrincieli, D. Lacamera, PEPsal: a performance enhancing proxy designed for tcp satellite connections, in *Vehicular Technology Conference*. VTC 2006-Spring, IEEE 63rd, vol. 6 (2006b), pp. 2607–2611

C. Caini, R. Firrincieli, Comparative performance evaluation of TCP variants on satellite environments, in *Proceedings of the IEEE International Conference on Communications* (2009), pp. 1–5

CCSDS, *Recommendations for Space Data System Standards, CCSDS 714.0-B-2. Blue Book*, Issue 2 (CCSDS, Washington, DC, 2006)

Cisco Website, http://www.cisco.com/web/strategy/government/space-routing.html

D.C. Cox, H.W. Arnold, R.P. Leck, Phase and amplitude dispersion for earth-satellite propagation in the 20 to 30 GHz frequency range. IEEE Trans. Antennas Propag. **AP-28**(3), 359–366 (1980)

L. Cui, S.J. Koh, X. Cui, Enhanced wireless TCP for satellite networks, in *International Conference on Wirelsss Communications, Networking and Mobile Computing* (2007), pp. 1813–1816

DOCSIS 3.0, MAC and upper layers protocol interface specification. CM-SP-MULPIv3.0-I14-101008 (2010)

DOCSIS 3.0, Physical layer specification, CM-SP-PHYv3.0-I09-101008 (2010)

E. Dubois, et al., Enhancing TCP based communications in mobile satellite scenarios: TCP PEPs issues and solutions, in *Proceedings of the Fifth Advanced Satellite Multimedia Systems Conference and the 11th Signal Processing for Space Communications Workshop* (2010), pp. 476–483

W. Dusi, W. John, Estimating routing symmetry on single links by passive flow measurements, in *(Draft Presentation for) the 1st International Workshop on TRaffic Analysis and Classification (TRAC) colocated with the 6th International Wireless Communications & Mobile Computing Conference (IWCMC 2010)* (2010)

ETSI EN 302 307 V1.2.1, DVB-S2 (2009)

ETSI TR 101 984 V1.2.1 (2007–2012): Satellite Earth Stations and Systems; Broadband Satellite Multimedia; Services and architectures

ETSI TR 102 676 SES, BSM, PEPs. V1.1.1 (2009)

ETSI TS 102 354 V1.2.1 (2006–2011): Satellite Earth Stations and Systems; Broadband Satellite Multimedia; Transparent Satellite Star; IP over Satellite Air Interface Specification

Fall 2010 Global internet phenomena report. Sandvine Corporation, (2010), www.**sandvine**.com/amp;/**2010**%20Global%20Internet%20Phenomena%20Report.pdf

FCC, Notice of proposed rulemaking (In the Matter of Flexibility for Delivery of Communications by Mobile Satellite Service Providers in the 2 GHz Band, the L-Band, and the 1.6/2.4 GHz Band), FCC 01–225, IB Docket No. 01–185 (2001)

FCC, Memorandum Opinion and Order and Second Order on Reconsideration, FCC 05–30, IB Docket No. 01–185 (2005)

S. Floyd, T. Henderson, The NewReno modification to TCP's fast recovery algorithm. RFC 2582 (1999)

J.D. Gibson, *The Communications Handbook* (CRC Press, Boca Raton, 1997)

Y.F. Hu, M. Berioli et al., Broadband satellite multimedia. Inst. Eng. Technol. Commun. **4**(13), 1519–1531 (2010)

Hughes Network Systems, IP over Satellite (IPoS) – The Standard for Broadband over Satellite (2007)

L.J. Ippolito, Propagation effects handbook for satellite systems design. NASA Doc. 1082 (1989)

ISO 13537:2010 Space data and information transfer systems – reference architecture for space data systems (2010)

ITSI EN 301 790 V1.5.1, DVB; Interaction Channel for Satellite Distribution Systems (2009)

R.C. Johnson, H. Jasik, *Antenna Engineering Handbook*, 2nd edn. (McGraw Hill, New York, 1984)

L. Joing, C. Zhigng, M.J. Khan, TP-Satellite: a new transport protocol for satellite IP networks. IEEE Trans. Aerosp. Electron. Syst. **45**(2), 502–515 (2009)

Lindberg-Walker, The emulation of satellite communication systems in load-stressing conditions, in *The Proceeding of the 2003 Conference of the American Institute of Aeronautics and Astronautics* (Monterey, 2003)

M. Luglio, M.Y. Sanadidi, M. Gerla, J. Stepanek, On-board satellite "split TCP" proxy. IEEE J. Sel. Areas Commun. **22**(2), 362–370 (2004)

M. Mathis, et al., TCP selective acknowledge options. RFC 2018 (1996)

Network Simulator (NS-2) (Online), http://www.isi.edu/nsnam/ns/. Accessed Nov 2010

O3B Networks, http://www.o3bnetworks.com. Accessed Nov 2010

E. Rendon-Morales, et al., Cross-layer architecture for TCP splitting in the return channel over satellite networks, in *Sixth International Symposium on Wireless Communications Systems* (2009), pp. 225–229

C. Roseti, M. Luglio, F. Zampognaro, Analysis and performance evaluation of a burst-based TCP for satellite DVB RCS links. IEEE/ACM Trans. Netw. **18**(3), 911–921 (2010)

I-PEP Specifications, Satlabs Group Recommendations, Issue 1a (2005), www.satlabs.org. Specific profile of SCPS-TP

W.R. Stevens, *TCP/IP Illustrated*, vol. 1 (Addison-Wesley, Reading, 1994)

Terrestar Networks, http://www.terrestar.com. Accessed Nov 2010

Transmission Control Protocol, DARPA Internet Program Specification, RFC 793 (1981)

J.L. Walker, B. Day, S. Xie, Architecture, implementation and performance of ground-based beam forming in the DBSD G1 mobile satellite system, in *AIAA 28th ICSSC Conference Proceedings* (2010)

Satellite Communications: Regulatory, Legal, and Trade Issues

Gérardine Goh Escolar

Contents

Note: All opinions and any errors in the text remain entirely those of the author and do not engage the United Nations or any other organizations with which the author is affiliated.

G. Goh Escolar (✉)
Lauterpacht Centre for International Law, University of Cambridge, Cambridge, UK

Iran-United States Claims Tribunal, The Hague, The Netherlands
e-mail: G.GohEscolar@uclmail.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_23

Abstract

This chapter focuses on the regulatory, legal, and trade issues related to satellite communications. The chapter first examines the regulatory and legal issues on three levels: the global or international arena, regional regulatory institutions, and national regulatory frameworks. Next, the chapter focuses on a discussion on the role of satellite communications in global trade and will review the regulatory, legal, and trade issues of satellite communications on the global scale. The last part will discuss issues related to the settlement of disputes that may arise from satellite communications, as well as the legal principles of responsibility and liability for any damage caused by satellite communications. Note that this chapter does not deal with the legal and other issues related to the International Telecommunications Union and radio frequencies, which have been dealt with in previous chapters of this book.

Keywords

Arabsat Asia-Pacific Satellite Communications Council (APSCC) • Association of Southeast Asian Nations (ASEAN) • Dispute settlement • EUMETSAT • European Space Agency • EUTELSAT • General Agreement on Trade and Tariffs (GATT) • General Agreement on Trade in Services (GATS) • INMARSAT • Intellectual property • INTELSAT • International law • INTERSPUTNIK • Legal issues Licensing • NAFTA • National legislation Outer Space Treaty of 1967 • RASCOM • Regional organizations Regulatory framework United Nations • Committee on Peaceful Uses of Outer Space (COPUOS) • World Intellectual Property Organization (WIPO) • World Trade Organization (WTO)

Introduction

The satellite communications industry is a complex multibillion dollar global industry that involves a diverse amount of regulatory, legal, and trade issues. This is because the technology involves international and national allocations of frequencies and safety and environmental provisions. The patterns of international regulation of satellite communications has shifted greatly in past decades with a movement toward privatization of satellite communication systems and the use of regulated competitive systems to bolster the range of service offerings and to reduce the cost of service. The pattern of this regulatory shift is reported in some detail.

Satellite communications is also an international undertaking that involves international treaty and regulatory provisions relating to strategic and defense-related consideration, the discrimination between air space and other space, and, in particular, to the activities of various United Nations entities and interpretation of a number of space-related treaties in the international court system. Many of the subjects discussed in this chapter involve complicated legal interpretations of international law, treaties, and conventions. Rather than seeking to address each of these areas in great depth, the approach has been to provide an overview of the issue and then indicate by reference where even more fine-grained information is available.

International Regulation of Satellite Communications

As a component of international space-based telecommunications, satellite communications are bound by the rules of international law generally[1] and international space law specifically (Salin 2000). Crucially, satellite communications are not primarily governed by the principles of international air law. Despite the academic debate on the delimitation on airspace and outer space[2], satellites clearly fall into the category of "space objects." As such, international space law applies.

Further, regulations and standards pertaining to satellite communications have been promulgated by international standard-setting organizations. These standards, which are derived from a recognized technical competence, deal with the nuts-and-bolts of launching, maintaining, and operating a satellite communication system. Further nuances of the international regulation of satellite communications are added by the activities of international satellite operators and multinational satellite consortia.

Satellite Communications in the International Legal System

The 1967 Outer Space Treaty incorporates many basic principles of international law and also includes regulations specific to activities in the outer space environment.[3] These include the following principles:

- States are obliged to act for the benefit and in the interests of all countries in conducting their space activities.
- Outer space shall be the province of all mankind.
- Outer space shall be free for exploration and use by all States.

---

[1] Article III, *Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies* entered into force 10 October 1967. 610 U.N.T.S. 205 (1967) [hereinafter "Outer Space Treaty"]. See text below.

[2] Vast literature has been devoted to the subject of the delimitation of airspace and outer space. See among others Dodge (2009), Gorove (1997, 2000).

[3] It is beyond the scope of this chapter to provide an elaboration that would do justice to this field of the law. For an excellent overview, see Lyall and Larsen (2009). A legal commentary on the Outer Space Treaty can be found at Hobe et al. (2009).

- Outer space shall not be subject to national appropriation by claim of sovereignty, use or occupation, or any other means.
- International law shall apply to activities in outer space.
- The placement of nuclear weapons or other weapons of mass destruction in orbit, installation of such weapons on celestial bodies, or the stationing of such weapons in outer space is prohibited.
- States bear international responsibility for national activities in outer space, whether such activities are carried on by governmental agencies or by nongovernmental entities.
- Activities of nongovernmental entities shall require authorization and continuing supervision by the State.
- A launching State is internationally liable for any damage caused by the space object it launched or procured the launch of.
- A State has jurisdiction and control over space objects that it has registered.
- States are to cooperate and mutually assist other States with activities in outer space, and shall conduct all space activities with due regard to the interests of other States.
- States are to avoid harmful contamination and adverse changes in the environment of the Earth in the conduct of their space activities, and must consult with other States where harmful interference may occur.

The Outer Space Treaty is supplemented by several other international treaties applicable to satellite communications, including the following:

- 1971 INTELSAT Agreement
- 1971 INTERSPUTNIK Agreement
- 1972 Liability Convention
- 1974 Brussels Convention
- 1975 Registration Convention
- 1975 European Space Agency Convention
- 1976 INMARSAT Agreement
- 1976 Intercosmos Agreement
- 1976 Arabsat Agreement
- 1982 EUTELSAT Convention
- 1983 EUMETSAT Convention
- 1994 International Telecommunications Union Convention

It must be noted that much of the basic framework of public international space law was created through the organs of the United Nations, and in particular through the work of the Committee on the Peaceful Uses of Outer Space (COPUOS).[4] Aside from the negotiation and drafting of the treaty framework applicable for activities in

---

[4] See COPUOS (2015). The website also provides a directory with links to all the relevant international documents, treaties, agreements and declarations.

outer space, work originating from the COPUOS has also led to the adoption of various relevant United Nations General Assembly (UNGA) Resolutions. Although such UNGA Resolutions are not binding as international law, they are indicative of State practice and *opinio juris*. As such, some principles framed in the UNGA Resolutions relating to activities in outer space are considered to have become customary international law. In the context of satellite communications, the two UNGA Resolutions that are of particular significance are as follows:

- The Declaration of Principles Governing the Use by States of Artificial Earth Satellites for International Direct Television Broadcasting
- The Declaration on International Cooperation in the Exploration and Use of Outer Space for the Benefit and in the Interest of all States, taking into account the Needs of Developing Countries

Certainly, customary international law dealing with activities in outer space,[5] as well as other soft law provisions, also applies to satellite communications.

It has been lamented by highly qualified publicists that the rate of lawmaking at the United Nations in relation to activities in outer space has considerably decreased in recent years. However, it is interesting to note that this deceleration has not affected issues related to space-based communications. Instead, there is a burgeoning increase in the development of technical standards and codes of conducts in recent years that are directly related to satellite communications (Jasentuliyana 1995).

Of particular interest is the interaction of public international law with the evolution of two characteristics of modern satellite telecommunications: commercialization and liberalization.[6] Certain principles inherited from the public international law roots of the law relating to satellite communications will necessarily have bearing on the more recent developments in the field. In particular, the principle that activities in outer space should be undertaken for the benefit and in the interests of all countries, irrespective of the fact that million- or billion-dollar investments may have been made by private commercial entities, appears to be untenable (Christol 1991).

Other legal principles jostle for attention in the increasingly liberal and commercial marketplace for satellite communications. Among these are the international responsibility and liability imposed on nation-states for space activities,[7] restrictions on military activities in outer space,[8] and principle of nonappropriation.[9] Certainly,

---

[5] On the topic of customary international law, see the judgment of the International Court of Justice in the *North Sea Continental Shelf Cases*, (Federal Republic of Germany vs. Denmark, Federal Republic of Germany vs. the Netherlands), Judgment, (1951) ICJ Reports 116.

[6] See for example Matte (1984).

[7] See, among others, Cheng (1998), von der Dunk (1991), Kerrest (1997), Jasentuliyana and Lee (1979).

[8] Haeck (1996). See also Cheng (2000), Goh (2004), Tannenwald (2004).

[9] See among others, Gorove (1969), Williams (1970).

as some highly qualified publicists have pointed out, there is “some kind of naïveté” in the idea that commercial entities that have invested vast sums in the construction, infrastructure, launch, operation, and maintenance of a satellite communication system would be amenable to sharing the benefits of this system with those who have invested nothing.

International Standards

Technical and policy standards impacting satellite communications are also established by other international and professional organizations. The harmonization of technical and other protocols has necessitated the establishment of clear standards to ensure interoperability. Professional and technical organizations such as the International Standards Organization (ISO) and the International Electrotechnical Commission (IEC) have established standards approved and in use by international, national, and private organizations dealing with satellite communications.

Standards have also been set by some regional and national organizations. These include the European Telecommunication Standards Institute (ETSI), the Standards Committee T1 Telecommunications (North America), and the Japanese Telecommunications Technology Committee. To avoid any possible overlaps or contradictions, the International Telecommunications Union has ensured interoperability and cooperation through the Interregional Telecommunication Standards Conferences (ITSC), which held its inaugural meeting in 1990.

International Satellite Operators: Regulatory and Legal Issues

What is readily apparent from the structure and framework of these international satellite operators (and their regional siblings discussed in the next section) is the cross-cutting nature of their operations. These organizations straddle the line between public regulatory agency and private commercial enterprise. While regulating satellite communications in line with their respective constituent agreements, many of these organizations are also satellite owners and operators. This leads to the interesting paradigm of the dual-nature public/private entity that promulgates the policies and regulations its commercial activities are then bound by.

INTELSAT

INTELSAT was established in August 1964 with the signature by 11 Member States of an Interim Agreement and a Special Agreement. A Multilateral Agreement and an Operating Agreement signed in 1971, followed by a Headquarters Agreement signed in 1976, set up its organization. Initially an international intergovernmental government, its membership complement consisted of States that were members of the ITU. Each Member State had one vote and subscribed to the capital of INTELSAT in proportion to its use of the INTELSAT space segment. As such, it combined the characteristics of a public intergovernmental organization and those of a private commercial entity. The case of INTELSAT is particularly noteworthy as a sterling example of trends toward globalization and commercialization.

INTELSAT was composed of its Assembly of Parties, the Meeting of Signatories, the Board of Governors, and the Executive Organ. Member States are represented by the national telecommunications agencies. Historically, only Member States had access to the INTELSAT system, although they were also free to have their own satellites, as long as these were compatible with the INTELSAT system. They were also free to resell capacity on the INTELSAT system to end users. Nonmember governments may become "Duly Authorised Telecommunications Entities" (DATEs) and contract with INTELSAT for the use of the system. Direct Access Customers, which are not Member States and which may also be investors in INTELSAT, came onto the scene slightly later.

INTELSAT exercised some regulatory control over its own, as well as other separate, satellite systems. INTELSAT's mission is to ensure, on a commercial basis, the provision of the space segment necessary for international public telecommunications services of quality and reliability. As a result of Articles III and XIV of the Multilateral Agreement, it mandates compulsory coordination for separate satellite systems. Where Member States use INTELSAT's services for the provision of domestic specialized telecommunications services, these must be coordinated with the INTELSAT system. It will be noted that Article XIV of the Multilateral Agreement in particular specifies the rights and obligations of Member States, including compliance with the principles of INTELSAT. Pursuant to Article XIV, this means that, for its domestic public telecommunications services, Member States must consult with the Board of Governors for recommendations, and for its international public telecommunications services, Member States must consult with the Assembly of Parties. This is to ensure technical compatibility of their facilities and operations with the INTELSAT space segment.

In the 1990s, it became clear that INTELSAT was successful in establishing a "global commercial telecommunications satellite system and Organisation ... to provide expanded telecommunication services on a non-discriminatory basis to all areas of the world." As a result of the changes in the economics of international telecommunications, the deregulation of the space market, and the intent of INTELSAT to expand its activities and business concerns, INTELSAT was restructured and privatized on July 18, 2001, becoming Intelsat Ltd., and is now internationally headquartered in Luxembourg. It was sold in January 2005 to four private equity firms, and then subsequently resold to other equity firm holdings, and also it merged with and took over the Panamsat system. Intelsat continues to generate revenue through satellite usage fees. A small International Telecommunications Satellite Organization was also established as of 2001 to address international concerns related to satellite services as safety and the provision of satellite services to developing countries.

INMARSAT

INMARSAT was founded in 1979 by the Inter-Governmental Maritime Consultative Organization. The INMARSAT Convention and Operating Agreements of 1976 were originally aimed at the provision of transponders for ship and maritime communications via satellite. In 1982, INMARSAT became an international maritime satellite organization. Under the name INMARSAT-II, it later became a global maritime satellite network system in 1990, providing mobile communications for users at sea, on the ground, and in the air.

The original INMARSAT Convention binds States, while its Operating Agreement had States and other entities as members. It combined, as did INTELSAT above, the characteristics of public service and private commercial activity. The Preamble of its Convention refers to Article I of the Outer Space Treaty. INMARSAT comprised an Assembly of Parties, a Council, and a Directorate. The Assembly of Parties was composed of all Parties, which made recommendations to the Council in regard to "the activities, purposes, general policy and long-term objectives of the Organisation." This Assembly included States as well as non-State entities. The INMARSAT Council comprised 22 members, representing its 18 largest contributing members and 4 geographical members elected by the Assembly. It managed the INMARSAT space segment. The Directorate, which reported to the Council, administered the Organization.

It must be noted that INMARSAT took a different approach than INTELSAT in allowing parties to contract as between themselves, while using INMARSAT facilities. As such, regulatory policies promulgated by INMARSAT may or may not have had an impact on inter-Party (extraorganization) dealings. Significantly, however, INMARSAT had quasipublic authorities, such as the authority to compel private competitors to provide early warning of any projected activities that would be in competition with INMARSAT.

With the pressure of increasing commercialization and competition, INMARSAT was privatized in 1999. It was thus the first of the international satellite treaty organizations to make this transition. INMARSAT was divided into a commercial entity, Inmarsat plc (headquartered in London), and a regulatory body, the International Mobile Satellite Organisation (IMSO). IMSO continues to provide public regulatory services such as the administration of the Global Maritime Distress Safety System (as established by the International Maritime Organization) and the public service component established through the Standards and Recommended Practices established by the International Civil Aviation Organization (ICAO).

INTERSPUTNIK

Intersputnik International Organization of Space Communications was established on November 15, 1971, in accordance with the Intergovernmental Agreement on the Establishment of the Intersputnik International System and Organization of Space Communications. It is an international organization, with its seat in Moscow, and vested with the right to execute contracts, acquire, lease, and alienate property, and to institute proceedings. It develops and maintains contacts and cooperation with other global, regional, and commercial satellite organizations. Intersputnik was registered with the United Nations on March 27, 1973. It has the status of permanent observer at the UN Committee on the Peaceful Uses of Outer Space, the ITU, and UNESCO.

It is a member of the Asia-Pacific Satellite Communications Council and the Global VSAT Forum.

Today, Intersputnik has 25 Member States, including countries in Latin America, the Arab Peninsula, and Asia. Its core business is the lease of satellite capacity to telecommunications operators, broadcasters, and other clients, as well as services for the establishment, engineering, and operation of satellite networks. The organization procures and deploys spacecraft in orbit. It is also an international organization that participates in international and intergovernmental forums. The Board and Operations Committee work in tandem with the Directorate to cover issues such as policy-planning, technical standards promulgation, determination of the amount of share capital, compliance with its international obligations, and establishment of its financial and commercial policies. Through its fully owned subsidiary Intersputnik Holdings, it also conducts commercial satellite communications activities through three daughter companies: Isatel Russia, Isatel Kyrgyzstan, and Isatel Tajikistan.

Regional Regulation of Satellite Communications

Aside from the international satellite operators mentioned above, there are several regional organizations using and regulating space-based satellite communications. These include

1. The Arab Satellite Communications Organization (Arabsat)
2. The Asia-Pacific Satellite Communications Council (APSCC)
3. The Association of Southeast Asian Nations (ASEAN)
4. Various European institutions such as EUTELSAT, EUMETSAT, and the European Space Agency (ESA)
5. The grouping of the North Atlantic Free Trade Agreement (NAFTA)
6. The Regional African Satellite Communications Organisation (RASCOM)

Arab Satellite Communications Organization (Arabsat)

Arabsat was founded by the 21-member Arab League in 1976. It is an intergovernmental organization with a paid capital of USD $500 million, paid up in various proportions by its Member States. Its establishment in 1976 initially aimed to design, execute, and operate the first Arab space system for satellite communications. The first of its first-generation satellites was launched in 1985. Today, Arabsat is manufacturing, launching, and operating its third-generation satellite network.

Arabsat carries more than 400 television channels and 160 radio stations, reaching an audience of more than 164 million in the Arab Member States alone. It offers broadcast, telecommunications, and broadband services, including data network solutions, telephony, and Internet provider trunking backbone connectivity, as well as broadband Internet access.

Arabsat comprises a General Assembly of Member States, a Board of Directors, and a Management Committee. The General Assembly negotiates and approves satellite communications policies for the region and the Arabsat network. The Board of Directors holds periodic meetings to secure, invest in, and maintain the satellite assets of the Arabsat Organization, as well as to implement policies passed by the General Assembly. The Management Committee is responsible for the day-to-day operations of the Arabsat network.

Association of Southeast Asian Nations (ASEAN)

The Association of Southeast Asian Nations (ASEAN) is a ten-member grouping of countries founded in 1967. Its declared aim is to "accelerate the economic growth, social progress and cultural development in the region through joint endeavours" and to "collaborate more effectively for . . . the expansion of their trade, including the study of the problems of international commodity trade, the improvement of their transportation and communications facilities and the raising of the living standards of their peoples."

In the context of satellite communications, the Report of the meetings of the Sub-committee on Posts and Telecommunications is of pertinence. This Sub-committee has purview over the Integrated Work Programme in Posts and Telecommunications, including domestic and regional satellite communications and radio frequency coordination in ASEAN. The ASEAN Radio Frequency Coordination Committee has also been established to coordinate the use, the administration, and the management and planning of radio frequency in the ASEAN region, especially in the border areas. The Secretariat of the Committee is situated in Jakarta. The Sub-committee proposes broad policies that are considered by the ASEAN Member States and that, if and when approved, are implemented by the Secretariat.

Asia-Pacific Satellite Communications Council (APSCC)

The Asia-Pacific Satellite Communications Council (APSCC) grew out of a proposal by the government of the Republic of Korea, presented at the United Nations Workshop on Space Communications for Development in the Asia-Pacific held in November 1992. The APSCC Constitution was adopted in 1994. Today, APSCC represents all sectors of satellite and space-based industries in the region under the purview of the United Nations Economic and Social Commission for Asia and the Pacific (ESCAP). It aims to promote satellite communications and broadcasting for the socioeconomic and cultural development of the ESCAP region.

APSCC provides the forum at which its members may exchange views on policies, technologies, systems, and services in relation to satellite communications. It counts satellite manufacturers and operators, launch service providers, satellite risk management companies, telecommunication carriers, and broadcasters among its members. It aids in the formulation of recommendations on policies, regulations, and technical standards within the ESCAP region and in relation to the global satellite communications arena. APSCC also works with Member States to minimize technical and regulatory barriers to the deregulation of the satellite communication industries in ESCAP. To that end, it cooperates with other international institutions such as the United Nations Office for Outer Space Affairs, ESCAP, the ITU, and the Asia-Pacific Economic Cooperation (APEC).

European Regulatory Authorities

The particular case of Europe merits special attention. The central regulation exercised by the European Union, together with its structures and institutions, cannot be examined to the detail they deserve within the scope of this chapter. Reference instead should be made to the plethora of learned writings on the topic.[10] This section will only look at the institutions that have a direct impact on the legal and regulatory framework of Europe on satellite communications.

The European Legal Framework on Satellite Communications

The European Commission (EC) has been very active in promulgating Community law pertaining to telecommunications, space activity, and satellite communications. The various EC Directives, Resolutions, Decisions, Recommendations, White Papers, Green Papers, Guidelines, and Communications issued since 1986 go a long way to show the EC policy on satellite communications.

In particular, the EC has focused on regulating the following issues:

- The harmonization of domestic regulations on satellite communications
- The liberalization of the Earth segment and related market sectors
- The competition between satellite operators
- The application of EC law and the development of the single European internal market

Further, the EC has also sought to ensure the separation of regulatory and operation functions in Member States, granting free and unrestricted access to the space segment capacity, while ensuring commercial freedom for space segment providers. It is significant to note that the European authorities may intervene in cases of restrictive or anticompetitive business practices, including corporate strategy development and monopolistic actions.

The harmonization of domestic regulations on satellite communications was established by the Telecommunications and the Satellite Green Papers first issued in 1987 and 1990. Measures undertaken include framework Council Directors on

---

[10] See among others, Hartley (2010), Kaczorowska (2010), Chalmers et al. (2010), Dashwood et al. (2012), Bishop (2009).

- *The Open Network Provision (ONP)*: All ONP conditions must be objective, transparent, and published, and should guarantee equal access under nondiscriminatory principles of Community law. Access to public telecommunications networks or services cannot be restricted except for reasons of essential requirements, and such restrictions must comply with EC law.
- *The Protection of the Satellite Communications*: This includes the protection of all forms and content of satellite communications, including copyright, the protection of databases and encrypted services, as well as the protection of neighboring rights. This established the author's exclusive right to authorize the diffusion of his/her work via satellite, and elaborates upon the applicable EC law relating to trans-boundary retransmission.
- *The Standardization of Satellite Communications Equipment*: This ensured harmonized procedures for certification, testing, quality assurance, and product monitoring for satellite communications equipment throughout the Community. It also harmonized domestic laws, regulations, and standards pertaining to safety, compatibility, interoperability, and interworking of telecommunications network equipment and terminal equipment.
- *The Reciprocal License Recognition for Satellite Services*: This ensured that licenses for satellite services issued by one Member State would be recognized in other Member States of the Community, replacing the system by which a satellite service provider would have to apply for different licenses in different domestic legal systems.

These measures were aimed at allowing the development of the European internal market, lifting trade and technical barriers across borders, while ensuring the freedom of access, use, and provision of satellite communications services. They were also aimed at the provision of a competitive environment for the progress of the European satellite communications industry as a whole.

The liberalization of telecommunications and satellite communications in the Community was aimed primarily at the following:

- *Separation of regulatory and operational functions*: This was meant to correct the monopoly that most national telecommunications operators had in the domestic market of Member States, and to ensure free market access for private operators.
- *Freedom of competition of space segment capacity providers*: This objective of this measure was to ensure open and free competition among providers of the space segment equipment and services.
- *Liberalization of infrastructure*: This intended to ensure a common approach to the telecommunications infrastructure in the Community. Liberalization was provided for so as to ensure competition not only in the equipment and services sector, but also in infrastructure provision and operation. Further, the application of the EC's competition rules allowed for fairness and transparency.
- *Access to the space segment*: This measure aimed to ensure direct access throughout all Member States, ensuring the joint reform of international satellite organizations and joint management of the space segment as an essential resource of the

Community and the European Union. It also aimed at providing a competitive environment for European industry players.

Regulation as to competition in the satellite communications industry was also given special attention by the Commission. In particular, action has been undertaken in regard of restrictive business practices, monopolies, dumping practices, and anticompetitive practices. This is achieved through the review of strategic alliances, service arrangements, as well as mergers and acquisitions of satellite communications entities. The domestic application of EC law in the context of the internal market also undergoes periodic review by the Commission to ensure full compliance with European standards.

EUTELSAT

The European Telecommunications Satellite Organization (EUTELSAT) was established in 1982, upon the opening of its Convention at the Paris Convention. Prior to the entry into force of the Convention, it operated under interim agreements entered into on a provisional basis. The establishment of EUTELSAT came as a regional impetus from European States to increase European participation in space and satellite-related activities and also as a European response to INTELSAT. Sixteen Member States of the European Conference of Postal and Telecommunication Administrations came together to establish EUTELSAT, which aimed to become a global leader in fixed satellite services.

EUTELSAT consisted of an Assembly of Parties, a Board of Signatories, and an executive organ. Its structure is similar to that of INTELSAT's, with the Board of Signatories in this case undertaking the responsibilities of INTELSAT's Board of Directors. Additionally, EUTELSAT's Board of Signatories was charged with ensuring compliance with the ITU Radio Regulations rules. Procurement was by international invitation to tender, with consideration given to the Parties and Signatories. EUTELSAT's practice in regard of coordination differed from that of INTELSAT. EUTELSAT recognized separate systems – although acknowledging that coordination was essential so as to ensure technical compatibility and interoperability, EUTELSAT recognized the separation between systems launched by its Member States and those launched by private European entities.

With the liberalization of the European satellite communications industry, the public operations and activities of EUTELSAT were transferred to a private company, Eutelsat S.A., in July 2001. In April 2005, Eutelsat Communications, the holding company of Eutelsat S.A., was established. Eutelsat continues to serve the space, telecommunications, and audiovisual industries of Europe, the Middle East, Africa, and some parts of Asia and the Americas.

EUMETSAT

The European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) is an intergovernmental organization established in 1983. At the time of writing, it had 26 European Member States and 5 Cooperating States. EUMETSAT's mission is to establish, maintain, and exploit European systems of operational meteorological satellites. Member States make mandatory contributions to the organization, which are proportional to their gross national income.

EUMETSAT works closely with the European Meteorological Infrastructure (EMI) to respond to the national meteorological needs of its Member States. It ensures the European contribution to the space-based component of the World Meteorological Organization (WMO), while continuing to establish bilateral and multilateral agreements from non-EUMETSAT countries and projects. It also conducts environmental monitoring and disaster management activities.

EUMETSAT comprises a Council, an Executive body, and a Director. The Council is composed of two representatives for each Member State. EUMETSAT's data policy has been harmonized to ensure equal and nondiscriminatory access for both paying customers and public authorities, in line with the European legislation.

The European Space Agency (ESA)

The European Space Agency (ESA) was established in 1975 and is an intergovernmental organization with 18 Member States. It aims to promote peaceful cooperation among European States in space research and technology, by elaborating a European space policy and harmonizing the national space policies of Member States. Headquartered in Paris, ESA coordinates the European space program and integrates national space programs of Member States especially in regard of the development of applications satellites. It also elaborates and implements a coherent industrial policy in regard of space activities of its Member States.

ESA is not an agency or institution of the European Union (EU). Non-EU countries such as Switzerland and Norway are members of ESA. However, it must be noted that there has been work on defining the legal status of ESA in regard of the EU. The EU and ESA cooperate on some satellite projects such as the Galileo satellite navigation system. The 2007 European Space Policy commits the EU, ESA, and their respective Member States to increase coordination in their space activities. Further integration of ESA into the framework of the EU may mean greater coordination and harmonization in the policies, regulations, and legislations relating to satellite communications.

It must be noted that ESA also comprises a Directorate of Telecommunications and Integrated Applications. ESA activities in regard of satellite communications aim to

1. Develop advanced technologies for both the space and ground segment
2. Develop and implement full systems for new capacities
3. Implement new missions in partnership with industry
4. Develop and implement new applications for satellite communications

In order to achieve these goals, the Directorate has executed a programmatic framework known as the Advanced Research on Telecommunication Satellite Systems (ARTES). Commercial entities within ESA Member States may submit proposals to the ARTES program, which provides for strategic analysis, funding, technical development, product demonstration, and deployment of the proposed satellite systems.

NAFTA

The North American Free Trade Agreement (NAFTA)[11] was signed by Canada, Mexico, and the United States on December 17, 1992. Upon receipt into the national legal system of the signatory States and the fulfillment of all necessary procedures, NAFTA entered into force on January 1, 1994. This section will only address chapters of NAFTA that pertain to satellite communications:

- Chapter 12 on Cross-Border Trade in Services
- Chapter 13 on Telecommunications
- Chapter 15 on Competition Policy, Monopolies, and State Enterprises
- Chapter 17 on Intellectual Property

Chapter 12 deals with cross-border trade in services, which refers to any kind of service in all their various phases. It does not apply to financial services, public procurement, public subsidies or grants, and air services. Chapter 12 stipulates that the following principles apply to satellite communications:

- National treatment of nationals of the other two parties
- Most Favored Nation treatment of nationals of the other two parties
- Best standard of treatment among the above

Significantly, licensing, certification, and legal consulting requirements must be coordinated between the parties.

Chapter 13 applies in particular to the transfer of data, electronic exchange of information, and the maintenance of intracorporate networks. For each of these categories, basic services are distinguished from value-added services. The provisions of Chapter 13 provide for open access and nondiscrimination for public networks. It also provides for common regulations on technical standards that operators must conform to. Monopolies are allowed insofar as these do not engage in anticompetitive actions. Generally, the chapter provides for harmonization to ensure interoperability and cooperation, as well as transparency to ensure competitive fairness. Chapter 13 does not apply to the regulation of direct broadcasting services and nonpublic network-connected telephony satellite services.

Chapter 15 stipulates that monopolies may not engage in anticompetitive conduct. While it does not prevent the designation of a monopoly or a State enterprise, prior notification must be given to the other two parties, and the monopoly in question may not behave in a discriminatory or anticompetitive manner.

Chapter 17 on intellectual property reflects the agreements achieved in the World Trade Organization as discussed below. In particular, it provides that unauthorized decoded distribution of program-carrying satellite signals is a violation of NAFTA. Further, copyright protection is afforded for computer programs and data

---

[11] For more information on NAFTA, see http://www.naftanow.org. Accessed 13 October 2015.

compilations. Parties are allowed to grant even more extensive protection, as far as these are not inconsistent with NAFTA provisions.

It should be noted that Part 8 of NAFTA provides for certain exceptions to the obligations outlined briefly above. Among these exceptions are national security and certain exceptions based on cultural industry. It should also be noted that NAFTA includes a set of reservations and exceptions, which may also impact upon satellite communications.

Regional African Satellite Communications Organisation (RASCOM)

The Regional African Satellite Communications Organisation (RASCOM) was founded on the growing awareness of the importance of telecommunications for economic development and productivity. Following several international consultative meetings, African leaders decided to pool their efforts so as to provide the continent with a satellite telecommunications infrastructure. A feasibility study conducted in some 50 African countries from 1987 to 1990 found that continental satellite telecommunications was the best system by which to meet Africa's telecommunications needs. The findings of this study were adopted by the African States in February 1991 in Abuja. In May 1992, the African States met in Abidjan and established the RASCOM.

RASCOM is an international intergovernmental organization run with commercial capital. Its mission is to design, implement, operate, and maintain the space segment of the African telecommunications satellite system. It aims to further African integration through the appropriate policies and technologies so as to provide an affordable infrastructure on a large scale to both urban and rural areas of the continent.

In order to achieve its objectives, RASCOM relies on three basic organs: The Assembly of Parties, the Board of Directors, and the Executive Organ. The Assembly of Parties is made up of the 44 African signatory governments to the RASCOM Convention. It negotiates and promulgates strategic and policy-oriented regulations and meets in ordinary sessions biannually. The Board of Directors has responsibility for the design, development, operation, and maintenance of RASCOM's space segments, as well as other activities that the Assembly of Parties authorizes it to undertake. It comprises representatives of the signatory governments and representatives of nonsignatory shareholders. The Executive Organ of RASCOM implements the decisions of the Board and is responsible for the daily operations of RASCOM.

National Regulation of Satellite Communications

International law is implemented through the receipt of its obligations into the domestic legal system. Further, State practice through national acts or the passage of national legislation plays an important role in the development of customary international law relating to satellite communications. It is therefore important to remember that the national regulation of satellite communications must be in line with the State's relevant international legal obligations, and yet States' national legislation may shape the future landscape of the international regulation of satellite communications.

A point of significance is that international law is not self-executing in every State. The dualist legal systems in countries such as the United Kingdom and Canada require the adoption of international law through the passage of national legislation. These systems are different than those of countries such as France and the United States, where the international law is self-executing due to the monist legal systems in those countries.[12]

It is beyond the scope of this short chapter to discuss individual pieces of national domestic legislation that deals with satellite communications. In view of the increasing commercialization of satellite communications, there is an increasing need at the national level to implement international standards through domestic legal regulations that provide for corporate responsibility, technical qualifications and standards, ethical practices, liability insurance, and licensing regimes. Some States have chosen to enact the national space legislation,[13] which inevitably impacts upon satellite communications through its space segment. Other States have elected to pass specific legislation on a plethora of issues that acutely impact upon satellite communications, such as telecommunications, direct television broadcasting, spacecraft operations, and database management. Aside from these sector-specific pieces of legislation, other more peripheral regulations such as copyrights and trademarks, labor laws, national security provisions, and engineering standards will also impact upon satellite communications.

Satellite Communications in Global Trade

The 1994 General Agreement on Tariffs and Trade (GATT) texts (GATT Secretariat 1994) included telecommunications as trade in services. Since then, regulation of international telecommunications and satellite communications has increasingly been recognized as governed by market access, competition, and deregulation. This is a marked departure from the public law protectionist framework that used to dominate. Through international and regional trade conventions, principles such as reciprocity, the Most Favored Nation (MFN) principle, national treatment, and transparency now operate to regulate the satellite communications industry. Satellite communications now play an instrumental role in trade-in services, foreign direct investment, and transnational trade policy, requiring transnational and international

---

[12] See also generally Buergenthal (1994).

[13] A list of the national space legislations, and other pieces of domestic legislation that may have an impact on satellite communications, can be found at the Website of the United Nations Office for Outer Space Affairs, online at http://www.oosa.unvienna.org/oosa/SpaceLaw/national/def-delim/index.html. Accessed 13 October 2015.

regulatory structures to cope with the expansion of the global commercial satellite communications industry and its attendant needs (Drahos and Joseph 1995).

This section will discuss the legal and trade frameworks in the context of the satellite communications industry.

The General Agreement on Trade and Tariffs (GATT) and the World Trade Organization (WTO)

The 1994 GATT texts, also known as the World Trade Organization (WTO) agreements, were the result of the 1986–1994 Uruguay Round negotiations and were signed at the April 1994 ministerial meeting in Marrakesh.[14] Among the 60 documents, the documents of particular relevance to the satellite communications industry are

1. The Agreement Establishing the WTO (Marrakesh Agreement 1994)
2. The 1994 GATT
3. The 1994 Uruguay Round Protocol to GATT
4. The General Agreement on Trade in Services (GATS)
5. The Agreement on Trade-Related Aspects of Intellectual Property Rights, including Trade in Counterfeit Goods (TRIPs)
6. The Understanding on Rules and Procedures Governing the Settlement of Disputes

One of the most significant decisions of the Uruguay Rounds was the conclusion of the Agreement establishing the World Trade Organization, which establishes a single institutional framework encompassing the GATT, all agreements concluded under it, and the results of the Uruguay Rounds. The 1994 GATT amended the original 1947 GATT, which was focused mainly on goods, to include tertiary-sector markets.

The WTO structure comprises a Ministerial Conference mandated to meet biannually, a General Council, a Dispute Settlement mechanism, and three Councils on services, goods, and intellectual property, respectively.

The General Agreement on Trade in Services (GATS)

The GATS mandates WTO Member States to liberalize and deregulate trade in services through continuing negotiations. Ministers returned to the round table for more negotiations on services pursuant to the November 2001 Declaration of the

---

[14] For a summary of the texts resulting from the Uruguay Round of multilateral trade negotiations, see the Website of the World Trade Organization (2015), online at http://www.wto.org/english/docs_e/legal_e/ursum_e.htm

Fourth Ministerial Conference in Doha with follow-up ministerial discussions taking place in Cancún (2003), Hong Kong (2005), and Geneva (2004, 2006, and 2008).

Satellite communications are directly concerned by the GATS and two of its Annexes.

Article I of the GATS provides the scope and definition for the Agreement. Essentially, trade in services refers to cross-border or transnational supply, foreign consumption, commercial presence abroad, and presence of a natural person abroad. Satellite communications may concern one or more of these modalities of trades. It is significant to note, however, that satellite communications provided by government authorities in the domestic national context is excluded from the GATS.

The second part of GATS concerns the applicable obligations, principles, and rules in the trading context of satellite communications services. It is important to consider these principles in light of the relevant obligations under public international law and international space law that was discussed earlier in this chapter. Of particular significance to the satellite communications industry are the following obligations:

- The application of the Most Favored Nation treatment (Article II)
- Transparency in publishing all relevant measures (Article III)
- No forced disclosure of legitimate confidential information (Article III bis)
- Increasing participation of developing States (Article IV)
- Free economic integration (Article V)
- Free labor markets integration (Article V bis)
- Compliance of domestic regulations with GATS principles (Article VI)
- Recognition of authorization, licenses, and certification granted abroad (Article VII)
- Compliance of monopolies and exclusive service suppliers with the Most Favored Nation treatment (Article VIII)

In the context of satellite communications, Article VIII is of particular import. While WTO Member States are not prohibited from regulating monopolies, each member is obliged to abide by the MFN treatment and must grant the same status to all other competitors, including foreign operators. This implied obligation to deregulate the domestic satellite communications market does not enjoy widespread observance today.

Certain exceptions are acceptable under GATS, including national security and whatever measures Member States deem necessary to protect its security interests and act in compliance with its obligations under the Charter of the United Nations. This ensures that a Member State's application of GATS to the satellite communications industry will comply with its public international space law obligations.

Part IV of GATS focuses on progressive liberalization, mandating "successive rounds of negotiations," to reduce or eliminate adverse effects on trade in services as a means of providing effective market access. In the present context, this again points to an obligation on WTO Member States to progressively liberalize the satellite communications industry and to eradicate hurdles to market entry through

Fig. 1 WTO Headquarters in Geneva, Switzerland

negotiations. (See the WTO Headquarters in Fig. 1 below which was initially the International Labor Organization building.)

Two Annexes to the GATS are specifically applicable to the satellite communications arena. The first is the Annex on Telecommunications, and the second is the Annex on Trade-Related Aspects of Intellectual Property Rights (TRIPs).

The Annex on Telecommunications

Satellite communications have progressively been integrated with commercial telecommunications services. The Annex on Telecommunications recognizes the integration of space-based communications with commercial telecommunications services, both as a tool for other activities and as an activity in its own right. The Annex specifically focuses on the "transmission of signals by any electromagnetic means," restricting its application to public telecommunications as opposed to radio, television, and cable broadcasting. Telecommunications transport services is defined as the "real-time transmission of customer-supplied information between two or more points without any end-to-end change in the form or content of the customer's information," bringing satellite communications within the purview of the Annex. Satellite-based communications using the low Earth orbit systems, including voice- and data-only systems such as mobile telephony and broadband data transfer systems, respectively, fall entirely within the framework of the GATS provisions.

In practical terms, this means that States are obliged to be transparent in the promulgation of domestic tariffs, technical standards, terminal connections, and licensing criteria. Further, access to public network or services should be nondiscriminatory, meaning that information must be free and unrestricted unless in clear contravention of security or confidentiality concerns. More particularly, special measures may be taken to increase the participation of developing States in programs of organizations such as the ITU, the United Nations Development Program (UNDP), and the World Bank. The harmonization of international standards to ensure global interoperability is also to be the linchpin of individual States' interaction with the ITU and other international organizations. It must be noted that to that end, a Decision on Negotiations on Basic Telecommunications has been undertaken by some States in order to promote progressive liberalization and deregulation of telecommunications markets and networks.

These standards and regulations aim to build a borderless global network of international telecommunications, as well as the liberalization of commercial communications networks and markets. While noteworthy strides have been taken by GATS in the area of value-added services in satellite communications and telecommunications, it must be noted that the GATS regulations do not affect "basic" telecommunication services. This means that competition is better provided for in value-added services rather than basic services and that the deregulation of basic telecommunication services still has some way to go toward true liberalization. Moreover, the exceptions provided for by GATS in the face of security and confidentiality concerns may provide governments with a way out of allowing competition into the domestic telecommunications market. These issues raise interesting concerns as well for the operations of satellite communication networks. Pursuant to the agreements made in Doha, negotiations are ongoing to agree on removing regulatory barriers and opening domestic markets to foreign competition.

The Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs)

The TRIPs Agreement is an elaborate stand-alone agreement. This section will address only those provisions of the Agreement that impact upon satellite communications. It will be noted in the preamble that the TRIPs Agreement recalls the conflicting interests of the Member States – action against counterfeit goods, as opposed to the special needs of "least-developed country Members in respect of maximum flexibility . . . to create a sound and viable technological base."

In the scope of this chapter, it must be noted that the basic provisions in relation to intellectual property rights are to be found in international agreements concluded prior to TRIPs and supplemented by it. These are

- 1883 Paris Convention for the Protection of Industrial Property (as amended in 1967 and 1979)
- 1886 Berne Convention for the Protection of Literary and Artistic Work (as amended in 1971 and 1979)
- 1961 Rome Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organizations
- 1989 Washington Treaty on Intellectual Property in Respect of Integrated Circuits

TRIPs, together with the World Intellectual Property Organization (WIPO), oblige Member States to extend the National Treatment and the MFN Treatment to all other members.

Some points are of note about the application of TRIPs and these Conventions to satellite communications. First, Article 10 of TRIPs indicates the applicability of the 1971 Berne Convention for compilations of data without extending to the actual data itself. Such compilations receive protection for 50 years from the date of production of the work. Second, the 1967 Paris Convention applies for trademarks in respect of services, as per Article 16 of TRIPs. Third, in regard of industrial designs, the provisions of the 1883 Paris Convention are of significance in regard to situations of commercial exploitation of satellite communications. Fourth, section "Conclusion" of TRIPs enunciates an approach akin to the trade secret concept, by which "undisclosed information" is protected from unauthorized disclosure without consent.

Section "Introduction" of TRIPs clarifies that the object of the Agreement is to achieve effective action, fairness and equity, decision on merits, judicial review, and Member States' freedom in the enforcement of their own domestic laws. Section "International Regulation of Satellite Communications" lists a full range of dispute settlement procedures and remedies, with provisional measures provided for in section "Regional Regulation of Satellite Communications." These three sections provide for the enforcement of intellectual property rights under TRIPs and are applicable to satellite communications where such activities fall under the TRIPs regime.

Dispute Settlement, Responsibility, and Liability Issues

What happens in the event that the regulatory framework does not prevent disputes from arising, or damage from occurring? This section will address issues related to the settlement of disputes arising from satellite communications, as well as the location and substantive obligations raised by legal principles related to responsibility and liability for such disputes and damage.

Dispute Settlement

In the specific case of satellite communications, there are several parties to consider when a dispute arises. These include

- The satellite operator
- The satellite owner
- The satellite user
- The satellite manufacturer
- The satellite launcher
- The satellite insurer
- The regulatory agency of the country linked to the satellite
- The third party

Perhaps unsurprisingly, there are very few instances of claims filed in the case of damage caused by a satellite, whether domestically or internationally. This may be due to various reasons, but in particular because of the prevalence of cross-waivers of liability and exculpatory provisions, which act as bars to litigation.[15] Litigation or other adversarial claims take a long time to resolve, which is counterproductive for business. Further, a closed field such as satellite communications tends to attempt the preservation of good relationships between parties, which may be threatened by such adversarial processes.[16]

Aside from a dearth of publicly settled disputes in satellite communications, there is also to date no international case brought before open court. The only provision for a dispute settlement mechanism is found in the 1972 Liability Convention, which provides for the formation of a Claims Commission to determine compensation in the case of damage caused by a space object. It will be noted that there is no provision for a dispute settlement mechanism in the case that there is a dispute as to the responsibility or wrongdoing or indeed any merits issue relating to the claim. Many highly qualified publicists have, however, noted that arbitration appears to be the favored method of dispute settlement in space activity-related disputes.[17]

Responsibility and Liability

Satellite communications is considered first and foremost an activity in outer space. This means that, by virtue of Article VI of the Outer Space Treaty, States bear international responsibility for the acts of their nationals, whether these activities are by public organizations or private commercial entities. Article VII of the Outer Space Treaty stipulates that States are liable for damage caused by activities in outer space. The 1972 Liability Convention stipulates a detailed system for liability and claims for damage caused by space activities.

The international responsibility of a State may be invoked only if the State commits an internationally wrongful act and if the act is attributable to it. In respect of international responsibility, States have generally ensured that they are in compliance with Article VI of the Outer Space Treaty by enacting domestic legislation requiring governmental authorization in the form of licenses for space activities by its nationals or on its territory. The structure provided by these domestic legislations and licensing schemes has several advantages for the State in question. First, it ensures that the relevant information about the space or satellite undertaking is provided to the regulatory authority. Second, it allows for the requirement of license renewal, which allows ongoing supervision of the space activity as required under Article VI of the Outer Space Treaty. Third, it provides the framework on which liability may be passed or shared between the government and entity involved. This

---

[15] On the topic, see Salin at p. 39 and Larsen (1992).

[16] Fewer than 20 lawsuits, for example, have been filed in the United States of America, arguably one of the most active countries in satellite communications. For an overview, see Meredith (1995).

[17] See generally Goh (2007).

last point is of particular significance in the case where private entities undertake commercial space activities, as is the case in many satellite communications projects. In many cases, the license which grants permission for the space activity usually carries with it a requirement for the private entity to prove the undertaking of sufficient insurance in the case of damage caused.

When damage occurs due to the space activity, a dual system of international liability is envisaged by the 1971 Liability Convention and imposed upon launching States. Absolute liability is prescribed for damage caused by the space object on the surface of the Earth or to aircraft in flight. Exoneration is only provided for in the case where there was an act of gross negligence or omission on the part of the claimant State, unless the launching State itself was in violation of international law. On the other hand, fault liability is prescribed for damage caused elsewhere than on the surface of the Earth to a space object or persons or property on board the space object in question. In both cases, damage refers to personal injury or damage to property, whether natural or juridical.

Two points are of particular interest in the Liability Convention. First, Member States of international organizations are jointly and severally liable for damage caused if that organization has acceded to the rights and obligations of the Convention. Second, the Liability Convention addresses liability claims as between States – and not between private individuals or entities. Of particular significance is the fact that the Convention does not apply to nationals of the launching State or to foreign nationals of the participating States. Only three categories of States are envisaged as in a position to make a claim: the State of nationality, the State of the territory on which damage occurred, and the State whose permanent residents have suffered the damage. The claim for compensation should be presented within a year after the damage or identification of the liable State, through diplomatic channels, another State, or the Secretary General of the UN. The exhaustion of local remedies is not necessary for a claim to be presented. Where claims do not produce a settlement, the Convention provides for the establishment of a Claims Commission.

In addition to the international liability outlined above in this section, another issue that may arise in relation to satellite communications is in relation to disputes involving private commercial entities and claims arising therefrom. Situations which may incur claims may include actions by such an entity in violation of international legal limitations (including the use of nuclear power), actions in violation of regional or national regulatory regimes (ranging from anticompetitive practices to labor laws and licensing requirements), and the acquisition, operation, modification, or termination of the enterprise (such as financing, shareholders' rights, dividends, bankruptcy, and so on).

Conclusion

This chapter aims to provide a concise overview of the many regulatory, legal, and trade issues that concern satellite communications. The overarching public service concerns, in light of growing commercialization and liberalization, has led to a complex labyrinth of rights and obligations on entities involved in the satellite communications field. Owners, operators, end users, regulatory agencies, and public institutions must keep this dual nature of satellite communications in mind when considering the regulatory and legal matters that arise. The increasing role of satellite communications in the field of global trade and services exchange also compounds the intricacy of the applicable regulatory frameworks. A heartening development is the global inclination toward harmonization and coordination. A coherent and practicable legal, regulatory, and trade framework for satellite communications can only be put into place with mutual collaboration and innovative foresight. Complementing this move toward international collaboration has been the adoption of "model space laws" by national legislatures that have set new standards in such areas as control of orbital debris, liability coverage, trade equity protections, and due diligence of regulatory review prior to launch.

Cross-References

▶ International Committee on GNSS
▶ Mobile Satellite Communications Markets: Dynamics and Trends
▶ Regulatory Process for Communications Satellite Frequency Allocations
▶ Satellite Communications and Space Telecommunication Frequencies
▶ Space Telecommunications Services and Applications
▶ Trends and Future of Satellite Communications

References

Agreement Relating to the International Telecommunications Satellite Organization (INTELSAT) with annexes and Operating Agreement. Entered into force 12 Feb 1973. 23 UST 3813, 4091 (1973)

Arabsat, http://www.arabsat.com/. Accessed 13 Oct 2015

Association of South East Asian Nations (ASEAN), http://www.aseansec.org/. Accessed 13 Oct 2015

R. Bender, *Launching and Operating Satellites* (Utrecht Studies in Air and Space Law/Nijhoff, Dordrecht, 1998)

B. Bishop, *European Union Law for International Business: An Introduction* (Cambridge University Press, Cambridge, 2009)

T. Buergenthal, Self-executing and non-self-executing treaties in national and international law, in *Collected Courses of the Hague Academy of International Law* (The Hague, 1994), p. 295

D. Chalmers et al. (eds.), *European Union Law: Cases and Materials*, 2nd edn. (Cambridge University Press, Cambridge, 2010)

B. Cheng, Revisited: international responsibility, national activities and the appropriate State. J. Space Law **26**, 7 (1998)

B. Cheng, Military use of outer space: article IV of the 1967 outer space treaty revisited, in *The Utilization of the World's Air Space and Free Outer Space in the 21st Century*, ed. by C.-J. Cheng, D.H. Kim (Kluwer Law International, The Hague, 2000), p. 305

C.Q. Christol, The sharing of access and resources by states of varying capacities, in *C.Q. Christol Law: Present, Past and Future*, ed. by C.Q. Christol (Kluwer, Deventer, 1991), p. 289

R.R. Colino, The possible introduction of separate satellite systems: international satellite communications at the crossroad. Columbia J. Transl. Law **24**, 13 (1985)

Convention Establishing the European Telecommunications Satellite Organization (EUTELSAT). Entered into force 3 July 1985. Misc. No. 25, Cmnd, 9069 (1985)

Convention for the establishment of a European organisation for the exploitation of meteorological satellites (EUMETSAT). Entered into force 19 June 1986. Bundesgesetzblatt, Federal Republic of Germany (1987) Teil 11, p. 256

Convention of the European Space Agency and Rules of Procedure of the ESA Council. Entered into force 30 Oct 1980. ISSN 1010-5697

Convention on Registration of Objects Launched into Outer Space (Registration Convention). Entered into force 15 Sept 1976. 28 UST 695 (1976)

Convention on the International Liability for Damage Caused by Space Objects (Liability Convention). Entered into force 1 Sept 1972. 4 UST 2389 (1972)

Convention on the International Maritime Satellite Organization (INMARSAT). with annexes and Operating Agreement. Entered into force 16 July 1976. 31 UST 1, 135 (1976)

A. Dashwood et al., *Wyatt and Dashwood's European Union Law* (Hart, Oxford, 2012)

Developing the high-speed telecommunications links (electronic highways) for the Community's 1992 market, European Commission COM (88) 341 (Brussels, 1988)

M.S. Dodge, Sovereignty and the delimitation of airspace: a philosophical and historical survey supported by the resources of the Andrew G. Haley Arch. J. Space Law **35**(1), 5–36 (2009)

P. Drahos, R.A. Joseph, Telecommunications and investment in the great supranational regulatory game. Telecommun. Policy **19**(8), 619–635 (1995)

E. Ducasse, *L'Europe des Télécommunications par Satellite: entre Libéralisation et Coopération* (European Centre for Space Law, Paris, 1993)

EUMETSAT, http://www.eumetsat.int/. Accessed 13 Oct 2015

European Commission Decision (International Private Satellite Partners). 15 Dec 1994. OJEC L 354/75

European Council Resolution on the Development of the Common Market for Satellite Communication Services and Equipment. Entered into force 19 Dec 1991. (92/C8/01) (OJ C8/1) (1992)

European Space Agency, Directorate of Telecommunications and Integrated Applications, http://www.esa.int/esaTE/index.html. Accessed 13 Oct 2015

EUTELSAT, http://www.eutelsat.com/. Accessed 13 Oct 2015

GATT Secretariat, *The Results of the Uruguay Round of Multilateral Trade Negotiations – The Legal Texts* (World Trade Organization, Geneva, 1994)

G.M. Goh, Keeping the peace in outer space: a legal framework for the prohibition of the use of force. Space Policy **20**, 259 (2004)

G.M. Goh, *Dispute Settlement in International Space Law: A Multi-door Courthouse for Outer Space* (Martinus Nijhoff, Leiden, 2007)

S. Gorove, Interpreting Article II of the outer space treaty. Fordham Law Rev. **37**, 349 (1969)

S. Gorove, Aerospace object – legal and policy issues for air and space law. J. Space Law **25**(2), 101–112 (1997)

K.M. Gorove, Delimitation of outerspace and the aerospace object – where is the law. J. Space Law **28**(1), 11–28 (2000)

L. Haeck, Aspects juridiques de certaines utilisations de léspace. Ann. Air Space Law **XXI**, 65–104 (1996)

T. Hartley, *The Foundations of European Union Law* (Oxford University Press, Oxford, 2010)

S. Hobe, B. Schmidt-Tedd, K.-U. Schrogl, G.M. Goh, *The Cologne Commentary on Space Law*, vol. I (Carl Heymanns, Cologne, 2009)

Inmarsat, see its website, http://www.inmarsat.com/. Accessed 13 Oct 2015

Intelsat, http://www.intelsat.com/. Accessed 13 Oct 2015

International Electrotechnical Commission (IEC), http://www.iec.ch/. Accessed 13 Oct 2015

International Organization for Standardization (ISO), http://www.iso.org/iso/home.html. Accessed 13 Oct 2015

International Telecommunications Union (ITU), http://www.itu.int/en/pages/default.aspx. Accessed 13 Oct 2015

Intersputnik, http://www.intersputnik.com/about/information/. Accessed 13 Oct 2015

R. Jakhu, International regulations of satellite communications, in *Legal Aspects of Space Commercialization*, ed. by K. Tatsuzawa (CSP, Tokyo, 1992), pp. 78–101

R.S. Jakhu, Safeguarding the concept of public service and the global public interest in telecommunications. Singap. J. Int. Comp. Law **5**, 71–102 (2001)

N. Jasentuliyana, Recent developments in the United Nations activities relating to outer space. J. Space Law **23**(2), 172 (1995)

N. Jasentuliyana, R.S.K. Lee, *Manual on Space Law*, vol. 3 (Oceana, New York, 1979)

A. Kaczorowska, *European Union Law* (Routledge-Cavendish, New York, 2010)

A. Kerrest, Remarks on responsibility and liability, in *Proceedings of the 40th Colloquium on the Law of Outer Space*, vol. 40 (Paris, 1997), p. 134

G. Lafferranderie, *Les telecommunications par satellites, aspects juridiques* (Editions Cujas, Paris, 1967)

P. B. Larsen, Cross-waivers of liability, in *Proceedings of the 35th Colloquium on the Law of Outer Space*, vol. 35 (Paris, 1992), p. 91

F. Lyall, *Law and Space Telecommunications* (Dartmouth, Aldershot, 1989)

F. Lyall, P.B. Larsen, *Space Law: A Treatise* (Ashgate, Surrey/Burlington, 2009)

Marrakesh Agreement Establishing the World Trade Organization of April 15, 1994, *The Legal Texts: The Results of the Uruguay Round of Multilateral Trade Negotiations* 4 (1999), 1867 U.N.T.S. 154, 33 I.L.M. 1144 (1994) (hereinafter “WTO Agreement”)

N.M. Matte, *Space Activities and Emerging International Law* (Centre for Research of Air and Space Law, McGill University, Montreal, 1984)

P. Meredith, Spacecraft failure-related litigation in the US: many failures, but few suits, in *Proceedings of the 38th Colloquium on the Law of Outer Space*, vol. 38 (Paris, 1995), p. 22

Ministerial Declaration on Trade in Information Technology Products. 18 Dec 1996, World Trade Organization. WT/MIN/(96)/DEC. 9–13 Dec 1996, https://www.wto.org/english/docs_e/legal_e/itadec_e.pdf. Accessed 13 Oct 2015

North American Free Trade Agreement (NAFTA), https://ustr.gov/trade-agreements/free-trade-agreements/north-american-free-trade-agreement-nafta. Accessed 13 Oct 2015

Regional African Satellite Communications Organization (RASCOM), http://www.rascom.org/. Accessed 13 Oct 2015

M.A. Rothblatt, The impact of international satellite communications law upon access to the geostationary orbit and the electromagnetic spectrum. Tex. Int. Law J. **16**, 207–244 (1981)

P.-A. Salin, *Satellite Communications Regulations in the Early 21st Century* (Utrecht Studies in Air and Space Law/Nijhoff, Dordrecht, 2000), pp. 11–12

I.H. Shefrin, The NAFTA: telecommunications in perspective. Telecommun. Policy **17**(1), 14–26 (1993)

M.L. Smith, *International Regulation of Satellite Communication* (Utrecht Studies in Air and Space Law/Nijhoff, Dordrecht, 1990)

N. Tannenwald, Law versus power on the high frontier: the case for a rule-based regime for outer space. Yale J. Int. Law **29**, 363 (2004)

Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies (Outer Space Treaty). Entered into force 10 Oct 1967. 610 UNTS 205 (1967)

United Nations Committee on the Peaceful Uses of Outer Space (COPUOS), http://www.oosa.unvienna.org/oosa/COPUOS/copuos.html. Accessed 13 Oct 2015

F. von der Dunk, Liability versus responsibility in space law: misconception or misconstruction? in *Proceedings of the 34th Colloquium on the Law of Outer Space*, vol. 34 (Paris, 1991), p. 363

R.L. White, H.M. White Jr., *The Law and Regulation of International Space Communication* (Artech House, Boston/London, 1988)

M. Williams, The principle of non-appropriation concerning resources of the moon and celestial bodies, in *Proceedings of the 13th Colloquium on the Law of Outer Space*, vol. 13 (Paris, 1970), p. 157

World Trade Organization, http://www.wto.org/. Accessed 13 Oct 2015

Trends and Future of Satellite Communications

Joseph N. Pelton

Contents

Abstract

Satellite communications technologies have achieved remarkable breakthrough efficiencies and increases in performance in nearly a half century. These developments, however, have occurred in parallel with large gains in performance by

---

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_24

other IT and telecommunications systems. Thus, these dramatic gains are not as apparent to the general populace as might have been the case if this explosion in performance had happened in isolation.

In many ways today's satellites are digital processors in the sky and specialized software defines how they perform and defines their communications capabilities. In fact, the innovations in satellite communications as well as the progression in all forms of telecommunications and computer processes have followed similar courses. In short, Moore's law that predicted a doubling of performance every 18 months has generally held true for all fields involving digital processing, whether it be computing, communications, video games, or even digital entertainment systems. What had been past is thus likely to be prologue. It is reasonable to anticipate continuing gains in terms of overall processing power, digital communications, and "intelligent" space communication systems.

In short, there are remarkable new technologies still to be developed in terms of space-based satellite communications systems, more powerful processors, new encoding capabilities, and new user terminal capabilities that can make user systems more mobile, more versatile, more personally responsive, more powerful in terms of performance, and yet lower in cost (J.N. Pelton, *Future Trends in Satellite Communication* (International Engineering Consortium, Chicago, 2005), pp. 1–19; Also see T. Iida, J.N. Pelton, E. Ashford, *Satellite Communications in the 21st Century: Trends and Technologies* (American Institute of Aeronautics and Astronautics, Reston), pp. 1–15, 2003).

As the world national economies become more global and as all parts of the globe, the oceans, and the atmosphere are exploited by human enterprise, the need for effective wireless interconnection via terrestrial wireless and satellite communications will expand. Further, the increased utilization of space systems to explore outer space – manned and unmanned – will increase the need for improved space communications systems. Clearly foreseeable technologies suggest that several more decades of continuing innovations are now possible. But technology will not be the only source of change for the satellite communications industry. Other drivers of change will include: (a) new service demands in both civilian and defense-related markets; (b) restructuring of commercial satellite organizations through acquisition, merger, and regulatory change; (c) new allocations or reallocation of frequencies and increased frequency interference; (d) convergence between and among the various satellite applications markets – both in terms of technology and structural integration; (e) constraints in orbital configurations; and (f) increased concerns with regard to orbital debris. Further, the growth of human activities in outer space may prove to be significant shapers of new satellite systems in the next 20–30 years (J.N. Pelton, *Future Trends in Satellite Communication* (International Engineering Consortium, Chicago), pp. 1–19, 2005; T. Iida, J.N. Pelton, E. Ashford, *Satellite Communications in the 21st Century: Trends and Technologies* (American Institute of Aeronautics and Astronautics, Reston), pp. 1–15, 2003).

Keywords

Application specific integrated circuits • Bits per Hz • Codec Encoding • Extremely high frequencies (EHF) • Integrated space and terrestrial systems • Millimeter and terahertz wave frequencies • Modems • Onboard processing • Orbital congestion • Orbital debris • Phased array antennas • Quantum computing • Rain attenuation • RF interference • "Smart" antennas • Super high frequencies • Transmission efficiency • Turbo coding

Introduction

Today's communications satellites represent a very impressive gain in performance when compared to those first deployed almost a half century ago. Contemporary satellites' solar arrays can generate well over 100 times more power, and advanced multibeam satellite antenna systems can deliver the equivalent of up to a 1,000 times usable bandwidth than that of the Early Bird Satellite – the world's first commercial satellite spacecraft. Deployable solar arrays have become larger in size, photovoltaic cells have improved in performance, and improvements in design have allowed the arrays to achieve maximum exposure to the sun. Battery systems have also improved with greater power density and longevity. Satellite antenna systems have evolved and improved in many different ways. These have included better pointing and focusing of radio frequency (RF) energy, multibeam antennas, frequency reuse strategies, and improved large-scale antenna manufacturing techniques. Overall, an ongoing series of technological improvements have increased the performance and lifetime of satellite systems in space and have made the user equipment on the ground easier to use, more accessible, and lower in cost (Pelton 2006).

The future suggests that many of these powerful trends will continue. There are, however, key challenges. One of these challenges that has been discussed in the previous section is that of integrating satellite communications systems with terrestrial wireless and broadband fiber and coaxial cable systems. The other challenge is to adapt satellite technology to a changing world. This could mean many things. It means effective use of satellite systems not only to communicate across the Planet but to points beyond throughout the Solar System (European Space Agency 2008). It could mean more integrated space applications so that user devices could provide not only voice, data, and video signals, but space navigation and location services, Earth imaging, weather and meteorological data, and other desired information on demand.[1]

The future will thus be shaped more by new service and market demands than new satellite technologies. Indeed regulatory shifts, industrial consolidation, constraints imposed by orbital debris, and even a change in financial and insurance markets could also dictate major shifts in the satellite communications industry.

---

[1] Op cit. J.N. Pelton, *Future Trends*, pp. 109–115.

Some might suggest that now satellite communications systems have reduced in size from giant 30 m multiton Earth stations to handheld transceiver devices; there is little further room for further innovations. But history has proven forecasters wrong many times in the past. Forecasters such as Thomas Watson, Chairman of IBM, once thought that the world would need only a dozen computers to be used by elite scientists. Others thought that trains would travel no faster than a 100 km an hour (or about 60 miles/h) because of wind resistance. It was suggested in the nineteenth century that patent offices could be closed because all important inventions had already been registered. Demand for new services and new capabilities in human society constantly gives rise to new technologies which in turn generate new applications and the process regenerates itself again and again. Sometimes enthusiasm for technology overestimates future trends as well. For the field of satellite communications, projecting demand for new services can outweigh technological innovation in achieving accurate forecasts. Indeed predictions based on technological innovation cannot only often be wrong, but frequently greatly overstated (Schnaars 1989).

The Path Forward

The world of satellite communications is quite complex and technical, but the dynamic range of physical systems within which the new satellite networks are defined is remarkably small. Antenna systems focus “power” and electromagnetic energy in the form of “radio frequencies” or optical signals. To increase performance, antennas must focus power more effectively or have access to more power or find a way to utilize available frequencies more effectively, either in higher frequency ranges or by more effective “reuse of the frequencies” or by both. These are the range of tools available to make satellite communications more effective. Of course one can invent more effective ways to send more “usable information” via a communications channel, whether that be a fiber optic link, a terrestrial wireless link, or a satellite. The way forward essentially lies along one of these pathways. This section thus explores the future in terms of more effective antennas, improved power, and more effective spacecraft design – including improved lifetime, reliability, and pointing systems, improved satellite orbital configurations, improved transmission capabilities, improved signal coding and decoding (i.e., complexity), and finally improved user transceivers.

Advanced Spacecraft Antenna Design

The key to a satellite antenna’s performance involves how well it can focus an RF or optical beam toward the designed reception or “catchment area.” This characteristic of the antenna to concentrate a signal is called antenna gain. A larger aperture antenna can create a narrower beam and thus there is less path loss due to the spreading of the signal between the satellite and Earth. Today’s largest aperture satellite antennas with diameters on the order of 20 m or more can be used in conjunction with a multifeed system to create many hundreds of highly focused and narrow beams that allow intensive frequency reuse. This type of advanced multibeam, large aperture communications satellite antenna can be observed in such spacecraft as the Viasat 1 and 2, Jupiter, Intelsat Epic, Terrestar, Skyterra, and Inmarsat Express satellites. This is because RF beams that are geographically separated from one another by a sufficient distance can use the same frequencies over and over again. The question naturally arises as to just how large can satellite communications space antennas grow without structural or cost barriers to their future expansion? (Pelton 1998; Iida and Suzuki 2001)

The answer to this question turns out to be rather complicated in that there are a variety of ways that one can create narrow beams for the purposes of reducing path loss and allowing intensive frequency reuse. These strategies can often be applied in parallel and thus are not necessarily mutually exclusive. The "best design" for the space antennas of the future might thus involve a combination of these various approaches.

- *Use of Higher Frequency Antennas with a Smaller RF Wavelength*: If one moves to higher frequencies and thus utilize smaller wavelengths, the effective capability of a satellite antenna and its "gain" changes exponentially as one moves to higher frequencies. Thus when a communications satellite employs a higher frequency the antenna's aperture can be smaller. Since the spacecraft antennas are transmitting and receiving smaller wavelength signals, a smaller antenna can achieve the same result – or effective throughput capacity – as another larger spacecraft antenna operating at a lower frequency and thus employing a larger RF wavelength. Indeed, since the "gain" of an antenna is inversely proportional to the square of wavelength, this makes a dramatic impact on the required size of the antenna needed to achieve the same effective performance. The largest satellite antennas today are for mobile satellite communications and the aperture size of this type of antenna is driven to a larger dimension because the antennas for down-linking signals to mobile users are typically in the range of 1,700 MHz to 2,500 MHz. These "lower radio frequencies" are used in part because the signals do not need to have direct line of sight to the satellite and can complete the link without necessarily having to "see" the user terminal that might be partially blocked by the top of a car or a telephone pole (Wakana 2003. Also see Hoeber). The down side of this consideration is that the antennas operating in these mobile satellite frequencies in the L band and UHF frequencies need to be larger to shape the longer wavelengths for transmission. In contrast, the satellites that use higher frequencies such as the Ka band (30 and 20 GHz), for instance, require direct line of sight. This will be even more the case with use of the Q/V or W bands or perhaps even terahertz wavelength in the future. This need of direct line-of-sight connection is thus a major constraint for satellite connections of the future and as higher frequencies are used. Since the aperture size of antennas is driven by the square of the wavelength, the aperture size of a parabolic antenna calculates to be 10^2 or 100 times smaller each time the frequency increases by a power of ten. The shortage of available bandwidth in the lower bands, however, is pushing satellite systems operation toward these higher frequencies for services especially those for applications other than mobile satellite communications – namely – fixed or broadcast satellite services. Unfortunately, there are problems with rain and other types of precipitation attenuation at these higher frequencies. This requires higher link margins. This in turn translates into requirements for either higher power or larger and higher gain antennas. Also the electronics technology is much more demanding in terms of requiring the generation of quite precise, tiny wavelengths and very high frequencies. This tends to drive costs higher. Although the parabolic antenna's aperture may be smaller, its contours must be much more exactly shaped to direct the smaller wavelength beam in exactly the correct way. Again this exacting contour also drives antenna fabrication costs higher. Further, the satellite must also be much more exactly pointed toward Earth so that the beams can be more precisely targeted. In short, while the higher frequencies and shorter wavelengths allow the spacecraft antennas to be smaller, the complications just cited can more than offset the advantages of the smaller aperture size and result in higher manufacturing costs. New electronics technologies for the EHF bands, strategies to address precipitation attenuation, and exacting manufacturing techniques all combine to drive up the costs. Eventually, however, these difficulties of migration to utilize these new and higher frequency bands are overcome. As more and more satellites and ground systems are manufactured and deployed in these new bands, the costs tend to go down. As the Ka-band systems are deployed the next horizon will be the next frontier, which are the frequency bands in the 48 and 38 GHz bands.

- *Phased Array Antenna Technology*: There is a new technology that is well suited to creation of larger-scale spacecraft antennas to improve satellite performance in the future. This technology is called phased array antenna systems. With this type of antenna an array of electronic components are combined to create "virtual" high performance and highly focused antenna beams. This type of antenna can support highly efficient multibeam transmissions. The result is an antenna system that allows effective reuse of available RF frequencies many, many times. There are two ways that this technology can be used. One way is simply by directly extending today's satellite technology. This approach continues to deploy a very large high-gain conventional antenna reflector but uses a phased array multibeam feed system to generate a very large number of beams by reflecting off of a large parabolic reflector (see Fig. 1 below) (Iida and Pelton 2003).
- As shown above, this could be a "tethered" large-scale reflector or "untethered." The feed system could use a phased array feed or horn feed array using more conventional technology.
- The second and more technically advanced way to proceed would be to design a "phased array antenna where the various electronic components actually 'electronically form' a beam for transmission or reception from a 'virtual reflector'." With this more advanced technology, one can create a virtually shaped electronic beam of any shape and of large-scale dimension. Since the beam in this case is "virtually created" the effective size of the beam can grow quite large by going

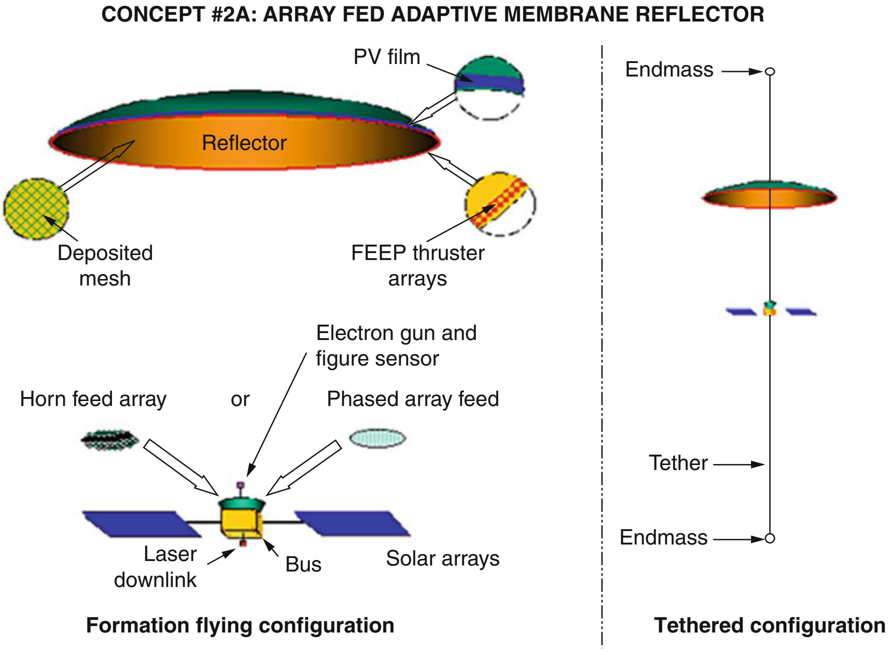
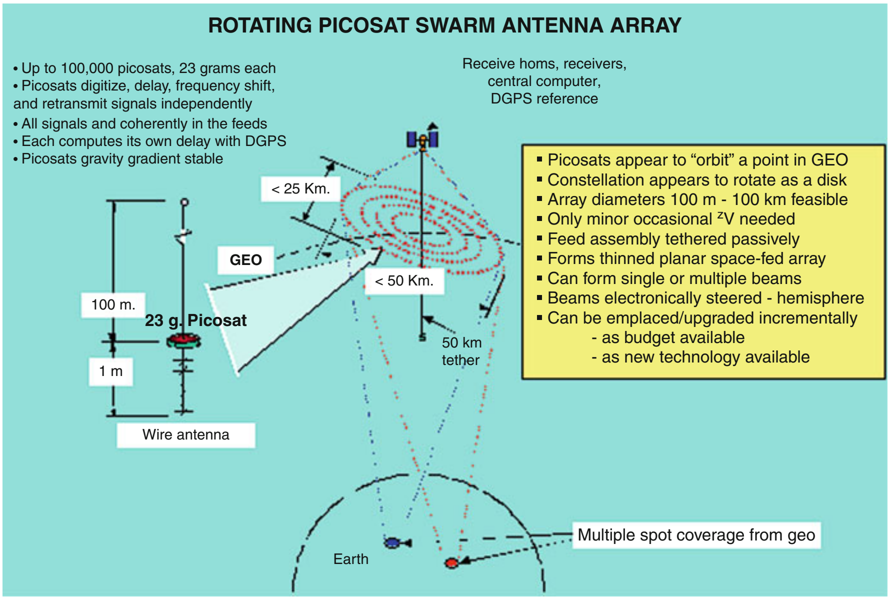


Fig. 1 Array fed adaptive membrane reflector – tethered or untethered (Graphic courtesy of Joseph Pelton and Ivan Bekey)

from something like a "six by six" phased array to a "twelve by twelve" or even a "hundred by hundred" phased array. A hundred by hundred phased array could, in theory, generate 10,000 different "pencil thin" beams and allow more than a 1,000-fold reuse of the same spectrum band available for satellite communications. The problem with this approach is that the technology is still at a very early stage and creation of a phased array antenna of this type is quite expensive.

The extension of phased array antenna technology could potentially go quite far. One concept is to deploy a large number of phase array components into space as a free-flying cluster. The microelements of a "virtual antenna reflector" could form a cluster of distributed "picosatellite array components" covering perhaps square kilometers and create beams that would create "picocells" on Earth (see Fig. 2 below) (Iida and Pelton 2003, pp. 188–190).

With such a device, the ability to reuse RF frequencies might climb to perhaps hundreds of thousands of times. At this time such concepts are just that. There would be many technical problems to be solved. These would include the issue of how to recollect all of the phase array components – perhaps with magnetic attraction – so as to minimize orbital space debris. There are also issues of interference with other satellite systems and terrestrial communications systems that utilize the same frequencies.



ROTATING PICOSAT SWARM ANTENNA ARRAY

- Up to 100,000 picosats, 23 grams each
- Picosats digitize, delay, frequency shift, and retransmit signals independently
- All signals and coherently in the feeds
- Each computes its own delay with DGPS
- Picosats gravity gradient stable

Receive homs, receivers, central computer, DGPS reference

- Picosats appear to "orbit" a point in GEO
- Constellation appears to rotate as a disk
- Array diameters 100 m - 100 km feasible
- Only minor occasional ^{Z}V needed
- Feed assembly tethered passively
- Forms thinned planar space-fed array
- Can form single or multiple beams
- Beams electronically steered - hemisphere
- Can be emplaced/upgraded incrementally
  - as budget available
  - as new technology available

Fig. 2 Conceptual design of a picosatellite array free-flying in space (Graphic courtesy of Joseph N. Pelton and Ivan Bekey)

There are other technologies that might be employed to boost future satellite communications performance as well. One future approach that might be used is what is called either "Scanning beam" or "Hopping beam" technology. This is a type of technology that allows a number of spot beams on an advanced satellite to work dynamically in the time domain using time division multiple access (TDMA) or code division multiple access (CDMA) multiplexing. In this type of antenna configuration, beams can be directed to different locations and bursts of data of varying durations (measured in a few milliseconds) can be sent to various destinations depending on levels of traffic demand. This technology allows streams of broadband traffic (voice, data, and video) to be sent in a burst to a particular location covered by a spot beam and then "hop" to the next location, and then "hop" to another location, and so on at extremely short time intervals. The advantage of this type of "hopping beam" is that the duration of the broadband blast of the digital data stream can be adjusted to the times of day as peak loads vary from time zone to time zone. Also if the satellite is operating in the higher frequency bands such as the Ka band (30 and 20 GHz), or in the future "Q/V" bands (48 and 38 GHz), or "W" bands (60 GHz) then the dwell time of the data blast in a particular spot beam can be adjusted to compensate for a heavy rain storm or other forms of rain attenuation. A dwell time of perhaps 10 min for a data transmission in a particular beam might be doubled or even tripled in duration in an area where there is heavy rain rate and thus severe rain attenuation. These types of concepts were tested in the US experimental communications satellite program undertaken by NASA known as the Advanced Communications Technology Satellite (ACTS).[2]

Regardless of the particular antenna designs of the future, the challenge will likely be to achieve the ability to generate more and more focused spot beams that can allow greater frequency reuse and lesser path loss by having transmission beams that "spread out" less as the signal travels between the satellite and Earth or the reverse direction. The key will be to add intelligence along with the improved antenna systems to allow the antennas to work more efficiently. This increased "intelligence" could be applied in the form of "beam hopping" or "beam scanning" so that digital transmissions within particular beams could better be matched to actual overall demand for communications services, changing peak load requirements, adjusting to heavy rain or other atmospheric conditions, or otherwise make satellite system more versatile to demand changes or system constraints.

Improved Transmission Systems and Onboard Processing Systems

The key to the efficient throughput of communication services hinges first and foremost on enhanced digital processing capabilities. The efficiency of satellite communications systems can be measured today most directly in terms of digital throughput or simply in bits per Hertz. Techniques such as interconnection of geographically separate spot beams, polarization discrimination, and operation at higher frequency bands – where wider spectrum bands are allocated to satellite communications – allow a satellite to expand available bandwidth. Digital encoding – and in particular more efficient coders and decoders (Codecs) and coding systems – plus improved digital processing, modulation and multiplexing techniques allow more bits to be delivered per Hertz. A decade ago a typical communication could provide 1 bit per Hz and could use a variety of techniques to reuse available spectrum by something like six to eight times. Today, through the use of advanced codecs communications satellites can derive something like 2.5 to even 5 bits per Hz. With advanced multibeam antennas they can reuse spectrum by factors on the order of 20–50 times or more. This is particularly critical in the case of mobile satellite systems that operate at the lowest frequencies and thus have the smallest amount of useable spectrum. This trend will continue.

The trends for satellite communications and terrestrial fiber optic connections, however, will follow different patterns for specific technical reasons that separate how communications satellites and terrestrial cable operate. In the case of fiber optic networks, there are two specific advantages of these terrestrial networks over satellite networks. The fiber optic networks operate, not in the RF frequencies but in the much higher light wave region of the electromagnetic spectrum. There is an incredibly large amount of spectra available for communications in the optical

---

[2] Advanced Communications Technology Satellite Overview, Nasa Goddard Research Center, acts.grc.nasa.gov and C.B. Cox, T.A. Coney, *Advanced Communications Technology Satellite (ACTS) Adaptive Rain Fade Protocol Performance*, acts.grc.nasa.gov/docs/4thKa_Cox_Coney.pdf

wavelengths. These networks typically use dense wave division multiplexing (DWDM) currently operating in these optical frequency bands. They can multiplex signals over and over again only a quarter of a nanometer apart in order to achieve tremendous broadband throughput speeds. Further, fiber optic cables have very little sources of external noise or interference and thus the quality of the signal remains very high over relatively long transmission distances. This means that fiber optic networks, due to broad available spectra and low noise, do not have to be nearly as concerned with transmission efficiency as is the case with satellite networks. Further, since satellites must interconnect uplinks and downlinks and increasingly interconnect different uplink and downlink beams, satellite systems need to use time-based multiplexing systems so there is time for digital processing associated with these switching operations. Fiber optic networks work almost exclusively with “wave division.” This is in part because fiber transmissions do not require “time division” intervals for processing purposes associated with interconnecting different beams and other complications. In short, fiber networks are quite unlike the case of satellite networks that must cope with the problems of beam interconnection, time delay spoofing, etc. These differences in multiplexing techniques are important as these competing approaches tend to separate the fiber world and the satellite world. The terrestrial wireless world of telecommunications, however, has similar constraints, particularly with regard to beam interconnection and multiple reuse of the same frequencies over and over again. Thus, the ever increasing worldwide demand for mobile services and the similarity of satellite and terrestrial wireless networks (including their dependence on processing time for “beam” or “cell” interconnection) helps to tie the terrestrial wireless and the satellite networks together. This similar approach to time-based multiplexing, as used by satellites and terrestrial mobile, works to ensure that the future standards for interconnection of satellites, terrestrial wireless and fiber optic networks will keep compatible protocols for universal, worldwide communications linkages. Clearly, a challenge for the future will be to keep all forms of telecommunications transmission media to interconnect as “seamlessly and compatibly” as possible (Sachdev 2004).

Improved Satellite Power Systems

The story of improved satellite power systems has had several components. First of all, solar arrays have greatly increased in size. Deployment systems for these arrays have grown more sophisticated to allow these very large-scale systems to unfold or otherwise be deployed from the compact configurations required to fit within the rocket fairings at launch. Secondly, the “efficiency” of photovoltaic (PV) solar cells performance in terms of converting solar energy into the power required to generate RF signals has also increased. Solar cells have improved from amorphous silicon solar cells, to structured silicon to gallium arsenide cells. Further, the number of gates or junctures where solar energy is captured have increased and moved up into the ultraviolet part of the spectrum where the maximum amount of energy is obtainable. In short, the efficiency of energy conversion has increased from around

7–30 % and soon may become close to 50 % in the most efficient systems. Not only have the solar cells improved in performance, but the ability of the satellites to display PV cells so as to achieve maximum solar radiation has also improved. The change from having solar cells mounted on the outside of cylinder-shaped satellites where the sun was “hidden” 40 % of the time to three-axis body-stabilized arrays has made a large difference. Today, the solar array can be constantly pointed toward the sun and even “angled” to get maximum illumination. Except when the satellite and its solar arrays are in Earth eclipse, the arrays are now deployed with great efficiency to soak up the most solar power that is possible.

The third major trend has been the increase in battery performance. There has been an increase in the “energy density” of battery systems and of their operational lifetimes. Batteries have increased in performance from Nickel-Cadmium batteries to Polymer Lithium Ion cells. The lifetime of batteries and solar cell arrays – both of which deteriorate in performance in orbit – are critical components in the ability of the spacecraft to continue to operate over long periods of time in the harsh environment of outer space. The power of communications satellites have increased from less than 100 W with the Early Bird (or Intelsat I satellite) to power systems that generate on the order of 15–25 kW. In short, overall output power performance has increased by a factor of some 150–250 times. These power systems also have much longer lives. These systems, however, are also much more massive. When one seeks to measure the net performance for satellite power subsystems in terms of watt per kilogram per year in orbit the net increase is more on the order of 20–50 times.

The future seems to promise further improvements. So-called rainbow solar cells, that have perhaps seven different PV junctions, may be able to achieve a net efficiency of 50–60 %. So-called quantum dot energy systems may also be able to achieve these levels of efficiencies at lower mass. These technologies are still in the laboratory or in prototype phase and thus not yet available to the commercial market (Ippolito and Pelton 2004).

The amount of solar power available in space is enormous if it could be effectively captured. One concept is to have some form of solar collector or concentrator to illuminate solar cells with higher intensity. There can be simple mirror surfaces associated with a solar array so that the cells can see the equivalent of two or three suns. There is, however, more advanced concepts that would deploy very lightweight film-covered collectors that could illuminate solar cells with great intensity. This technology is being developed in conjunction with plans to create solar power satellites so that solar cells could see the equivalent of hundreds of suns or even thousands of suns. This same technology, however, could be deployed in conjunction with a large-scale communications satellite platform. These large-scale and very low mass solar concentrators could be designed and deployed at much less cost than high performance solar cells (Fig. 3).[3]

---

[3] NASA Research on Light Weight Solar Concentrators at Marshall Space Flight Center and Goddard Research Center, nmp.nasa.gov/ds1/tech/scarlet.html also see http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20090026381_2009013967.pdf

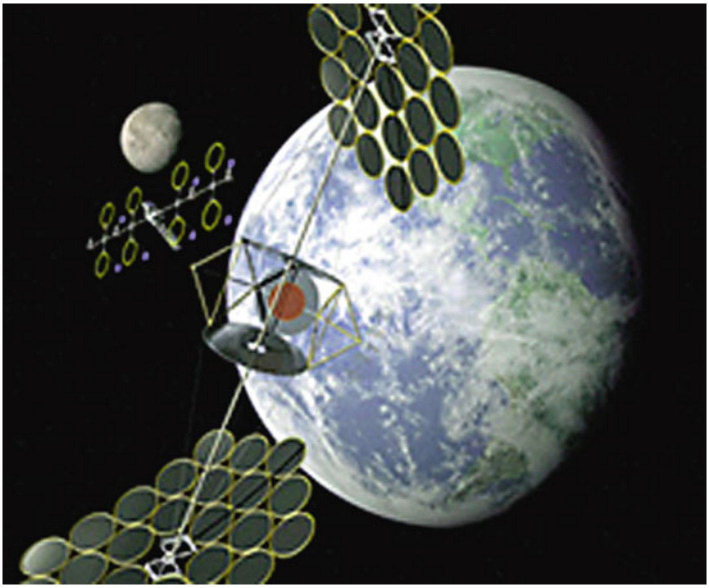


Fig. 3 Design concepts from NASA for lightweight solar concentrators (Photo courtesy of NASA)

There have been various studies to examine whether solar- and battery-powered satellites might "peak" in performance, in terms of maximum power and lifetime. These studies seek to compare large-scale solar arrays and battery systems versus other power sources such as nuclear energy, regenerative fuel cells, etc. These various studies have produced different optimization formulas and projected results, but many believe that for power systems above 25–40 kW nuclear energy or regenerative fuel cells might prove more cost-effective.

Certainly many space projects have used isotope-based SNAP generators for long-term and high-powered missions, but there is always the concern with safety (both at launch and reentry) with nuclear power sources. There are more ambitious longer-term research projects to develop nuclear reactors using thermal or ionized gases to provide propulsion systems. In these cases, nuclear power could also be applied to generate electrical power to support a variety of missions in the space applications field. Such systems are still under development by most of the major space agencies. There are development issues related not only to mass-to-power performance ratios and usable lifetime, but most importantly nuclear generators in space give rise to concerns of the safety of radioactive materials and their safe disposal.

There are, however, certainly technologies with great potential that do not involve the risks of nuclear power. Two of the most promising are regenerative fuel cells that can produce reliable power for extended periods of time with high mass-to-power ratios and are able to operate independently of sun exposure. Recent progress to use fuel cells for Earth-bound energy requirements for buildings and cars suggest that these may be effectively used in space within the next 10 years.

Even more immediate space-based energy systems involve those that create further efficiencies with solar power. One innovation is that of very lightweight solar concentrators that can be used not only with solar power satellites (SPS) but as a means to concentrate the equivalent of radiation of many suns on advanced solar cells or so-called quantum dots. Quantum dot technology which is still in the laboratory promises the ability to be perhaps three times more efficient in converting solar radiation into electrical power. A quantum dot can be defined as a special type of semiconductor whose “excitons” are confined from moving in any dimension. This constraint ends up giving quantum dot units properties that lay between those of conventional semiconductors and the behavior of an individual molecule. Nanotechnologists are seeking a wide range of applications for “quantum dots” in medical imaging, energy systems, and other fields.

Nearer term applications of quantum dot technology might involve using near infrared and perhaps higher frequency quantum dots as a retrofit to existing silicon solar cells to enhance performance.

More Effective and Reliable Spacecraft Design

The main drivers of satellite communications design and performance will likely be antenna design, increased power and digital encoding, and multiplexing techniques that will allow more throughput. Nevertheless, it is also important to have an effective design for the spacecraft to maintain reliable long-term operations and design features that can allow the satellite to be manufactured more quickly, at lower mass for lower cost launch or other sorts of improvements.

Satellite Orbital Configurations and Improved Spacecraft Orientation and Pointing

The last few decades in satellite communications development can be largely summarized by the following trends. These have been to make spacecraft systems more powerful and to deploy larger aperture antenna systems with higher and higher gain and the ability to reuse available spectrum through the interconnection of cellular-like spot beams. These innovations plus the ability to launch and deploy larger spacecraft in space plus the gains that digital communications and digital compression techniques have brought to the satellite communications industry have dominated gains in satellite capacity and allowed many new applications to evolve, particularly in fixed, mobile, and broadcasting services. Although there have been various types of smaller satellites – variously described as microsatellites, nanosatellites, picosatellites, etc. – these spacecraft today, despite innovative use of digital processing techniques, represent far less than 1 % of in-orbit capacity.

The future expansion of satellite system capacity and the ability to provide consumers with low cost transceivers was thought for some time to simply require the deployment of even larger aperture, multibeam antennas in space. The problem is that with current and projected launch capacity this may be quite difficult to achieve unless one starts to evolve toward one of several options. These would be: (a) the deployment of "parts" of systems that are either assembled in low Earth orbit and then flown into Geo orbit in order to create large-scale satellite platforms with antenna apertures in excess of 30 m; (b) the creation of "networked" antenna systems that fly in some form of formation or are cabled together to create a "virtual antenna system"; or (c) creation of a large-scale constellation operating in available spectra in the super high frequency (SHF) or the extremely high frequency (EHF) bands. However, in order to provide significant capacity via small satellites there is a need for a very large number of satellites in a constellation. This was the concept originally proposed some 20 years ago for the Teledesic satellite network. Some of the current concepts of this type for Internet-optimized constellations are now in the process of being implemented. These include the One Web and the SpaceX constellations. The problem with this type of approach to achieving a large amount of satellite capacity for Internet-optimized service is an increased risk of orbital debris collision and problems of interference with GEO satellite networks.

In spacecraft design there are always various forms of trade-off in terms of optimizing system capacity, lower cost and smaller user antennas, system lifetime, and strategies to cope with issues such as precipitation attenuation. One of the key constraints that would be posed by very huge antennas with apertures in excess of 30 m is that the pointing accuracy of the space antenna would need to be very precise. Consideration would need to be taken of such aspects as thermal expansion of the antenna due to exposure to solar radiation, etc. In these design trade-off considerations, it is clear that much higher frequencies in the EHF (i.e., 30 GHz and above) allow the space antennas to be smaller in size, but on the other hand the problems of precipitation attenuation (especially rain) becomes much more severe and greatly complicates keeping user antennas on the ground small in size and low in cost.

New Ground User Systems

The predominant trend in all forms of digital communications, information processing, and digital entertainment (i.e., the ICE industries, or the information; communications; and entertainment enterprises) have been to develop consumer-oriented, distributed systems that have moved closer and closer to the edge. This means that computers, entertainment systems, and communications devices have become smaller, more compact, more user-friendly, and lower in cost so that essentially all forms and types of consumers can own and operate these devices. Satellite communications have obviously followed this same trend. In the 1960s, computers were massive and highly expensive devices that required a team of experts to operate and were thus highly centralized. Satellite earth stations were much the same. A typical Intelsat Standard A Earth Station of that era cost in excess of $10 million (US), required a team of 50 or more people to operate, and involved the precise pointing of a 30 m (93 ft) antenna that weighed many tons. Today there are handheld computers that have much greater computational power than the first digital computers. Likewise handheld satellite transceivers sold directly to the consumer can communicate directly to in-orbit satellites. Consumer video games, computers, and digital satellite phones cost in the hundreds of dollars and require a minimum of training for consumers to use these products.

This trend toward miniaturization of user-friendly consumer devices and digital instruments that can be acquired at low cost will undoubtedly continue. This means that the trend will continue toward "wearable devices" that are even more compact. Assuming that health-related issues involving RF radiation can be successfully overcome, it seems possible that embedded communications devices that are capable of linking to in-orbit satellites may represent the next tier of development.

There are clearly a number of technical challenges to overcome. There continues to be a rapid population growth with some seven billion people on the Planet and everyday more and more people are seeking access to broader band communications to support entertainment, video messaging, and voice and data communications. Further, more and more people are seeking broadband mobile connections either via terrestrial mobile services or satellite communications connections. This expanding demand for broadband mobile services requires new solutions. These can be in the form of more efficient digital compression techniques, new ways to reuse available frequencies more intensively within terrestrial or satellite wireless systems, migration to higher frequency spectrum bands, or some combination of these solutions. Smaller consumer devices with smaller antenna size make all of these efforts more challenging. Application specific integrated circuit (ASIC) devices have contributed greatly to the ability to miniaturize consumer handheld communications devices, but new breakthroughs in quantum computing will be needed to achieve new levels of miniaturization. Such quantum computing-level breakthroughs may also help to reduce power requirements as well. This reduction of power would help not only with the problem of portable power supplies but also would assist with health-related concerns as well (Hagar 2007).

Integrated Satellite and Terrestrial Services and New Market Demand

Breakthroughs in quantum computing, nanomaterial engineering, and the next generation of ASIC technology can all help to further the development of future communications devices used in terrestrial mobile systems as well as satellite-based networks, but the key may also be found in the integration of terrestrial cable and mobile systems with satellite networks in new and innovative ways. In short, the key way forward, given frequency constraints and the continuing expansion of broadband services, would seem to involve integration of all forms of terrestrial and wireless links. This means that seamless interfaces between satellites and all other forms of telecommunications is key. In the future, satellites must be able to interconnect to terrestrial wireless for localized services, fiber optic systems for intrabuilding risers, urban-wide area networks, and long-distance "trunking interconnections" of all kinds. In this new world, satellites will be prime providers for broadcast and multicasting applications, for global, regional, and even some forms of localized mobile services, and for services in rural, remote, and island areas as well as for military and strategic applications. Satellites will not become obsolete, but rather will adapt to changing conditions.

The key to this type of integration of broadband services is seamless digital networking standards – most likely based in Internet Protocol standards. Terrestrial wireless, satellite communications networks and fiber optic cable systems all have their strengths and seamless interconnectivity allows each of these transmission media to be optimized. The consolidation of IP-based protocols on a global basis will allow an intellectual platform upon which this integration of various types of transmission systems will be increasingly possible. Likewise, the Internet and TCP/IP will allow for all the space-based application satellites to be integrated as applications that can be accessed on "smart phones," tablets, and other consumer devices.

Originally, satellite communication links, especially those to and from geosynchronous satellites, positioned almost a tenth of the way to the Moon at 35,870 km (or 22,230 miles) above the Earth's surface encountered major problems with communications operating via TCP/IP because of transmission latency or delay. The original design of TCP/IP was to interconnect computers on the Internet. Detected delays were considered to be the result of network congestion and the links automatically timed-out and thus went into recovery mode. Over time, a number of changes have been made to optimize satellite operation using TCP/IP – especially for GEO satellites with the greatest latency. In this case, special IP over Satellite (IPoS) standards employ "spoofing," reset of timers to accommodate satellite delay, and other techniques to allow satellite links to operate at increasingly high efficiencies. Also adjustments have been made to accommodate to IP Security (IPSec) and virtual private network (VPN) security measures when using satellite links. Part of the problem is that the Internet Engineering Task Force (IETF) and the International Telecommunication Union (ITU) have developed different methods to achieve efficient network interconnections. Part of the future challenge for the satellite industry is to be able to adapt to ITU and IETF standards and requirements rapidly, effectively, and economically.[4]

Telemetry, Tracking, Command, Monitoring, and Autonomous Operations

In coming years, it seems likely that satellite systems will continue to become larger and more sophisticated – either larger spacecraft or large constellation of small satellites. As this evolution continues, communications satellites will tend to assume

---

[4] Tech Republic White Paper, *IP Over Satellite*: *Optimization vs Acceleration End to End* (2010), whitepapers.techrepublic.com.com/abstract.aspx?docid

more complex roles with regard to onboard switching, onboard signal processing, and other functions that were once performed exclusively on the ground. This evolutionary movement, particularly toward onboard processing, will make satellites more capable and better able to cope with interbeam connection, rain attenuation, and other advanced functions. This trend toward greater complexity in space will make the role of telemetry, tracking, command, and monitoring more difficult. Especially the software and engineering needed to achieve rapid fault detection when there might be something like a 1,000 beams and millions of possible interbeam connections can become enormous. The most demanding role for the future in terms of designing advanced communications satellites will be the development of computer code to rapidly detect a particular fault in onboard link interconnections. For these future TTC&M roles, especially fault detection and detection of interference, there will be a trend to use artificial intelligence to assist with these functions. Likewise, there will be increasing efforts to apply computer programs and artificial intelligence to control as many of the operations of the satellite over its 10–18 years lifetime through what is called "autonomous operation."

Communications satellites operate 24/7 throughout the year and are maintained by a team of human engineers and technicians to monitor all of these operations and to engage in rapid fault or interference detection, tasks which are increasingly uneconomic. One of the greatest technical challenges of the future for satellite system design thus will not only be the development of complex multibeam antenna systems and onboard processing, but all of the new types of onboard intelligence that this will imply. Thus, it seems that the communications satellites of the future will have largely automated tracking, telemetry, command and monitoring and fault detection systems, and AI-based autonomous operations. This means that human-originated commands will be the exception and battery discharges, activation of redundant receivers, shutdown of nonfunctioning switches, recording of billing information, and hundreds of other operations that were once controlled by ground operators on a continuous basis will become activities that are increasingly assumed by onboard computers. Nevertheless, there will still be a need to monitor many satellite operations and maintain satellite operations centers. Autonomous operations and artificial intelligence will, however, serve to prevent these centers from growing exponentially in size and thus serving to make satellite operations uneconomic.[5]

Future Trends for Markets and Regulatory Systems

The future development of satellites is not dependent on technological development alone. Market demand and regulatory actions are likewise strong drivers of the satellite communications industry. Today's market trends suggest that new

---

[5] R. Sherwood, S. Chien, D. Tran, B. Cichy, R. Castano, A. Davies, G. Rabideau, *Next Generation Autonomous Operation on a Current Generation Satellite* trs-new.jpl.nasa.gov/dspace/bitstream/2014/7497/1/03-1398.pdf also see *JAXA Study A System Study for Satellite Operation and Control in Next Generation Systems* track.sfo.jaxa.jp/spaceops98/amp;/nfe_nakayama.html

growth will be shaped by demand, primarily in the three prime areas. These are: (a) entertainment and broadcasting services (i.e., video, high-definition and 3D and 4D digital television, plus broadcast radio that is also coupled with emergency vehicular services); (b) mobile communications services in areas not well covered by terrestrial wireless services (i.e., air, maritime, and remote land areas); and (c) gaps in communications services not well covered by fiber optic networks and terrestrial wireless broadband systems. Satellite broadband still remains key for Internet connections in many developing countries. New types of Internet-optimized satellite systems such as O3b (i.e., the Other Three Billion) and One Web are designed to bring broadband Internet and Voice over IP services to the parts of the world where effective terrestrial telephone and data networks are still lacking.

This is not to suggest that there may not be other market niches for satellite communications systems. Store-and-forward satellite systems that provide messaging services and business to business (B2B) services are also often tied to space navigational services for trucks, trains, buses, and ships. These "messaging" satellites represent one key type supplemental satellite service. One of the unknowns about future market demand relates to what might be called integrated space applications. In the world of "smart phone" applications, it seems increasingly likely that applications to support interactive navigation, immediate weather data updates, and remote sensing applications will evolve over time. Today remote sensing, Earth observation, meteorological and space navigation systems are delivered through separate space-based satellite systems and the provision of information from these type satellites are largely through separate and "stove-piped" telecommunications networks. In the coming years, these systems can and likely will be integrated via Internet linkages to become just additional applications available via handheld "smart devices" or ultimately maybe via embedded chip technology. All of these changes will serve to reshape the structure of the satellite communications industry. Mergers, acquisitions, and market integration via Internet applications will break-down traditional industry divisions. This will mean at one level that companies will integrate across transmission technologies such as fiber, cable television, terrestrial wireless, and satellites. On another level, companies in one space-related service such as satellite communications can and likely will be diversified into other space applications such as space navigation, space-based messaging, remote sensing, and real-time situational awareness.

In addition to market and service demand, regulations will also play a critical role. One of the most obvious areas will be that related to frequency allocations and the regulatory addition of new capabilities (i.e., new RF allocations that could interfere with satellites). This is already the case in terms of interference between telecommunications satellites and high altitude platform systems (HAPS) operating in the Ku-band. Such HAPS may be deployed over urban areas to provide television, remote sensing, or wireless broadband communications. Today one of the major constraints to the expansion of satellite communications services involves the lack of available spectrum and the lack of new orbital locations in GEO orbit for new satellite communications networks.

The demand for increasingly broadband services and expanded terrestrial wireless services to support mobile applications will only exacerbate this problem. This will push technology to develop larger multibeam satellite systems to increase frequency reuse and improved ASIC (Application Specific Integrated Circuits) transceivers to operate more effectively in this spectrum-limited environment. Today the International Telecommunication Union (ITU) plays a key role in the allocation of spectrum and the establishing of recommendations to limit intersystem interference. The ITU is limited in role in many ways. Nations often establish through footnotes limitations on frequency spectrum allocations within their national borders. The ITU has no special enforcement powers such as fines or penalties for those who do not fully implement its recommendations. In the coming decades, this lack of enforcement powers and saturation of orbital locations and spectrum shortages could limit the growth of the satellite industry in a serious way.

Another serious regulatory issue involves the increasing spread of orbital debris. This spread is particularly troubling in the low earth orbit, but increasingly it is also a concern in the medium earth orbit and the GEO orbit as well. Controls designed through the UN Committee on the Peaceful Uses of Outer Space and ITU and the Inter-Agency space Debris Coordination (IADC) Committee, plus national efforts at due diligence in these areas, are starting to have some favorable impact, but the continued deployment of new systems may also make this issue a potential brake on the future development of the satellite communications industry and other space applications. New so-called MegaLEO constellations in LEO orbit are a particular concern in this regard.

With the latest high-definition tracking systems some 22,000 objects the size of human fist can be tracked in the Earth Orbit. The hope is that improved due diligence efforts to eliminate sources of orbital debris and continued decaying of materials that drop from low earth orbit can bring the problem of "space junk" under control before cascading effects from orbital collisions can create a blizzard of hazardous materials in space.

There are currently efforts to create a global database to track the orbits of satellites starting with the GEO ring where the most communications satellites are currently in operation. To date, Intelsat, SES Global, and Inmarsat have agreed to input data and Echostar, Telesat, and Eutelsat have indicated intent to participate (Chan and DalBello 2010).

Other Drivers and Opportunities

The future of satellite communications involves more than just spacecraft, launchers, communications equipment, and new telecommunications equipment. The future success of the industry will require financial markets that will provide the capital to acquire new systems and technology. It will require a flexible insurance industry that can allow risks of various kinds posed by telecommunications markets and system failures to be overcome. It will also need an educational system that will produce the future engineers, financial analysts, regulators, and business people who can steer the industry forward.

Sometimes unsuccessful satellite ventures such as Orbcomm, Iridium, ICO, Globalstar, and Teledesic help to shape a more viable satellite market several decades later. This can be achieved by providing key data as to market demand, technological challenges to be overcome or better ways to achieve frequency efficiencies, interference reduction, or more appealing consumer products. Today, launch systems remain one of the key obstacles to satellite communications' economic efficiencies. Insurance and risk management costs often represent over 10 % of communications satellite system costs. The lack of new frequency allocations, orbital congestion, and orbital debris can today all serve as major barriers to industry growth. In the future, environmental concerns could also serve as a brake to new system deployment. It is key to recognize that these other type factors can pose a "limit to new growth" when considering the future of the satellite communications industry.

New Initiatives in Space Communications

Satellite communications technology and operations represent a still very rapidly evolving field. The telecommunications satellite industry is not only driven by a wide range of technologies to develop improved space systems and user equipment but also by competing communications systems – particularly fiber optic networking. Other technologies are also driving the curve. The future of satellite communications will thus also be shaped by artificial intelligence, robotics, terrestrial wireless systems, high altitude platform systems, quantum computing, laser communications, new multiplexing systems, and a host of other technologies. The future of humankind is ultimately based in space and thus space communications will one way or another continue to be a part of that future. If one considers the development of satellite communications in terms of throughput capabilities, power, antenna gain, lifetime, and costs, it has improved by a factor of over a 1,000 times in the last half century and there is sufficient technology in the pipeline that it could improve another 1,000 times in another 50 years. The key to considering the future, however, is not by projecting the rate of technical innovation but to seek to understand basic market trends and to interpret what kinds of applications people will need to meet future societal, environmental, and economic needs in the decades ahead. Totally new markets could well evolve.

Keys to the Future of Satellite Communications

Predicting the future is always difficult but there are clear indicators as to trends and opportunities. The following "keys" are currently observable.

Off-World Communications

The most uniquely suited new market for satellite communications seems likely to be cislunar communications to Moon-based colonies as well as links to Mars or inhabited asteroids, artificial space colonies, or the satellites orbiting other planetary bodies in the solar system. The technology to support communications in support of today's scientific satellites already probing the Solar System will provide a head start in this direction. The most logical extension of capability in this respect would seem to be laser-based communications since the lack of an atmosphere on the Moon, asteroids, or the satellites of other planets makes light communications quite viable over the great distances of the Solar System. Laser beams are highly focused and thus are much less subject to path loss due to spreading. When the transmission distances are millions of miles the ability to reduce path loss is crucial.

Light attenuation within the Earth's atmosphere would suggest that laser communications would be directed toward Earth orbiting satellites perhaps in geosynchronous orbit. There have been other suggestions as to how to most efficiently establish such links. One suggestion is to have a solar sail-oriented satellite that could be positioned or "levitated" above one or both of the poles so that a signal could be relayed directly to anywhere in the Northern (or Southern) latitudes. The various space agencies, and NASA in particular, have invested a good deal of research as well as state-of-the-art space communications hardware in intraplanetary and even interplanetary relays.

Smart Satellites and Advanced Encoding

The rapid advance in computer technology that was predicted by the so-called Moore's Law anticipated a doubling in capacity every 18 months and, in general, this exponential growth in computer performance has continued for some 30 years. The satellite communications industry has followed a similar curve of accelerated performance. This has been particularly true since the conversion from analog to digital satellite communications. In essence, today's communication satellites are, in fact, "software defined hardware." Although these are elaborate devices designed to operate in the harsh environment of space, the communications function is today essentially the result of very fast processing of digitally encoded information.

Advanced coding capabilities such as "turbo coding" allow the digital processing to be more and more efficient. Just a few years ago the most efficient communications satellites could process about 1 bit of information per 1 Hz of available spectrum. Today with more efficient coder/decoders (i.e., codecs) and more efficient modulator/demodulators (i.e., modems) efficiencies of 2.5 up to 5 bits per Hz of available spectrum are achievable. In future years, the efficiencies of digital satellite communications with onboard processing and improved codecs may be able to achieve even higher efficiencies.

The limit on performance that derives from more efficient encoding of information is strongly determined by the amount of noise or system interference. This is most precisely defined in digital communications as "bit error rate" (ber). Efficiency is achieved by the use of more and more efficient codes. Particularly in the case of heavy rain and associated rain attenuation of signal clarity one might use something like 4 bit encoding. In such conditions, the use of even higher efficiency 8 bit or 16 bit encoding is not possible with today's communications satellites. However, in clear sky conditions and with other forms of interference not present, much more efficient coding is indeed possible. In the future, there is likely to be "onboard processing" that can restore up-linked signals to pristine quality. This ability to process signals onboard the satellite and again as they are received as down-linked transmissions can allow the transmission efficiency to rise.

It is significant to note that satellite systems and broadband wireless technologies are most keenly focused on processing and encoding efficiencies, as opposed to fiber optic networking systems. This is because in the fiber world, there is almost limitless spectrum with the use of dense wave division multiplexing (DWDM) and close to zero bit error rate (i.e., virtually no meaningful interference). Under these conditions, there are no strong incentives for the development of high efficiency encoding and increased bit/Hz throughput. With virtually unlimited available spectrum for fiber optic networks, there is little incentive for more and more efficient use of available bandwidth. In the world of satellites and broadband wireless communications, of course, the reverse situation holds true.

Integrated Satellite and Terrestrial Networks (i.e., The "Pelton Merge")

The world of communications today is driven toward integrating different types of transmission networks. The top objective is thus to combine "seamlessly" different transmission media to serve a wide range of consumer needs. The convenience of mobility drives demand for broadband wireless services, including communications satellites and in the future high altitude platform systems (HAPS) and unmanned autonomous vehicles (UAVs) that provide platforms for communications and broadcast services. On the other hand, fiber optic and coaxial cables can support very high efficiency and cost-effective services to fixed locations – particularly when heavy routes of traffic are involved. These divergent communications requirements lead to the need for operational and technical standards to allow these wire and wireless networks to link easily, at low cost and with high quality. This is what is meant by a "seamless connection."

This objective is easy to define and understand. Yet the ability to achieve "seamless" interconnection remains difficult. The difficulty stems from three key factors:

1. The world of fiber and coaxial cable multiplex signals in the wave division domain because of the vast amount of spectrum this provides for broadband services, while wireless services, including satellites, operate in the time domain. This is because of the cellular type frequency reuse required to divide spectrum

into small cells in order to allow multiple reuse of available spectrum. The interconnection of the signal used in these various cells requires digital processing time and thus multiplexing in the time domain. Interconnection is possible, but the differences between satellite and terrestrial fiber networks add cost, complexity, and technical challenges in terms of smooth interconnection.

2. The second complicating factor between the world of fiber and satellites is related to the need to use high efficiency codecs to pack more information into available spectrum in the world of wireless services – as discussed above. Again the technical differences make "seamless interconnection" more difficult and thus more expensive.
3. The third factor comes from the world of Internet that was initially designed to run on terrestrially "wired" networks such as local area networks (LANs). As noted earlier, in the original design, any significant delay in transmission was assumed to be network congestion rather than the transmission delay associated with geosynchronous satellites. Other aspects of the design such as IP Security (IP SEC) architecture were not designed to accommodate the transmission architecture of satellites. Initially where IP-based traffic was routed over satellites these problems tended to make satellite transmissions very inefficient and secure virtual private networks (VPNs) difficult to establish because the IP SEC process of "stripping off of header" information to preserve privacy led to confusion in satellite routing.

Over time, new standards related to IP over Satellite (IPoS) were developed and these issues of compatibility between satellite transmission and IP-based traffic were largely resolved by resetting "clocks" to accommodate satellite transmission delay and by other measures designed to accommodate to IP SEC requirements associated with VPNs.

In the future, the design of compatible terrestrial and satellite networks to accommodate IP-based traffic will be a major challenge to make wire and wireless networks fully compatible, cost-effective, and of high quality. It is recognized by most planners today that a combination of wire and wireless systems will be needed to accommodate diverse networking and broadcasting needs and especially the demand for mobility. The achievement of the so-called Pelton Merge still remains a difficult-to-achieve goal that is dependent on improved interface standards optimized to meet the needs of IP-based traffic, low cost fiber optic cable networking design, and broadband mobile services provided via wireless and satellite systems.

Advanced Launch Capabilities, In-Orbit Servicing, and Advanced Platforms

For years, the major brake on the development of the satellite communications industry has been with regard to cost and reliability of launch services. Although satellites have increased in capacity, power, lifetime, and cost efficiency, launch vehicle costs in terms of kilograms of payload to orbit have remained rather static.

Recently, new commercial lift systems developed by SpaceX, Virgin Galactic, and others seem to offer shorter term hope for lower launch costs. Also systems developed by India and China have served to provide some new economies as well.

There is also hope that breakthrough technology in terms of tether lift systems, advanced ion engines, and electrical propulsion or even space elevators might in future years increase not only the reliability of launch systems, but also dramatically decrease launch costs. There has also been increased R&D and now even new space tugs that are able to move satellites from LEO to GEO orbit or perhaps soon provide retrofit and repair to satellites that require new batteries, refueling, or even new satellite antennas or other components that have failed or are inadequate on in-orbit satellites. These on-orbit capabilities are discussed later in the Handbook. Such improved launch systems or space tugs could, of course, hasten the design and deployment of large-scale satellite platforms or satellite clusters that could provide greatly expanded communications capability to orbit as well.[6]

Space Safety and Orbital Debris

For years, the focus of space safety systems and related technology was on human space flight. In recent years, however, there has been an increasing focus on how to make space activities in all its forms safer and the peaceful uses of outer space better guaranteed. The UN voluntary guidelines with regard to space debris are one important step in this direction. The recent UN COPUOS initiative on the "Long-Term Sustainability of Space" is aimed at addressing this issue in broader terms that go beyond space debris, and thus the effort is seeking ways to keep space from being militarized and finding new ways to ensure that access to space by all nations can better be assured.

Conclusion

The nature of satellite communications, in terms of technology, operations, institutions, finance, and markets, has changed greatly in the last 50 years. Satellite communications is by far the most successful of satellite applications, at least as measured by annual income. Part of the success of the satellite communications industry has been its constant evolution to find new markets. Today "submarkets" of satellite communications include fixed satellite services; direct to the home television satellite services; broadcast satellite services (for both television and audio); land, aeronautical, and maritime mobile satellite services; and store-and-forward or machine-to-machine satellite services. One could even suggest that satellite navigation services constitute a form of satellite communications. This constant expansion

---

[6] Op cit, Iida and Pelton.

of such “submarkets” has served to expand global revenues in the field for nearly a half century.

The improvement of the technology has, of course, likewise expanded the success and the reach of satellite communications. So-called technology inversion that has allowed more powerful, capable, and longer-lived communications to work to ever smaller, more compact, and mobile user terminals at lower cost has expanded the reach of satellite communications. Instead of just a few large earth stations operating in the realm of satellite communications, there are today literally millions of modest to quite small ground systems (including handheld units) supporting a rich array of services. There is no reason to expect that technology innovation will slow as digital satellite communications systems, with increased levels of “intelligence,” are deployed in future years. This chapter has examined a number of technologies of the future that might be anticipated or even are being deployed today.

Finally, competition, innovation, and flexibility of institutional arrangements have allowed satellite communications markets and technology to grow and expand. This competitive environment has led to an active pattern of change. This change has seen – and continues to see – mergers, acquisitions, evolution of new entities, and new arrangements for financing and risk management in the field. This openly competitive environment has tended at times to give birth to new problems such as contending with orbital debris; interference between satellites, particularly in the geosynchronous (or Clarke) orbit; difficulties related to assignment of orbital locations; and the need for additional support to satellite communications in developing economies. One hopes that technical and operational innovation – and perhaps new economic and institutional arrangements – can provide required answers to such challenges in future decades.

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Economics and Financing of Communications Satellites
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ History of Satellite Communications
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Regulatory Process for Communications Satellite Frequency Allocations
- ▶ Satellite Antenna Systems Design and Implementation Around the World
- ▶ Satellite Communications and Space Telecommunication Frequencies
- ▶ Satellite Communications Antenna Concepts and Engineering
- ▶ Satellite Communications Modulation and Multiplexing
- ▶ Satellite Communications: Regulatory, Legal, and Trade Issues
- ▶ Satellite Communications Video Markets: Dynamics and Trends
- ▶ Satellite Earth Station Antenna Systems and System Design
- ▶ Satellite Orbits for Communications Satellites
- ▶ Satellite Radio Communications Fundamentals and Link Budgets

- ▶ Satellite Transmission, Reception, and Onboard Processing, Signaling, and Switching
- ▶ Space Telecommunications Services and Applications
- ▶ Technical Challenges of Integration of Space and Terrestrial Systems

References

J. Chan, R. DalBello, Data sharing to improve close approach monitoring and safety of flight, in *Space Safety Regulations and Standards*, ed. by J.N. Pelton, R. Jakhu (Elsevier, New York, 2010)

European Space Agency, Interplanetary communications action 1: requirements analysis (2008), telecom.esa.int/telecom/www/object/index.cfm?fobjectid=29325.

A. Hagar, Quantum computing, in *Stanford Encyclopedia of Philosophy* (2007), stanford.library.usyd.edu.au/archives/win2007/new.html

C. Hoeber, Annual satellite update, in *Aerospace America Magazine*, Sep 2014, 46

T. Iida, J.N. Pelton, The next thirty years, in *Satellite Communications in the 21st Century: Trends and Technologies*, ed. by T. Iida, J.N. Pelton, E. Ashford (American Institute of Aeronautics and Astronautics, Reston, 2003), pp. 188–190

T. Iida, Y. Suzuki, Satellite communications R&D for the next 30 years, in *Proceedings of the 19th AIAA International Communications Satellite Systems Conference*, Toulouse, vol 233, Apr 2001

L. Ippolito, J.N. Pelton, Satellite technology: the evolution of satellite systems and fixed satellite services, in *Communications Satellites: Global Change Agents*, ed. by J.N. Pelton, R.J. Oslund, P. Marshall (LEA Associates, Mahwah, 2004)

J.N. Pelton, Telecommunications for the 21st century. Sci. Am. **278**(4), 80–85 (1998)

J.N. Pelton, *Basics of Satellite Communications*, 2nd edn. (International Engineering Consortium, Chicago, 2006), pp. 1–10

D.K. Sachdev, *Business Strategies for Satellite Systems* (Artech House, Norwood, 2004), pp. 1–15

S.P. Schnaars, *Megamistakes: Forecasting and the Myth of Rapid Technological Change* (Free Press, New York, 1989)

H. Wakana, Mobile service update, in *Satellite Communications in the 21st Century: Trends and Technologies*, ed. by T. Iida, J.N. Pelton, E. Ashford (American Institute of Aeronautics and Astronautics, Reston, 2003)

Future of Military Satellite Systems

Joseph N. Pelton

Contents

Abstract

The use of space systems to support military activities and enhance defense-related capabilities has increased exponentially since their first application in 1965 with the Initial Defense Satellite Communications System. Although the first major application was for communications services, space-based defense capabilities have now expanded to provide a wide range of other types of services. Today these applications include navigation, targeting, mapping, remote sensing, surveillance and meteorological tracking, and prediction. In short, 50 years of expanded space-based capabilities for military and defense-related

---

J.N. Pelton (✉)
International Space University Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_107

services seem destined to be followed by 50 years of even greater capabilities. Thus, there will be expanded competence in terms of new types of space hardware and new applications. Further, entirely new capabilities will be added. These will likely include expanded use of artificial intelligence and increased focus on cybersecurity and space situational awareness, on-orbit servicing, and perhaps even active orbital debris removal.

This chapter examines all of these trends and discusses whether some of these trends will relate to improved commercial satellite capabilities, particularly in the context of dual use of commercial networks and hosted payload systems.

Keywords

Artificial intelligence • Cybersecurity • Dual use • Fiber optic satellite ring • Global Information Grid (GIG) • Global Positioning System (GPS) • Hyperspectral Imaging • Internetworking • Internet of Things • Intersatellite Links (ISLs) • Laser Light Communications • Meteorological Satellites • NATO • Orbital space debris • Satellite Services to the Edge • Space Situational Awareness • Transformational Satellite System

Introduction

In 1965, the Initial Defense Satellite Communication System was launched as one of the three first Satcom networks ever launched into orbit around planet earth. Intelsat and the Soviet Molniya systems for domestic communications were the other two launched in 1965. Since that time, defense satellite networks have played a significant and ever-growing role in military operations and activities. These defense-related space systems have expanded from telecommunications and broadcasting to an ever-expanding role. Today, tremendously more capable satellites are utilized by defense forces not only for communications but also for weather monitoring and forecasting; for earth observation, remote sensing, and surveillance; and for navigation, mapping, and targeting. Satellite communications in support of tactical military communications now provide a wide range of broadband communications and terrain-based geo-location services to extremely small micro-terminals and handheld devices. This allows broadband and video-based services to be delivered right to the battlefield edge. These space-based systems can provide not only "communications" services but also up-to-date weather forecast, topographical maps, and information about the location of potentially hostile forces. The most current systems can also collect information and provide command and control for drones, unmanned aerial systems (UAS), and high altitude platform systems (HAPS).

These types of highly capable satellite systems blanket the world and are interconnected with terrestrial and wireless radio networks. Such sophisticated space-based systems, with all these capabilities, are typified by the US-based Global Information Grid (GIG), but parallel capabilities are also being developed in Europe, Japan, and even in other countries such as China, India, and Russia, albeit not currently to the level represented by the GIG.

And current space-based capabilities do not stop there. There are even sensors in satellites such as the US- deployed Global Positioning System (GPS) system that are designed to monitor nuclear explosions. There are also networks to maintain situational awareness in space to detect rocket launches and to monitor orbital debris. Defense space-based systems are often designed with onboard protective systems against radiation and electromagnetic pulses as well as armor against micrometeorites and other possible hazards. Thus, one of the more recent trends has been to utilize commercial satellite systems, a process that is called dual use, to support a number of military or defense services such as to provide television and radio programs to overseas troops or video imaging from drones. These commercially based satellite services are typically lower in cost, since commercial systems are not radiation-hardened or involve elaborately protected hardware, and can reasonably be used for non-tactical purposes.

The future will, at one level, be more of the same, only better in terms of higher throughput, smaller and more mobile systems, higher resolution imaging, and better meteorological systems and navigation accuracy. At another level, there will be true innovations in both the hardware and the applications/services. The hardware will likely include optical communications intersatellite links (ISLs) in space as well as so-called megaLEO satellite networks in space that will be optimized for Internet Protocol services. Even broader band services linked to artificial intelligent networks can not only be linked to ground command and control centers, troops, and hardware but also to robotic and cybernetic units programmed for a number of tasks that range from disaster relief and recovery to actual war-fighting operations. One of the continuing concerns with regard to all of these innovations will be cybersecurity and the possible “infecting” of artificially intelligent systems to the extent that there could be a loss of control of various classes of weapons systems and robotic systems.

Finally there will be ongoing concerns with regard to space situational awareness of all types of potential space weapons systems There are also continuing and indeed growing concerns about orbital debris that could become an even larger problem with the deployment of large-scale LEO constellations, particularly in the 500–1200 km altitude range and especially in the polar regions.

Advanced Satellite Capabilities

One of the most recent developments in satellite communications is the advent of high-throughput satellites (HTS) that are 10–50 times more capable than satellites of only a few years ago. These satellites are not only deployed for commercial systems but in defense systems as well. Each one of the six US Defense Department’s Wideband Global Satellites (WGS), for instance, has nearly 5 Gb/s of switchable throughput capacity that is available in both the X-band and Ka-band. Each of the WGS satellites represents, for instance, 12 times the capacity of a DSCS III military satellite that was deployed only a few years ago. Although this is a US system, it has direct participation by Australia, Canada, Denmark, Luxembourg, the Netherlands, and New Zealand which are all equipped with terminals that can connect to these WGS satellites (Wideband Global Satcom Satellite 2015).

This six satellite wideband system is able to provide broadband services to armed forces around the world and to small highly mobile units. The very flexible XTAR system that operates in the X-band is also highly reconfigurable to provide services to conflict areas on land, littoral, or ocean areas on demand. It is likely that new satellite defense systems that are deployed by the USA, Europe, Japan, and other countries will deploy new satellite hardware that has the following characteristics: (i) increasing broadband capabilities delivered to ever smaller and mobile terminals at the “edge” of networks in the field, (ii) networks that are optimized for Internet services and designed to have lower transmission latency, and (iii) systems that are optimized to connect to an optical ring of satellites that can support broader band and lower latency connections.

Networking to the Edge

During past conflicts, the headquarters and command centers in overseas locations had key information from the surveillance satellites, weather satellites, or navigation and positioning satellites, but there were frequent problems in communicating that information directly to the front where hostilities were being waged. With satellite systems such as XTAR and WGS and with the dual use of commercial satellite systems such as Iridium, Globalstar, and Inmarsat, it is no longer true that there is limited defense-related information provided to the “edge” of conflict zones. The increased capability of new defense satellite systems will only add to the capability to provide broadband, video, and imaging data to the front in future years. Today, this capability is limited to the most technologically advanced defense networks, but in the future, this type of capability will become more common around the world. Dual use of the most advanced commercial satellite systems will also augment systems such as WGS and XTAR and allow ever smaller micro-terminals and handheld devices to be used.

Internet-Optimized Satellites

Today in the commercial world, new satellite networks involving very large low earth orbit (LEO) networks are being planned by commercial entities. These networks such as One Web and SpaceX are sometimes referred to as megaLEO constellations. The concept is to provide low-cost broadband services in the developing countries in the more underserved parts of the world in the equatorial regions and to also provide much lower latency transmission speeds. The orbits that are being planned for such networks in the 500–1000 km range are 20–40 times closer to the surface of the world and thus for a ground station to satellite to ground station hop can involve transmission times that are 40–80 times faster than a link from ground to GEO orbit and back to the ground.

The first such systems, like o3b, One Web, and SpaceX, are designed to support commercial service requirements, but such new LEO constellations might also be used for military/dual-use applications. The commercial spectrum and the military frequency allocations are, of course, different, but there can be crossover use in dual-use systems, and these applications can be expected to continue in the future years. As all forms of communications and networking move to be more and more Internet Protocol based and global in nature, there is interest in and desire to use satellites that are Internet friendly and thus much faster in end-to-end transmissions.

There are many challenges here. These challenges include finding clever ways to minimize interference between LEO constellations and GEO satellite networks, avoiding collisions with orbital debris that has not been removed from orbit, and creating systems that are fully reliable and cost effective. The satellite networks can be made to be reasonably low cost, but the user terminals on the ground must be able to receive signals effectively from satellites that are passing rapidly overhead. This means installing relatively expensive ground systems that can track the overhead satellites or using relatively low gain terminals that can continuously receive the signal via tracking or by other means yet to be clearly understood and certainly not yet developed.

Another way of explaining this is that it is likely that the development of new technology to support low-cost ground systems that can operate effectively and reliably to low earth orbit constellations is a very large challenge. Large-scale constellations that would involve very rapid transition from beam to beam and even satellite to satellite constitute a major reliability and continuity of service issue. The Iridium and Globalstar LEO constellations that involved the switching of beam to beam once a minute, and from satellite to satellite about every 8–10 min, posed a very difficult challenge in terms of smooth and reliable transitions.

The creation of large-scale LEO constellations with a 1000–4000 satellites will pose major concerns in a number of areas. These concerns include reliable switching between beams and satellites as these satellites move into view and then disappear over the horizon; significant challenges in the design, performance, and reasonably low cost of user transceivers; and finally with regard to minimizing interference between such LEO and MEO satellites and those satellites in geosynchronous orbit.

Optical Intersatellite Links and the GIG

There is another approach to creating a military defense satellite network that could be optimized for Internet Protocol service. This could be accomplished by reducing satellite transmission paths and thus minimizing latency. There was a great deal of engineering and design effort invested in the concept of creating an optical ring of satellites in medium earth orbit. Such a new broadband infrastructure was envisioned by the US defense communications system under the name of the "transformational satellite network."

The concept of the Transformational Satellite System (TSAT) was to provide orbit-to-ground laser communications. Throughput for the five-satellite constellation as initially envisioned could be up to 40 Gb/s. The problem was in getting funding for the estimated total program, which at the time of its cancelation was estimated to have cost of up to $26 billion for the entire constellation.

The so-called TSAT program was at one point seen as key to global net-centric operations and as such as the next stage in the development of space-based systems to support the Global Information Grid (GIG). As envisioned, the five-satellite constellation ring in medium earth orbit would allow the GIG to accommodate broadband users without terrestrial connections. In short, the objective was to achieve improved connectivity and data transfer capability and greatly increase the quality and throughput of satellite communications for the US and allied warfighters. Secretary of Defense Robert Gates, responding to US Presidential guidance to curtail large-scale new military programs, canceled this ambitious program in the US military budget request for fiscal year 2010 (Brinton 2009).

The explicitly stated goal of the Transformational Satellite System before it was canceled was "to provide improved, survivable, jam-resistant, worldwide, secure and general purpose communications as part of an independent but interoperable set of space-based systems." TSAT was thus conceived as being able to replace the US Department of Defense current DSCS satellite system and supplement its Advanced Extremely High Frequency (AEHF) satellites. The network was also seen as being able to support NASA civilian space program communications as well (Transformational SATCOM) (Fig. 1).

Fig. 1 The Boeing design for the now canceled TSAT optical communications satellite network

The idea of an optical ring that could also support optical links from space to earth and earth to space, although canceled as part of the US military budget, is still very much alive and well in the current planning of a new commercial laser ring satellite network as envisioned in Laser Light Communications (addressed in chapter "▶ New Millimeter, Terahertz, and Light-Wave Frequencies for Satellite Communications" on Millimeter Wave and Light Communications systems). Currently this commercial laser light communications system is being presented as a civilian commercial communications satellite network and not as a network to support military or defense links. It is significant to note that while there are separate radio waves (i.e., frequency spectrum allocations) for civilian communications and defense/military purposes, there is currently no process to allocate light wave spectrum for any type of allocation. This means that combining military and civilian/commercial communications services for optical-based laser services would be fully authorized and in no way restricted under current regulations for frequency allocations.

The Future of Dual Use and Hosted Payloads

This provides a good segue into the issue of dual use of commercial and defense satellite networks moving into future decades of satellite planning. Optical fiber networks on the ground and optical laser communications via satellite constellations in the future can be expected to support combined commercial and defense communications links as "dual-use" networks and become technologically expedient and economically efficient. The migration to very high radio frequency and optical spectrum, as noted above, could only serve to accelerate this current trend.

The other trend of deploying Internet-optimized satellite networks and especially the deployment of larger-scale satellite constellations also would seem to open the door to more opportunity to have hosted payloads. These could be designed and engineered to be added as supplemental payloads on constellations in low and medium earth orbit satellites networks.

Large GEO satellites have sufficient mass margin to host experimental packages for next-generation satellites, and thus Intelsat satellites have hosted such payloads as the Internet Router in Space experiment, as designed by Cisco (Cisco's Space Router 2010).

Large-scale constellations such as the Iridium NEXT generation are, on the other hand, able to host scores of secondary packages that provide complete global coverage. The second generation from Iridium system Aireon LLC, a joint venture with between Iridium and Canada's air traffic agency NAV CANADA with support from the US Federal Aviation Administration (FAA) and suppliers Harris Corporation and ITT Exelis, will use hosted payloads to provide space-based monitoring and control of aircraft. Thus, the current plan is to include, as hosted payload on Iridium Next spacecraft, space-qualified Automatic Dependent Surveillance-Broadcast (ADS-B) receivers. These will be built into each of the 81 satellites in Iridium's NEXT spacecraft to provide fully global and continuous space-based monitoring and control of aircraft, even over oceans and remote regions where it is not currently possible. The payloads are based on Harris Corporations's AppStar reconfigurable platform. In addition, 58 of the satellites will carry an AIS (Automatic Identification System) payload for tracking of maritime traffic exact earth coordinates. These exactView-RT payloads allow for real-time ship tracking data with revisit times of 1 min. The payloads are based on Harris Corporations's AppStar reconfigurable platform (Iridium Next) (Fig. 2).

The two examples of Iris on Intelsat 22 and the hosted payloads on Iridium Next that derive from the commercial satellite world have already been demonstrated in the defense satellite world as well. In short, hosted payloads are a growing trend in both the commercial and military/defense worlds. In some cases, military payloads might indeed fly on commercial satellites. Such "dual use" at the hardware level rather than just combined use still allows economies and a level of protection as being on a commercial carrier rather than just being a separate piece of military hardware.

Intelsat and Boeing have in fact teamed to fly a defense-related payload for the Australian military. In particular, they cooperated to build and install a 20-channel UHF-hosted payload on the Intelsat-22 spacecraft launched in 2012. This UHF package now gives the Australian Defence Forces (ADF) continuity of service and augmentation of their UHF capacity. The ADF has estimated that over the lifetime of the Intelsat 22, it will save $150 million by using a hosted payload rather than developing and launching a free-flyer UHF satellite as an independent project (Hosted payloads) (Fig. 3).

The US Department of Defense, in order to maintain communications in this region, has also contracted with the Australian Defence Forces for 10 of these 20 channels (DoD, Australia 2012).

Fig. 2 Iridium next satellite platform will carry at least two hosted payloads

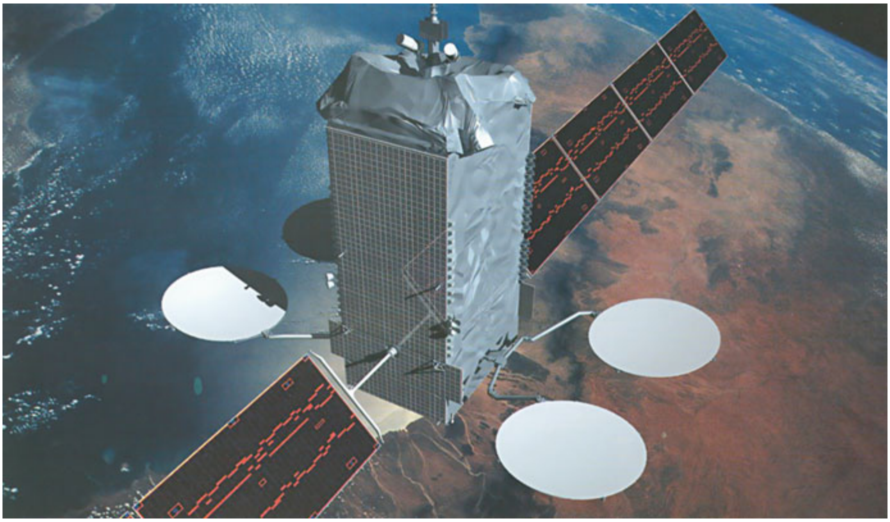


Fig. 3 Intelsat 22 with hosted UHF payload supporting military communications for Australia and United States (Graphic courtesy of Intelsat)

The chapter on hosted payloads will provide greater detail about currently planned initiatives for hosted payloads that are being actively pursued in military satellite system planning in a number of countries. This concept seems to provide a triple advantage in terms of cost economies. This is to say that such economies can be realized at three levels: (i) spacecraft design and manufacture, (ii) launch cost savings, and (iii) perhaps most significantly in terms of operational cost savings. The problem is that such hosted payload programs for critical military or defense programs may not have a recovery strategy if the satellites hosting such payloads should experience a significant loss of power, stabilization/pointing capability, or relevant transponders. In short, if the satellite system serving as host fails or has insufficient power to support the hosted platform, what is the backup plan? Hosted payloads have generally been successful to date. The backup plan and resilience questions will come to the fore when unexpected failures occur. Certainly one would be wise to invest some of the savings that can come from hosted payload operations in restoration capabilities.

Cybersecurity

The design of the interconnected Global Information Grid is based on the ability of all of its component parts to be free of cyber intercept and thus cybersecurity of its satellite components – which are potentially vulnerable to cyber attack. The Global Information Grid (GIG) as now envisioned is an all-encompassing communications project of the US defense forces, but on occasion it is also interconnected to NATO and its information and satellite networks.

The GIG is defined as a “globally interconnected, end-to-end set of information capabilities for collecting, processing, storing, disseminating, and managing information on demand to warfighters, policy makers, and support personnel” (Management of the Department of Defense Information Enterprise 2009). This strategic network is, of course, encrypted to protect it against unauthorized access. And it is the wireless and satellite-based part of this network that is most vulnerable to attack. Today there is a significant buildup around the world of cybersecurity efforts in most of the major countries in the world. As a result of cyber attacks against civilian, business, governmental, and defense sites that have increased over time, the United States and China are seeking a formal agreement to not engage in cyberspace intrusions against each other. Experts have speculated that if such a Sino-US agreement could be reached, it might become the basis of a network of agreements around the world (Sanger 2015). As a result of President Xi’s State visit to Washington, D.C. on September 24 and 25, 2015, high-level agreement on such a cybersecurity arrangement was reached in theory. Discussion of implementation details involving cabinet-level coordinative processes is now going forward. This type of trendsetting agreement is designed to protect civilian, business, governmental, and military-related websites, but its effectiveness and longer-term viability will only be proven in practice over time (Nakashima and Mufson 2015).

What is indeed clear is that satellites, in addition to terrestrial and submarine systems, need to be protected against cyber attacks in a number of ways. First and foremost, there needs to be encrypted systems to protect against fraudulent commands that can be sent to satellites by other than their system operators. In earlier times, there have been verified reports of even teenage hackers taking control of satellites as simply a prank. There are techniques that are used by system operators such as Intelsat to not only encrypt their commands to satellites, but a further requirement of a further authorization command from another command station location. Some 20 years ago, one of the major space agencies around the world did not have encryption protection on their satellites, but today all governmental agencies and private satellite operators have protective systems in place. Even tighter controls, such as independent authorization from two geographically remote locations for all spacecraft commands, is prudent practice.

What is now clear is that satellites need tight cybersecurity to protect both the spacecraft itself and all the information that they carry. A recent report on this subject contained the following alarming statement: “A group of Russian-speaking hackers is exploiting commercial satellites to siphon sensitive data from diplomatic and military agencies in the United States and Europe as well as to mask their location. . ..” This group that is known as “TURLA,” which is based on the sophisticated malware that it utilizes, has also targeted Chinese, Japanese, and Russia sources as well as pharmaceutical research labs and other sensitive and protected information centers – as reported by Kaspersky Labs, a noted cybersecurity firm. This operation, which has been running from at least 2007–2015, has reportedly concentrated on “hijacked satellite connections to obtain data and to cover its tracks” according to a Kaspersky Labs spokesperson (Nakashima 2015).

In the past decade, there has been perhaps a tenfold increase in governmental and private cybersecurity staffing seeking to encrypt and create passcode and firewall protection against cyber criminals and cyber spies, but this activity has not been concentrated with regard to various types of transmission media. Most recently there has been a focus on wireless transmission media such as supervisory control and data acquisition (SCADA) systems, microwave, and satellite links. There has also been increasing focus on so-called dark webs that trade in various aspects of computer crime, stolen credit card information, and the marketing of technology that can be used for cybercriminal activity. These efforts increasingly expose the degree to which dual use of commercial systems for military and defense-related activities creates vulnerable inroads into systems that are designed to be protected against unauthorized access. Thus, future military and defense-related satellite systems, whether dedicated military networks or commercial systems that are being used for strategic purposes, need to be encrypted and protected against all forms of cyber attacks (Pelton and Singh 2016).

Convergence on the Ground: Disaggregation in Space

This idea of consolidation of control facilities on the ground is one of the concepts that currently seem to be leading military planning for space systems. This means that as more and more capability is added in space for various types of communications and broadcasting, surveillance, remote sensing, navigation, nuclear explosions, and missile tracking, there can still be consolidated tracking, telemetry, and control facilities on the ground. Such consolidation on the ground can indeed be commercialized to save costs and allow interagency consolidation of control facilities. This type of thinking is currently being led by US military planners, but it is also being considered and implemented in Europe. In fact, in some cases, strategic space systems for Europe have been commercialized both in space and on the ground (Diamante 2015).

Advanced Coding, Processing, Autonomous Control, and Artificially Intelligent Systems Employed in Space Defense Networks

The design of today’s most advanced military and defense satellite is based on many factors, but advanced coding, processing, autonomous control, and artificial intelligence are perhaps among the most important in terms of adding capability and capacity at reduced cost and increased efficiency. These moves to greater efficiency, however, involve the move of controls from human response to automated decision-making. This concern about machine-coded response and how this might “go wrong” has been iterated many times in the Karel Capek play RUR, the “Terminator” movies, the Michael Crichton novel “Congo,” and most recently by space and technology leader Elon Musk. This is the ongoing concern of military space planners that are increasingly torn by the push toward greater efficiency and microsecond response times, on the one hand, versus the need for careful consideration and tempered response to crisis conditions, on the other.

There is thus focus on creating the ability to countermand controls and automated decision-making so that space systems of the future do not inadvertently create an unintended hostility or outbreak of war due to a defective “heuristic algorithm” similar to that reflected in “HAL” the computer in the film “2001: The Space Odyssey” made famous by Sir Arthur Clarke.

Orbital Space Debris and Military Satellite Networks

The general in charge of the NORAD tracking system for space situational awareness in late July 2015 said: “The satellite infrastructure that the Department of Defense (DoD) relies on for operational awareness is inefficient and is badly in need of modernization. The status quo isn’t acceptable, and changes must begin now” (Gen. John Hyten 2015).

Currently the US Space Defense system tracks about 22,000 orbital debris elements, primarily in low earth orbit and polar orbits. The focus is on these areas because it is in these orbits where most space debris objects currently are located. Furthermore it is where the most valuable US Space Defense assets are currently in orbit. Although the current ground and space-tracking systems are capable, they are not able to keep track of all potential threats in real time, especially with new plans to launch so-called megaLEO systems with potentially thousands of small satellites in LEO orbits in the range of 500–1200 km altitudes. The launch of 700 or so One Web and perhaps over 4000 Space X LEO satellites in gigantic constellations in the next few years will compound problems of accurate space situational awareness.

The tracking satellites, such as the one shown in Fig. 4, are currently used to track and monitor all missile launches, active and operational spacecraft, as well as orbital space debris. This seems almost like a losing battle as the number of space objects in earth orbit only keeps increasing on a net basis – despite orbital decay of some satellites. The recent innovation of building and launching small satellites using only low-cost off-the-shelf (OTS) components compounds the problem of more and more space objects in earth orbit. Further, as new spacecrafts are launched, it takes some time, such as on the order of an hour or two, to program the new orbits into software systems to accurately predict the orbit of these new spacecrafts.

Security of Space-Based Defense Systems

The Initial Defense Satellite Communications system (DSCS) was launched about a half century ago in 1965, and at that time, space systems – separate from rocket and missile weapon systems – were simply not part of military or defense infrastructure. Today, virtually every aspect of modern military and defense systems is highly dependent on various types of space systems. Thus, there are mobile, rapid stop

Fig. 4 US Air Force satellite utilized to maintain space situational awareness

on-the-move, fixed, and broadcast communications satellite systems. There are meteorological, surveillance, remote sensing, and command and control systems. There are also navigation, missile, satellite and space debris tracking systems, and nuclear explosion monitoring systems. The modern defense armies, navies, and air forces are today greatly dependent on space systems on which hundreds of billions of dollars are expended.

Today, just 50 some years later, the security of these systems against hostile attack are vital to national and regional defense systems. There is some level of defense of these space systems by radiation hardening and armor-protection against orbital debris and micrometeorites, but this type of protection is not sufficient against a missile strike or even a high-powered laser or directed energy weapon. Thus, there are other protection strategies such as seeking to disguise or otherwise not disclose the orbital location of prime assets or to move certain functions to dual-use commercial facilities so that defense services are merged with commercial systems.

Conclusion

The speed of development of new space systems continues to accelerate. The latest high-throughput satellites, megaLEO constellations, advanced coding, processing, autonomous control, and artificially intelligent systems are adding great efficiency and capability to a wide range of space systems. New antenna design, automation, and consolidation of ground control systems will fuel new efficiency. Dual-use commercial systems and integration of space systems with UAVs, aircraft, and terrestrial networks will likewise add new capability and also serve to save cost. The various innovations discussed in this chapter are under active R&D in the USA, Europe, Japan, China, and even in other countries such as Israel, India, and the Republic of Korea.

There are several important new challenges for new and future military and defense satellite systems. One of the greatest challenges will involve the reliable design and operation of global laser communications rings to allow broadband and effective Internet Protocol-optimized services. These may first emerge as commercial systems that migrate to military systems. This is the reverse of the development concepts that were first envisioned with the TSAT program. The other challenges include coping with cybersecurity protection for space-based national security systems and systems and strategies to cope with orbital space debris and the threat they pose to vital strategic space systems.

The development of on-orbit servicing could also allow the ability to provide active orbital debris removal and thus provide a way to repair satellites, extend lifetimes, upgrade satellite capability, and also remove derelict spacecraft or upper stage rockets that threaten key space assets. There is a legal/regulatory problem with such new on-orbit capabilities in that these systems could be considered to be a space weapon. In short, a system that can link up with an element of space debris and deorbit could also do the same for military communications satellites or other military space systems.

Cross-References

- ▶ Innovations in Hosted Payload Satellite Services
- ▶ New Millimeter, Terahertz, and Light-Wave Frequencies for Satellite Communications
- ▶ Trends and Future of Satellite Communications

References

T. Brinton, Pentagon cancels T-Sat program, Trims Missile Defense, *Space News* (6 Apr 2009), http://spacenews.com/pentagon-cancels-t-sat-program-trims-missile-defense/#sthash.ClJurD5V.dpuf. Last accessed 18 Sept 2015

Cisco's Space Router to Transform Satellite Communications:Company, Debuts First VoIP Call From Space (7 Dec 2010), http://www.cisco.com/web/strategy/docs/gov/IRISpr_120710.pdf. Last accessed 14 Sept 2015

C. Diamante, Government needs for convergence, disaggregation will require commercial support, (23 Sept 2015), http://www.intelsatgeneral.com/blog/government-needs-for-convergence-disaggregation-will-require-commercial-support/#sthash.WiGGSLp8.dpuf. Last accessed 27 Sept 2015

DoD, Australia sharing hosted UHF payload space on Intelsat 22 satellite (Apr 2012), http://connection.ebscohost.com/c/articles/74288959/dod-australia-sharing-hosted-uhf-payload-space-intelsat-22-satellite

Gen. J. Hyten, memorandum dated July 29, 2015 concerning space situational awareness quoted from *Intelsat General Newsletter*, July 2015

Hosted payloads and synergistic design alter military satellite landscape 2012, http://defensesystems.com/Articles/2012/02/28/cover-story-military-satellite-communications.aspx?admgarea=DS&Page=2. Last accessed 14 Sept 2015

Iridium Next 2015, http://space.skyrocket.de/doc_sdat/iridium-next.htm. Last accessed 14 Sept 2015

Management of the Department of Defense Information Enterprise, DoDD 8000.01, dated 10 Feb 2009

E. Nakashima, Russian Hackers find sophisticated way to steal data, *Washington Post*, 10 Sept 2015, P. A-11

E. Nakashima, S. Mufson, Obama, Xi reach accord on hacking, *Washington Post*, 26 Sept 2015, P. A-1 and A-6

J.N. Pelton, I. Singh, *Digital Defense: A Cyber Security Primer* (Springer Press, New York, 2016)

D. Sanger, US and China seek arms deal for cyberspace, *Washington Post*, 20 Sept 2015, pp. A-1, A-11

Transformational SATCOM 2012 (TSAT) Transformational Communications Satellite (TSAT) Advanced Wideband System, http://www.globalsecurity.org/space/systems/tsat.htm

Wideband Global Satcom Satellite, *U.S. Air Force Fact Book*, Posted 25 Mar 2015, http://www.afspc.af.mil/news/index.asp. Last accessed 18 Sept 2015

Part II

Satellite Precision Navigation and Timing Section

Introduction to Satellite Navigation Systems

Joseph N. Pelton and Sergio Camacho-Lara

Contents

Abstract

The "youngest" of the major satellite applications is the field of satellite navigation. This field of measurement and ranging through the use of satellite positioning systems first started in the context of scientific research. This initial use of satellites was simply for positioning and location. These activities, that were first based on using Doppler frequency shifts as a satellite orbited above, were largely scientific and not strategic. These types of activities included geodetics (i.e., such as the measurement of continental drift over time) or the collection of scientific information and data from atmospheric land- or ocean-based sensors where the specific locations of the sensors were important.

The real strides in the development of satellite navigation, however, came when space systems were developed for the specific purpose of precise targeting of missiles and various other types of weapon systems. The Soviet Union/Russian

---

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

S. Camacho-Lara
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Santa María Tonantzintla, Puebla, México
e-mail: sergio.camacho@inaoep.mx

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_10

Federation satellite navigation system, known as GLONASS, and the US-based Navstar system – with its Global Positioning Satellite (GPS) network – developed space-based systems that provided unprecedented capabilities to determine locations on the ground or oceans with great accuracy. Early systems such as Argos that relied on Doppler shift technology were only accurate within a precision of hundreds of meters. Today's advanced satellite navigation systems, however, are accurate for measurements that can be indicated with a precision only a few meters and within centimeters if utilizing reference stations.

The fact that GPS and GLONASS satellite signals are freely available in space for all to use has spurred the development of low-cost receivers for much more than strategic or military usage. Today, various types of civilian use of navigational and positioning satellites have become popular on a global basis. The latest development in application-specific integrated circuit (ASIC) chip technology has fueled the growth of applications based on the use of these precise space-based navigational and positioning systems. The availability of these highly capable but increasingly low-cost chips – small enough that they could be included in handsets such as cell phones or included with the electronics available on various types of vehicles (i.e., cars, trucks, buses, trains, airplanes, ships, etc.) – represented a real breakthrough. The development of specialized computer chips to perform satellite navigation calculations has increased the number of users of this technology from a relatively small population of several thousands to tens of millions. In the coming decade, the proliferation of international satellite navigation systems in orbit (i.e., USA, Russian Federation, China, India, Europe, and Japan) plus the continued development of ever lower cost of satellite navigation chips for receivers will continue the popular expansion of these increasingly "easy to use" and versatile systems.

Wide and easy access to low-cost consumer receivers for satellite navigation services strongly suggests that this trend of expanded use will continue to surge. Thus, within a decade, use of these devices will increase to the hundreds of millions if not billions of people. Atomic clocks with incredible temporal precision today allow satellite navigation systems to be used by military organizations, governmental geospatial scientists, and scientists to determine locations with great accuracy. But these applications are now just a small fraction of total usage. The deployment of the new international satellite navigational systems and smaller and lower cost ground receivers will support an ever-increasing civilian consumer market for an ever-expanding range of everyday applications. These consumer applications include driving a vehicle to a desired location, safely sailing a boat, going mountain climbing or hiking without getting lost, or simply finding out where you are within a city.

The precise time keeping ability of today's satellite navigation systems also means that these spacecrafts can also serve as a global timekeeper for computers and scientific experiments. The time stamp from a satellite navigation satellite can also be used not only for scientific or regulatory purposes but for other applications such as security and banking systems as well. These satellites are also used to support the synchronization needs of communications satellites. In short, navigation and positioning satellites have also become in many ways the world's timekeeper (Cesium clocks and global timekeeping, http://www.rfcafe.com/references/general/atomic-clocks.htm. Last accessed 14 Jan 2016).

The official US discontinuation of the so-called selective availability feature for the Navstar satellite system has accelerated the use of the GPS network for highly sensitive applications such as assisting in the takeoff and landing of aircraft. The fact that selective availability could be reactivated has nevertheless been one of the concerns and invoked reasons why other countries have now proceeded to develop and launch their own satellite navigation systems. In short, when nations believe that certain space infrastructure represents a strategic asset, there is a strong motivation to deploy such a system rather than depending on other nations to own and operate such networks.

Despite the strategic importance attributable to GNSS services, considerable progress has been made to achieve international cooperation compatibility and standardization among the six systems now in operation via the International Committee on GNSS (ICG) that now meets regularly under the auspices of the UN Committee on the Peaceful Uses of Outer Space.

Keywords

Argos • Chinese BeiDou Navigation Satellite system • Chinese Compass Navigation Satellite • Doppler shift • European Galileo Navigation Satellite System • European Space Agency • Global Navigation Satellite System (GNSS) • Global Positioning Satellite System (GPS) • Global Navigation Satellite System (GLONASS) • Indian Regional Navigation Satellite System (IRNSS) • International Committee on GNSS (ICG) • Japan Aerospace Exploration Agency (JAXA) • Japanese Quasi-Zenith Satellite System • Japanese Regional Navigation Satellite System • National Aeronautical and Space Administration (NASA) • Navstar • Ranging • Selective availability • UNISPACE III

Introduction

Satellite navigation and positioning systems continue to evolve. The initial systems that relied on Doppler shift as a low-Earth orbit satellite moved across the sky were useful for such applications as collecting data from ground- or ocean-based sensors or tracking a ship as it made its way across an ocean. The development of extremely precise atomic clocks that were precise within picoseconds has allowed the development of a whole new type of space infrastructure that allows precise location and positioning that is orders of magnitude better than the first satellite systems. A network of over 20 such satellites in a well-designed constellation allows users of ground receivers to instantly calculate where one is located on the Earth's surface. The propagation time between various visible satellites can be calculated with great precision – especially when four or more satellites can be viewed at exactly the same time.

When such satellites using propagation time calculations were first developed in the 1980s, the intended purpose was strategic and, in fact, focused on the precise targeting of missiles and other types of weapons systems. This military application initially justified the billions of dollars that were involved in designing, manufacturing, and deploying scores of the GLONASS and GPS satellites. Once the satellites were up and transmitting signals back to Earth, people began envisioning a host of new uses for these devices. Initially, these devices that were produced in low quantities were quite expensive, but as the demand for satellite navigation devices grew and their production increased to higher and higher levels, the costs began to fall. Today, the latest in computer chip technology has lowered the cost of production of receivers sharply and also let the devices to become very compact and even integrated with mobile phones simply by including an integrated circuit developed to process satellite navigation signals.

Initially, the satellite navigation systems designed for targeting were limited to the Soviet Union/Russian Federation GLONASS system and the US Navstar System that is composed of GPS spacecraft. Today, there are a growing number of satellites systems for navigation being developed and deployed by other space-faring nations. The navigation satellite systems in existence and those being developed are referred to collectively as Global Navigation Satellite Systems (GNSS). These satellite networks are being deployed in different types of orbits and some are oriented toward regional coverage as will be discussed in later chapters, but the combined use of these various satellite networks will allow for increasingly precise location and timing measurements. Most of these satellites will be providing their signals for free, although encrypted signals will sometimes be used to support strategic applications or to collect special value-added data for which clients will pay fees. From an economic perspective, the problem does become that of how can one system collect fees for its services when the same or similar information can be obtained from other countries' networks for free? In some cases, the economic model is based on the collection of sales tax from the sale of receivers optimized for certain navigation satellites or the collection of value-added taxes for services associated with particular satellite navigation systems.

The proliferation of satellite systems for navigation and positioning has led to the recognition that some degree of coordination and standardization might be advisable to allow the various systems to work together efficiently. To this end, the Third United Nations Conference on the Exploration and Peaceful Utilization of Outer Space (Report of the Third United Nations Conference 1999) (UNISPACE III), held in 1999, called for the enhancement of, universal access to and compatibility of, space-based navigation and positioning systems. This call resulted in the development of a new international coordination entity, the International Committee on GNSS, (ICG) to exchange information on the various satellite navigational systems and sort out such issues as the optimum frequencies to use to not only support satellite navigation but also minimize interference among the satellite systems. The ICG, as discussed below, first met in 2006 and has now developed processes to coordinate the various satellite networks and to facilitate the development of units capable of receiving signals from multiple satellite networks.

Today, the global market for satellite positioning and navigation is the most rapidly growing satellite application in terms of percentage growth. Commercial applications of these satellites and their ground receivers as estimated in statistics of the Organization of Economic and Commercial Development (OECD) has grown into a market that was nearly US$15 billion per annum as of year-end 2009 and may grow, in terms of total wholesale and retail sales of equipment and services – including new space systems to a global market of US$30 billion per annum by year-end 2014 or 2015 (OECD 2011). Global statistics in this area are difficult to assemble, and others have put these numbers at a lower level, but all agree this is the most rapidly growing space-based business. Certainly, satellite navigation is now second only to satellite telecommunications in terms of a measurable commercial market.

Technology of Satellite Navigation Systems

The first satellite navigation system was Transit, a system deployed by the US military in the 1960s. Transit's operation was based on the Doppler effect: the satellites traveled on well-known paths and broadcast their signals on a well-known frequency. The received frequency will differ slightly from the broadcast frequency because of the movement of the satellite with respect to the receiver. By monitoring this frequency shift over a short time interval, the receiver can determine its location to one side or the other of the satellite, and several such measurements combined with a precise knowledge of the satellite's orbit can fix a particular position.

There are still satellites that use Doppler frequency shifts for position determination such as the polar-orbiting French Argos satellite positioning system. Argos transmitters' messages are recorded by a constellation of satellites carrying Argos instruments and then relayed to dedicated processing centers. This system has been operational since 1978 and was initiated jointly by France and the USA. While Argos transmitters have become known to the public through their use in tracking yacht races, their primary mission remains the collection of data for the scientific community. By measuring temperature, pressure, humidity, and sea levels, Argos takes the pulse of the planet and its atmosphere.

The technology used in the major satellite navigation networks represented by the US GPS system, the Russian GLONASS system, the Chinese BeiDou/COMPASS system, the Indian Regional Navigation Satellite System, the European Galileo system, and the Japanese Quasi-Zenith Satellite System all hinge on ultraprecise atomic clocks and the relative transmission times from a network of satellites in a structured orbital constellation.

The satellite broadcasts a signal that contains orbital data (from which the position of the satellite can be calculated) and the precise time the signal was transmitted. The orbital data is transmitted in a data message that is superimposed on a code that serves as a timing reference. The satellite uses an atomic clock to maintain synchronization of all the satellites in the constellation. The receiver compares the time of broadcast encoded in the transmission with the time of reception measured by an internal clock, thereby measuring the time-of-flight to the satellite. Several such measurements can be made at the same time to different satellites, allowing a continual fix to be generated in real time.

Each distance measurement, regardless of the system being used, places the receiver on a spherical shell at the measured distance from the broadcaster. By taking several such measurements and then looking for a point where they meet, a fix is generated. However, in the case of fast-moving receivers, the position of the signal moves as signals are received from several satellites. The basic computation thus attempts to find the shortest directed line tangent to three to four oblate spherical shells centered on the satellites emitting the signals. Satellite navigation receivers reduce errors by using combinations of signals from multiple satellites and multiple correlators and then using techniques such as Kalman filtering to combine the noisy, partial, and constantly changing data into a single estimate for position, time, and velocity. This "estimate," however, is actually quite precise.

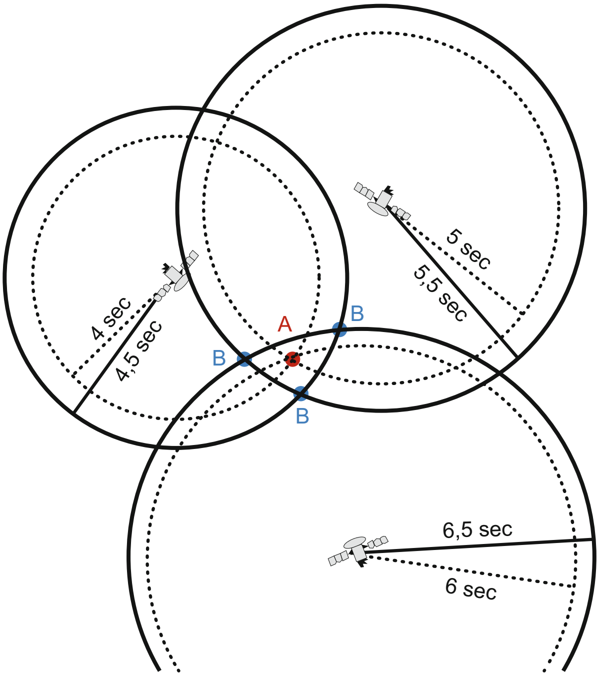
The intersecting "virtual spheres" represented by the transmission time from the network of satellites serve to create a positional location on the Earth. The "virtual spheres" from three satellites can identify a location but four or more satellites can determine a unique location even more accurately. More information on atomic clocks and the precision of today's navigation and timing satellites is included in later chapters in this section.

These various satellite networks represent very complex systems requiring an investment of billions of dollars to design, build, and launch a large constellation of very sophisticated satellites, but the solid trigonometry involved is just a matter of rapid calculations called trilateralization. This process is depicted in Fig. 1 below where the location "A" on the Earth's surface has been identified.

There are two main technical problems with the satellite navigation systems today and they both hinge on the need to transmit radio frequency signals from the satellite to the ground receivers. The first problem is that interference from other satellites or terrestrial sources can affect performance. Specifically, jamming could create a major problem to these satellites as they do not emit high gain signals because the signal is very broadly spread across the Earth. The second problem is that a variety of frequencies are used in satellite navigation systems. This started with the GPS and GLONASS systems using different frequencies. This makes receiving units that are engineered to receive signals from multiple systems more expensive. The International Committee on Global Navigation Satellite Systems (ICG) and its Working Group A (WG-A) on Interoperability and Compatibility are, among other things, addressing these radio frequency issues.

As a final note, it should be noted that satellites equipped with cesium atomic clocks can provide with great accuracy not only location, three-dimensional positioning, and navigational velocity but also time synchronization. The cesium clocks, which are remarkably accurate "scientific instruments," are one of the major contributors to the cost of navigational satellites.

Fig. 1 Satellite-based determination of position via trigonometric trilateral calculation



International Development of Satellite Navigational and Positioning Systems

The Navstar GPS Satellite system and the Russian GLONASS satellite system, now almost completely repopulated, already provide two comprehensive global navigation and position systems that have enough satellites to provide the capability to determine a precise location over the entire surface of the world as well as incredibly precise timekeeping. The latest generation of GPS Block III satellites, and improved ground control system, represent a $5.5 billion investment that will make the system more resistant to interference, better able to support precision applications such as aircraft takeoffs and landings and even more precise time synchronization (Newman 2012). The GLONASS system had in the 1990s and the early 2000s lost a number of its original satellites, but a new generation of satellites has now been launched and the network has some two dozen satellites, fully functional spacecraft in the restored network (Global Navigational Satellite System (GLONASS) 2009).

Despite the complete global coverage of these two systems, a rather amazing array of additional satellite navigation and positioning satellites, costing many billions of dollars, have either been launched or new spacecraft designed for this purpose are in final aspects of design, engineering, manufacture, or launch. These additional systems from China, Japan, India, and Europe plus the US and Russian systems are discussed in detail in later chapters.

Applications and Markets

The ability to locate, navigate, and position using satellite technology – essentially in real time – has turned out to be a very valuable capability. Initial applications in such field as geodetics – measurement of motion of the Earth's tectonic plates – or being able to know the location of sensors sending up data about ocean or weather conditions, were very much oriented toward scientific, geological, or ecological conditions. The possibility of targeting missiles using space systems led the governments of both the USA and Soviet Union to make the multibillion investment in global constellations of positioning and navigational satellite systems. It also led the USA to design the Navstar system to include a "selective availability" capability that would in times of warfare prevent anyone without specific codes to accurately target weapon systems using this constellation. Selective availability, in essence, was a degradation of the signal that was received by a user without the secret codes. President Clinton, however, signed an Executive Order to confirm formally a pledge not to use selective availability. This Executive Order was in part based on the fact that increasingly vital applications involving aviation traffic and airport safety depended on GPS accuracy and in part based on the fact that new capabilities involving calculations based on known locations (reference stations) could actually defeat the "selective availability" of spurious information in any event.

The scientific and defense-related applications of satellite navigation systems continue to increase but the civil applications for industrial and consumer services have skyrocketed in recent years. The following listing of ways that these satellite networks can be used today is not complete, but is nevertheless indicative of the uses that are now being made. Specialized data bases have been developed for very specific applications that consumers are now purchasing as "online applications."

- Cross country skiing
- Driving instructions to unfamiliar or unknown addresses
- Emergency recovery and restoration operations – repair crews, emergency vehicles, etc.
- Fire-fighting assistance in coping with unfamiliar terrain and with forest fires
- Golf course layouts and specific dimension of course hazards
- Hazardous marsh, fen, and difficult terrain identification
- Hiking and mountain and rock climbing information
- Location determination for automotive breakdowns for tow trucks
- Location reference points and "icemarks" for explorers in arctic regions and on glaciers
- Mapmaking
- Ocean hazards and location of pollution zones
- Oil spill containment and coping with pollutants, red tide, etc.
- Police, Narcotics enforcement, and illegal immigrant enforcement activities
- Precision agriculture
- Rescue operations in mountains, wilderness, arctic regions, etc.
- Routing of trucks

- Sailing and boating locations on lakes, rivers, canals, and on ocean voyages
- Synchronization services for computer networks, communications, and energy systems
- Security systems that are based on codes and real-time positioning of users
- Surveys and establishment of legal boundaries
- Timekeeping for scientific experiments
- Transit system maps and locations of stations for virtually every city in the world
- Transportation design, construction, repair, and maintenance
- Utility repair crews
- Urban planning

The navigation, position, and timing satellites can be used for knowing where you are, where you are going, and the best way to get there. These satellites can also be used not only for map making and surveying and the like but also to know the exact time and to synchronize computer networks or communications systems. The real importance of this indicative list is that almost any activity is enhanced by flexible and reliable broadband telecommunications and the instant ability to know one's location and the ability to navigate to where one needs to go safely and effectively. The gradual development of improved software for satellite positioning and navigation systems adds precise knowledge as to the location of roads, highways, railroads, utilities, canals, rivers, bridges, shoals, reefs, cliffs, marshes, and the layout of towns, cities, and even amusement parks, golf courses, colleges, and industrial and governmental complexes.

The constantly expanding utility of satellite navigational systems and the software that provides the location and current timing of practically everything has made these space systems very valuable. First responders and emergency aid workers; military troops; all forms of transportation services; workers involved with deployment, repair, and maintenance of utility systems; urban planners and map-makers; and indeed civilian consumers going about their daily lives – either at work or at play – increasingly find that satellite navigation is a sine qua non. Each day the number of people who find satellite navigation almost as important as mobile communications continues to grow. As these functions are wedded together in cell phones, the applications only increase (Biba 2009).

Satellite navigation, as an identifiable economic activity, has thus grown rapidly in recent years and continues to lead growth in the space industry in terms of a percentage increase in revenues. The remarkable growth of space navigation services all trace back to the ability to make the technology available to global consumers at a very low cost and in a "demystified way."

The development of specialized computer chips that have embedded in them the ability to perform the specialized calculations needed to allow position location on a near instantaneous basis was a key step. The next important step in the satellite navigation, positioning, and timing industry was the inclusion of this technology in consumer devices and particularly in cell phones. Today, third and fourth generation cell phones come with embedded chips that allow consumers to know instantly where they are and to be synchronized to the Internet. Most recently, it has been the

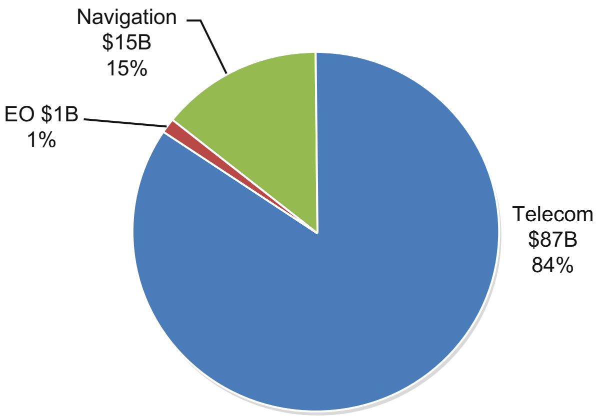


Fig. 2 Distribution of satellite services revenues (in US$ billions) between the three main types of satellite applications in 2009 (Source: Op cit, *The Space Economy*, OECD 2011)

development of applications (i.e., particularly GPS-related "apps" on the latest generation of cell phones) that can be downloaded to mobile devices that has contributed to an explosion of satellite navigation and positioning uses not only in the USA but all across the globe (Fig. 2).

International Coordination, Standards, and Regulatory Issues

At the UNISPACE III conference, one of the clear trends in space development that emerged was the rapid proliferation of space navigation and positioning systems. The work of the "Action Team" established by the United Nations Committee on the Peaceful uses of Outer Space resulted in what is now known as the International Committee on GNSS (ICG). The ICG was formed through the voluntary participation of all the GNSS operators in order to create a useful forum for technical coordination and interference reduction in the operation of satellite navigation systems. The ICG is also now assisting in formulating internationally accepted satellite navigation and positioning standards and to stimulate key new civil applications around the globe (International Committee on the GNSS 2009). The functioning of the ICG and the specific activities of its Working Group A and its three other working groups are detailed in "International Committee on GNSS" (International Committee on the GNSS 2009).

One of the great commercial successes in the global space economy that has emerged primarily in the last decade is that of the dual-use global satellite navigation, timing, and positioning services. At first, military and strategic reasons provided the financial and economic justification to develop the technology and to launch very expensive and complex spacecraft with onboard super accurate cesium atomic clocks. These clocks are essential to the capability to determine locations and times, and thus speeds, with a very high accuracy. This new space technology allowed a large number of new applications to emerge where determining one's location and transit vectors can be used in a remarkably diverse number of ways. In the meantime, advances in software and hardware technology have combined with the improved signals from the GNSS to provide more accurate measurements at lower costs. One can expect that, as the GNSS operators make the satellite systems interoperable, a single receiver will be able to process the signal from the satellites of any of the GNSS. This will result in even lower costs for receiver equipment and the possibility of viewing ten or more satellites at the same time, resulting in more accurate measurements in urban and mountainous environments.

The economic applications related to scientific, military, government, and particularly industrial and consumer-based applications continue to expand, and exponentially so. The current size of the satellite navigation, positioning, and time industry in economic terms is over US$15 billion and is expected to grow to perhaps $30 billion in 2–3 years. By making other economic areas more efficient and by creating new ones, the economic impact of GNSS will be in the hundreds of billions of dollars (US).

Conclusion

One of the great commercial successes in the global space economy that has emerged primarily in the last decade is that of the dual-use global satellite navigation, timing, and positioning services. At first, military and strategic reasons provided the financial and economic justification to develop the technology and to launch very expensive and complex spacecraft with onboard super accurate cesium atomic clocks. These clocks are essential to the capability to determine locations and times, and thus speeds, with a very high accuracy. This new space technology allowed a large number of new applications to emerge where determining one's location and transit vectors can be used in a remarkably diverse number of ways. In the meantime, advances in software and hardware technology have combined with the improved signals from the GNSS to provide more accurate measurements at lower costs. One can expect that, as the GNSS operators make the satellite systems interoperable, a single receiver will be able to process the signal from the satellites of any of the GNSS. This will result in even lower costs for receiver equipment and the possibility of viewing ten or more satellites at the same time, resulting in more accurate measurements in urban and mountainous environments.

The global installed base of GNSS devices is currently in a state of rapid flux. The number of units is predicted to almost double to seven billion by 2019. This is equivalent to almost one GNSS receiver for every person on the planet. The use of smartphones for global navigational space systems (GNSS) services continues to dominate with there being over 3.3 billion as of year-end 2015. This is followed in terms of number of GNSS devices applications of units used for road-related services with there being about 0.3 billion of these in vehicles.

The percentage growth of GNSS units will continue to be quite different at the regional level with there being a predicted growth of about 8 % per year within Europe and the USA for the next few years. The primary region of global market growth in terms of total numbers will be Asia-Pacific, which is forecasted to grow

11 % per year, from 1.7 billion in 2014 to 4.2 billion devices in 2023. The Middle East and Africa will grow at the fastest percentage rate – 19 % per year – but starting from a low base.

The breakdown of applications is dominated by location-based services (53 %) and road-based services (38 %). Other applications include surveying, rail, agriculture and fishing, maritime, timing synchronization, and aviation. Although aviation is only about 1.0 % of the use, it is a very crucial application because it is key to flight safety (GNSS Market Report 2015).

The economic applications related to scientific, military, government, and particularly industrial and consumer-based applications continue to expand, and exponentially so. The current size of the satellite navigation, positioning, and time industry in economic terms is very difficult to estimate although it is perhaps as large as $30 billion if not more. The difficulty is in knowing what economic activity to be considered as being directly attributable to a GNSS or positioning, navigation, and timing (PNT) service or to the industry or governmental service it supports. As the number of GNSS equipped devices more or less double by 2019, it is hard to estimate whether the associated market applications will increase proportionately in terms of measured value. By making other economic areas more efficient and by creating entirely new applications and services possible, the economic impact of GNSS will clearly be enormous and may in time even outstrip satellite communications services.

Cross-References

- ► Current and Future GNSS and Their Augmentation Systems
- ► Global Navigation Satellite Systems: Orbital Parameters, Time and Space Reference Systems and Signal Structures
- ► International Committee on GNSS

References

E. Biba (2009) Inside the GPS revolution: 10 applications that make the most of location. Wired Magazine http://www.wired.com/gadgets/wireless/magazine/17-02/lp_10coolapps?currentPage=all. Last accessed 14 Jan 2016

Global Navigational Satellite System (GLONASS), (2009) *Access Science,* http://accessscience.com/content/Global-Navigation-Satellite-System-(GLONASS)/YB980695. Last accessed 14 Jan 2016

GNSS Market Report, (2015), http://www.navipedia.net/index.php/GNSS_Market_Report#Previous_Report_Issues. Last accessed 31 Dec 2015

International Committee on the GNSS,(2009) www.oosa.unvienna.org/oosa/SAP/gnss/icg.html. Last accessed 14 Jan 2016

OECD, *The Space Economy at a Glance* (2011), http://www.oecd.org/sti/futures/space/48301203.pdf. Last accessed on 14 Jan 2016

Report of the Third United Nations UNISPACE Conference, (1999), www.un.org/events/unispace3/dailypro/Progpart1.pdf , Last accessed on 20 may 2016

Global Navigation Satellite Systems: Orbital Parameters, Time and Space Reference Systems and Signal Structures

Rogerio Enríquez-Caldera

Contents

R. Enríquez-Caldera (✉)
Centre for Space Science and Technology Education for Latin America and the Caribbean, México Campus (CRECTEALC), National Institute of Astrophysics, Optics and Electronics (INAOE), Coordinación de Electrónica, Tonantzintla, Puebla, Mexico
e-mail: rogerio@inaoep.mx

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_93

Abstract

Global navigation satellite systems (GNSS) are part of the most complex modern space systems humankind has created, and therefore their orbits, orbital parameters, and their two main terrestrial mappings are firstly described. Different frames of space-time reference systems are treated as part of such descriptions.

Communication systems engineering are important sections to allow for a GNSS precise fix positioning. All signal structures and data streams are treated for a clear understanding permitting the reader to see how ranging is obtained from space.

Theoretical and practical error budgets are considered to give the reader a perception of limitations during scientific and/or technical user campaigns or for simple common life enjoyment.

Keywords

MEO orbit • Reference systems • International atomic time (TAI) • Signal structure • Frequency issues • Relativistic time dilation • Modulation techniques • Pseudorandom codes • Carrier frequencies • C/A and P(Y) codes • Data stream • Pseudo-range • Accumulated delta range • Sources of error and error budgets • Differential GNSS • Kinematic and real time kinematic systems

Introduction

Global navigation satellite systems are comprised of tens of satellites placed in middle Earth orbits (MEO), somewhere around 20,000 kms from the Earth. Scientific, technological, and technical aspects of modern human capabilities were weaved into GNSS as in no other systems and are finding applications in practical everyday life. The use of these space systems, originally developed for military purposes, is increasingly contributing to the well-being of humankind. Space topics, such as aeronautics, propulsion, tracking, communications, handheld multiuser receptors, signal processing, electronic and inertial sensors, estimation theory, navigation, mapping, electromagnetic propagation, relativistic physics, reference systems, timing, and many other disciplines, constitute the elements of these GNSS, a wonderful human creation which is closest to a ubiquitous system.

Keeping in mind the above, the reader would, first of all, appreciate the complexity involved in sending a constellation of spacecraft to specific orbits, envisioning in perspective the trajectories followed by each and all satellites that comprise a global navigation satellite system and to time the arrival of their signals to a ground receiver.

In order to have a complete global concept of how GNSS systems work, one must understand different frames of reference around the globe by considering both the dimensional space description and the abstract reference time.

Through basic communication engineering, we can understand how electromagnetic signals from space and Earth objects can be joined to give space-time ranging calculations, thus making possible many position- or time-dependent applications in an ample variety of fields including scientific, engineering, technical, social, and even the very life-sustaining capabilities of a society. Thus, the basis for GNSS signal structure and modulation communications are presented as well as those techniques that permit a user to achieve a desired positioning precision.

Since reception of satellite signals can be obtained from either lumped or distributed receiver-transmitter systems, static or dynamic environments, clear, partial, or almost completely obstructed access for electromagnetic waves, a description of error sources and ways of dealing with measurements made in each type of situation is a must.

GNSS Orbits

To be able to cover the terrestrial globe at any given static or dynamic spot at any given time, GNSS are not placed in GEO orbits. For the same purpose, GNSS could not be comprised of a single spacecraft but rather a well-established set of spacecraft forming a proper and well-calculated system. The specific kinematic geometry for the set of satellites that form such system is called constellation. The very first GNSS was the well-known GPS and can be taken as a good example of a constellation placed on MEO orbits shown in Fig. 1.

Modern display and tracking of the actual real time orbits can be followed at NASA's science@NASA Satellite Tracking (http://science.nasa.gov/realtime/jtrack/3d/JTrack3D.html/). Choosing any GPS satellite one can obtain a practical idea of what a MEO orbit means with respect to Earth.

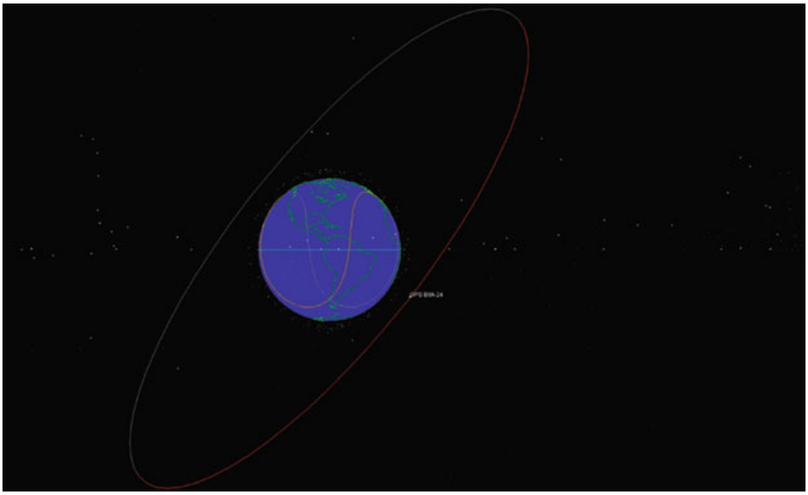


Fig. 1 Example of a GPS satellite MEO orbit (http://science.nasa.gov/realtime/jtrack/3d/JTrack3D.html/)

Now, following the GPS example, it can be seen that for a constellation of 24 GNSS satellites at MEO orbits with an intended global coverage of at least four satellites above the celestial vault for a given point on Earth, there ought to be four satellites in a particular MEO orbit (defining an orbital plane) and there ought to be six orbital planes. Satellite separation in each plane is about 90° and, in the GPS case, planes differ from each other 60° and the origin plane is taken as being 55° respect to the equatorial plane.

Up to date, there are two well-established GNSS: GPS (Mohinder et al. 2001) owned by the USA and GLONASS (Global Navigation Satellite System: GLONASS 1998) owned by the Russian Federation. There are two other GNSS under development and planned to be active in the near future: GALILEO (Wilson 2006) owned by the European Union (EU) and COMPASS (Chen et al. 2009) owned by the People's Republic of China. The Quasi-Zenith Satellite System (QZSS) is a proposed three-satellite regional time transfer system and Satellite-Based Augmentation System (SBAS) for the Global Positioning System that would be receivable within Japan. The first satellite "Michibiki" was launched on 11 September 2010. Full operational status is expected by 2013. The GPS Aided Geo Augmented Navigation (GAGAN) is another planned regional SBAS by the Indian government. These systems will improve the accuracy of a GNSS receiver by providing reference signals (see chapter "▶ Current and Future GNSS and Their Augmentation Systems").

MEO Orbits

Table 1 shows all orbit data related to the GNSS (Shaw et al. 2008; Revnivykh 2008; Verhoef 2008; China Satellite Navigation Project Center 2008).

Reference Systems

Despite the fact that one can visualize what 90° of separation between satellites in a particular plane and 60° among all planes means, such angles should be referred to a specific frame of reference. The most widespread 3D reference system is that constructed from three straight lines with right angles among them: the Cartesian coordinate system (CCS) because its geometry and associated algebra are easy to understand.

From practical human-size measurements, our world can be thought as being Cartesian in a spatial square box of 100 m each side, meanwhile, for global Earth-size measurements, a geometrical abstraction reference named geoid has been defined, and finally, for planetary-size measurements, a sphere centered in the Sun is the most common idea of a reference.

When quasi-spheres are involved in mobile outer space trajectories, two common frames of reference are used: one related to the center of the Earth (conventional terrestrial system CTS) and the other one related to a celestial or astronomical reference system (ARS) the former being not an inertial frame of reference the latter one almost fulfills the inertial frame requirement.

Table 1 MEO orbits for GNSS

GNSS/space segment	GPS	GLONASS	Galileo	COMPASS/BeiDou 2
Constellation (space vehicles)	24 operational/ 3 spares 31 GPS-M	24	27 operational/ 3 spares	30/5
Orbit	MEO	MEO	MEO	MEO/GEO
Orbital period	11 h 56 min ~ 12 sidereal h	11 h 15 min	14 h 05 min	12 h 50 min
Altitude/orbital radius (Km)	~20,200/ 26,600	19,100/ 25,508	~23,616/ 29,600	21,500/27,840 35,785/42,164
Planes	6	3	3	3
Planes position (° from equator) latitude/ separation	55/60	64.8/120	56/120	55/120
Satellites per plane/degrees of separation	4/90	8/45	9/40 + 1	10/36
Declination				58.75E,80E,110.5E,140E, 160E
Reference system/time	WGS-84/GPST	PZ-90/ UTC(SU)	GTRF/GST	China geodetic system (CGS 2000) ~ ITRF/China UTC (BDT) ~ UTC

Figure 2 shows the CCS, CTS, and ARS reference systems and their respective coordinate axis known as *horizontal datum and vertical datum* (Rogers 2003).

Relationships among CCS, CTS, and ARS values are given in terms of coordinate transformations (Xu 2003). To view the effects of such transformations the reader could use one of the most practical and modern informatics tool: the Google Earth (http://www.google.com/earth/learn/beginner.html) where one can switch views from CTS to CCS and vice versa using the TOOLS command and changing datum views at the 3D view window tab. Thus, vertical datum (orthometric heights H) and horizontal datum (latitude φ and longitude λ) are shown either in degrees for the CTS or in meters for the CCS under the universal transversal Mercator (UTM) projection (Carnes 2007) (In Fig. 2, H is changed from the geoid reference to the ellipsoid reference or geometrical height h).

Ephemeris

From the previous concepts, it can be seen that satellite position coordinates, for non-GEO orbital satellites, change with time. In Fig. 3, the 3D position vector r of a dynamical satellite can be most properly written as $r(t)$ which is known as the *satellite's ephemeris* (Beutler 2005).

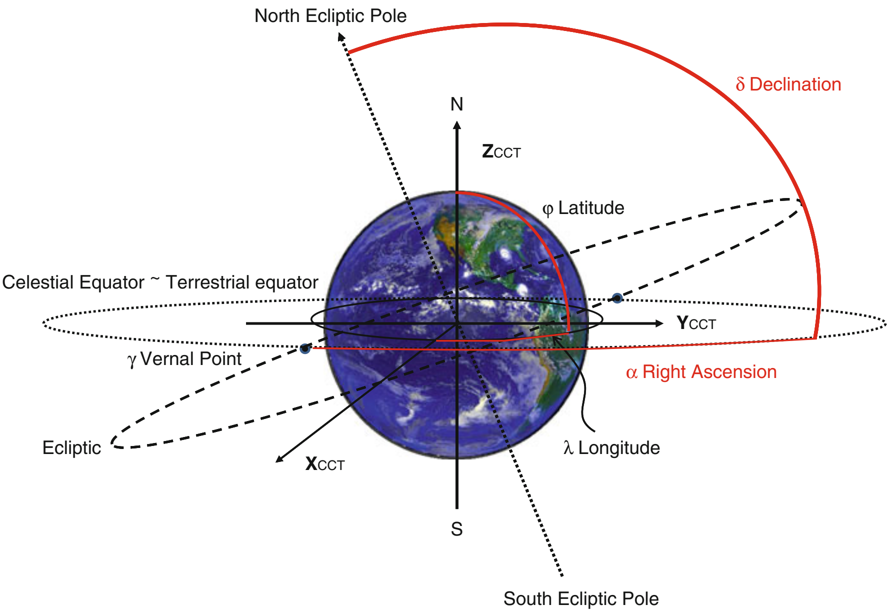


Fig. 2 Horizontal datum and vertical datum in the CCS (X, Y, Z), CTS (λ, φ), and ARS (α, δ) reference systems

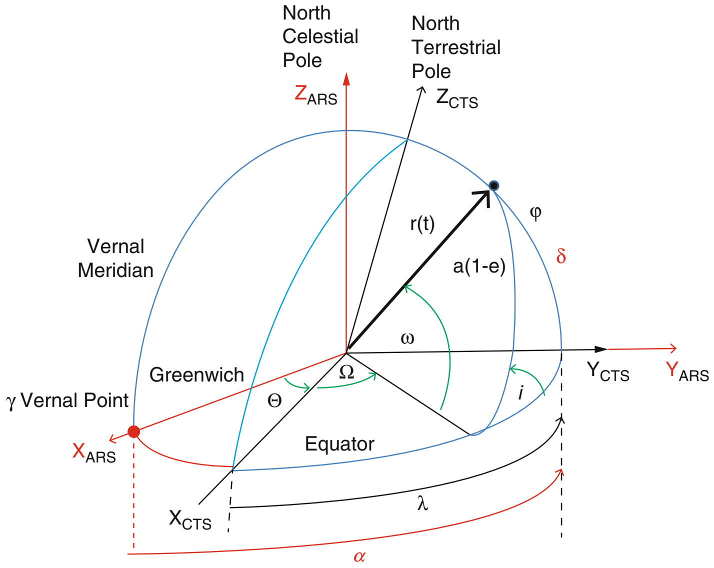


Fig. 3 Satellite ephemeris

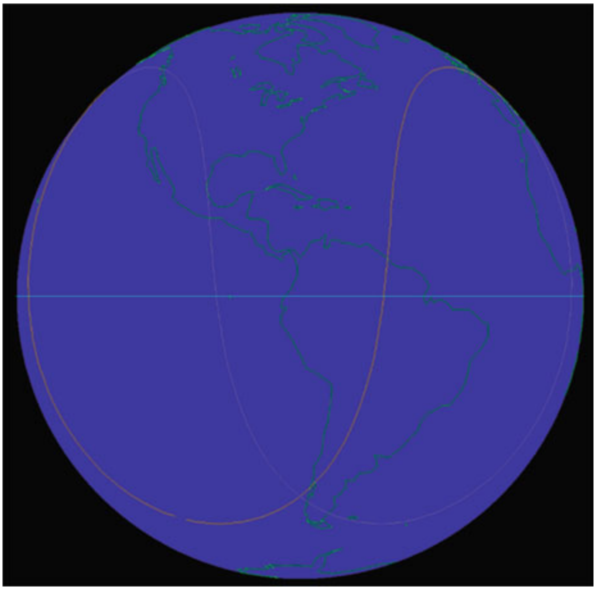


Fig. 4 Ground track projection (http://science.nasa.gov/realtime/jtrack/3d/JTrack3D.html/)

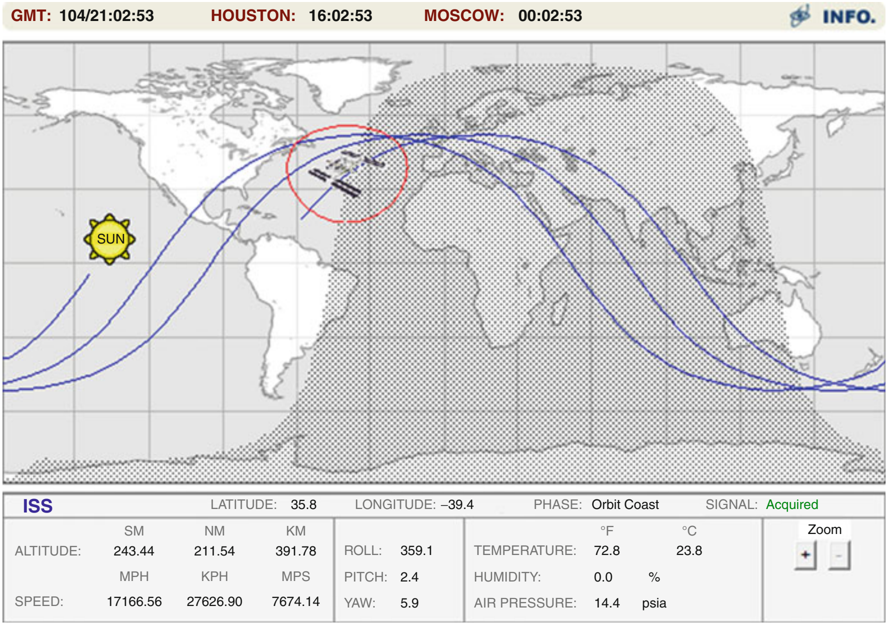
Ephemeris are calculated using celestial mechanics and can also be predicted using estimators from observed data. Satellite observations are a task for the tracking ground systems which are part of the so-called terrestrial segment of any satellite system. For the purpose of information completeness, satellite systems are composed by two segments, the space segment, composed by the satellite constellation, and the terrestrial segment, composed by the ground control and tracking subsystem. The terrestrial segment also includes a subsystem that is formed by all the GNSS users. Figure 4 shows all necessary GNSS segments.

From Fig. 3 ephemeris are given by: right ascension (α), declination (δ), argument of latitude (ω), and range $\| r(t) \|$.

Ground Tracks

Figure 5 represents an ephemeris projected over the surface of the terrestrial globe for a single orbit and for all the time that takes that orbit to be completed. Those projections are known as ground tracks and in fact are closed paths that resemble sinusoidal trajectories.

Ground tracks together with geographical maps are useful to actually see ephemeris and ease satellite tracking or, in a more general case, to visualize the position of any space vehicle around Earth. Figure 5 shows a typical ground track map consisting of a flattened 3D map where a specific ground track has been drawn. There could be as many ground track maps as necessary to properly and clearly display as many



GMT: 104/21:02:53 HOUSTON: 16:02:53 MOSCOW: 00:02:53

ISS LATITUDE: 35.8 LONGITUDE: –39.4 PHASE: Orbit Coast SIGNAL: Acquired

	SM	NM	KM
ALTITUDE:	243.44	211.54	391.78
	MPH	KPH	MPS
SPEED:	17166.56	27626.90	7674.14

ROLL: 359.1
PITCH: 2.4
YAW: 5.9

	°F	°C
TEMPERATURE:	72.8	23.8
HUMIDITY:	0.0	%
AIR PRESSURE:	14.4	psia

Fig. 5 Typical ground track map (http://spaceflight.nasa.gov/realdata/tracking/index.html)

satellites as are being tracked. Obviously, these maps are 2D but allow the visualization of the past, the present, and the future of satellite paths.

Polar Plots

There is another 2D kind of satellite movement representation: polar plots. A polar plot is nothing else than a polar projection of the whole sky dome above a particular observer looking up the zenith and oriented toward north. Because this is a 2D projection, the vertical datum, i.e., the satellite's range is not considered. Thus, the horizontal datum of satellites above the observer is represented on one hand by a compass giving the satellite's right ascension and on the other by circles that represent satellite's declination angle starting from the zenith and ending at the observer's horizon. Figure 6 depicts an example of a polar plot with four satellites in view.

It should be noted that polar plots just show the position of satellites for a given time. Nevertheless, history of trajectories can be consulted as well as future positions using the memory and/or predicting tools proper of the modern informatics system.

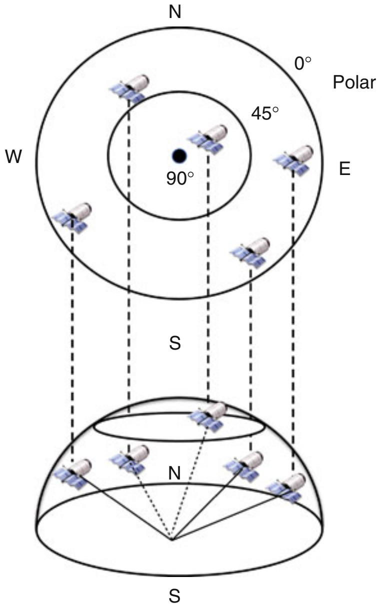


Fig. 6 Polar plot with four satellites above an observer

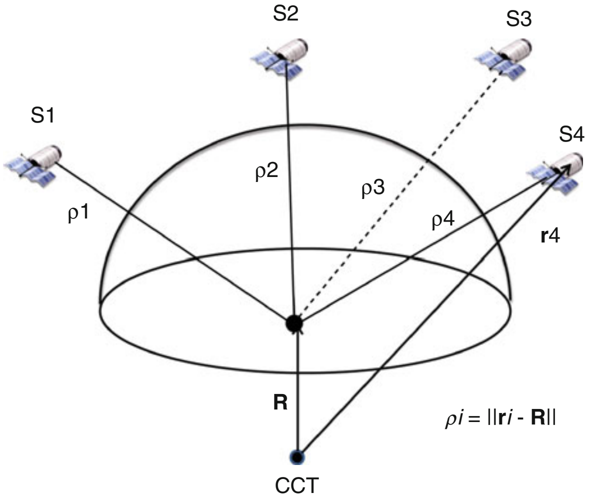


Fig. 7 Point positioning

Point Positioning

Point positioning is the process of finding the vector R of a specific user in terms of the satellite range, ρ, and the instantaneous satellite ephemeris $r(t)$ or simply r. Point positioning is shown in Fig. 7.

Thus, the user vector R is given by the equation

$$R = r - \rho$$

Differential GNSS

Sometimes R is not the quantity we are interested in but rather a position with respect to another point taken as a new reference. In such case, we can talk about differential positioning. Figure 8 shows this concept.

The differential GNSS (DGNSS) equation is easily obtained from Fig. 8 as:

$$\Delta R = R_2 - R_1 = \rho_1 - \rho_2$$

where $R_i = r - \rho_i$ and r is the ephemeris to the very same satellite used for positioning fix.

There is an advantage to using DGNSS and that is that the uncertainties in the determination of the time of arrival of signals from the satellites "in view" of the GNSS receiver are reduced, and therefore DGNSS is used when high-precision positioning is required.

However, the differential concept extends to other ways of calculating positions on Earth. There are many ways of doing differential positioning (Wells 1987): differences between observations points, differences between satellites in view, differences between epochs of satellites, and some other linear combination of observations. It is even possible to define a network of local references. The purpose of such a network is to obtain data redundancy, thus allowing for diminishing positioning errors.

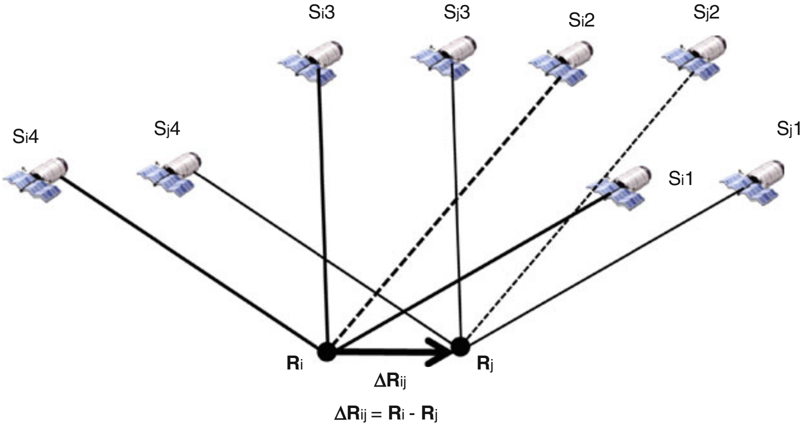
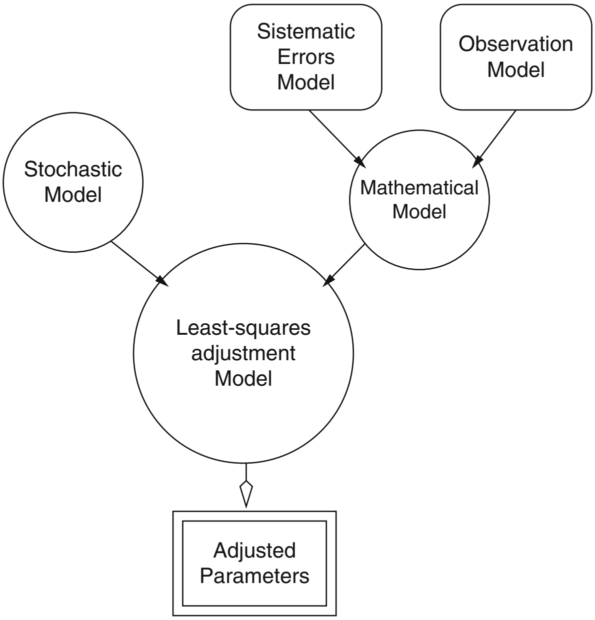


Fig. 8 Basic differential positioning

Fig. 9 Least-squares adjustment elements



An advanced technique to combine observations is known as *bundle*, and it has applications where a very high precision is required in practical measurements (Triggs et al. 2000).

Any least-squares adjust deals with two equally important models: stochastic and mathematical models (Fig. 9).

The n observations are the primary input to the adjustment and are denoted with the vector $\mathbf{L}_b$. The variances of these distributions comprise the stochastic model which gives information about (relative) precision of the observations, accomplished through the variance-covariance matrix of the observations $\mathbf{\Sigma L}_b$. Because in surveying the observations belong to a normal distribution, the vector of observations is a sample from the multivariate normal distribution. If $\mathbf{L}_P$ denotes the vector mean of the population, $\mathbf{L}_b$ can be expressed as:

$$\mathbf{L}_b \sim N\left(\mathbf{L}_p, \Sigma\mathbf{L}_b\right)$$

The adjustment requires either the cofactor matrix $\mathbf{Q}_{\mathbf{L}b}$ of the observations, which corresponds to the scaled variance-covariance matrix:

$$\mathbf{Q}_{\mathbf{L}b} = \left(1/\sigma_0^2\right) * \Sigma\mathbf{L}_b$$

or the weight matrix $\mathbf{P}$:

$$\mathbf{P} = \mathbf{P}_{\mathbf{L}b} = \mathrm{Q}_{\mathbf{L}_b}^{-1} = \sigma_0^2 \mathbf{\Sigma}_{\mathbf{L}_b}^{-1}$$

The scale factor σ_0^2 is called the a priori variance of unit weight.

The original observations are modified to yield the so-called model observation. In most surveying applications, the mathematical model is *nonlinear*. If the observations are explicitly related to the parameters, such as in:

$$\mathrm{F}(\mathbf{X}_a) = \mathbf{L}_a$$

where $\mathbf{X}_a$ are the u adjusted parameters (unknowns) and $\mathbf{L}_a$ is the vector of n adjusted observations. This is the *observation equation model*, and it has particular advantage that each observation generates one equation, allowing its implementation seamlessly on a computational algorithm.

If $\mathbf{X}_0$ is a vector of known approximated values of the parameters, then the parameter corrections $\mathbf{X}$ are:

$$\mathbf{X} = \mathbf{X}_a - \mathbf{X}_0.$$

If $\mathbf{L}_b$ denotes the vector of observations, then the residuals are defined by:

$$\mathbf{V} = \mathbf{L}_a - \mathrm{L}_b$$

and using both previous equations, the mathematical model can be written as

$$F(\mathbf{X_0} + \mathbf{X}) = \mathbf{L}_b + \mathbf{V}.$$

This nonlinear mathematical model can be linealized around a known point of expansion ($\mathbf{X_0}$) giving:

$$\mathbf{L}_b + \mathbf{V} = F(\mathbf{X}) + \partial\left[\frac{F(\mathbf{X})}{\partial\mathbf{X}}\right] dx\Big|_{\mathbf{X}=\mathbf{X_0}}$$

If $F(\mathbf{X})|_{\mathbf{X}=\mathbf{X_0}} = \mathbf{L_0}$, that is, the value of the observations from the approximated parameters $\mathbf{X_0}$ (the point of expansion), and every partial derivative

$$\partial\left[\frac{F(\mathbf{X})}{\partial\mathbf{X}}\right] dx\Big|_{\mathbf{X}=\mathbf{X_0}}$$

is expressed as the product of the design matrix $\mathbf{A}$ by the parameter matrix $\mathbf{X}$, then:

$$\mathbf{L}_b + \mathbf{V} = \mathbf{L_0} + \mathbf{AX}$$

which finally gives

$$\mathbf{V} = (\mathbf{L_0} - \mathbf{L}_b) + \mathbf{AX} = \mathbf{L} + \mathbf{AX}$$

where the difference $\mathbf{L}$ is often called the misclosure.

The previous linealized equations are called observation equations and can be written as

$$ {}_{n}\mathbf{V}_{1} = {}_{n}\mathbf{A}_{uu}\mathbf{X}_{1} + {}_{n}\mathbf{L}_{1} = \mathbf{0} $$

The least-square estimate $\widehat{\mathbf{X}}\,\mathbf{E}$ is based on the minimization of the function $\mathbf{V}^{\mathrm{T}}\mathbf{PV}$. A solution is obtained by introducing a vector of *Lagrange multipliers*, λ, and minimizing the function:

$$ \Phi(\mathbf{V},\lambda,\mathbf{X}) = \mathbf{V}^{\mathrm{T}}\mathbf{PV} - \mathbf{2}\lambda^{\mathrm{T}}(-\mathbf{V} + \mathbf{AX} + \mathbf{L}) $$

The solution starts by noting that P is a squared matrix and can be inverted. Thus, the expression for the residuals is:

$$ -\mathbf{V} = \mathbf{P}^{-1}\lambda $$

and the solution for the Lagrange multiplier is:

$$ \lambda = -\mathbf{P}\left(\mathbf{A}\,\widehat{\mathbf{X}}\,\mathbf{E} + \mathbf{L}\right) $$

from where the estimate vector $\widehat{\mathbf{X}}\,\mathbf{E}$ follows

$$ \widehat{\mathbf{X}}\,\mathbf{E} = -\left(\mathbf{A}^{T}\mathbf{PA}\right)^{-1}\mathbf{A}^{T}\mathbf{PL} $$

Kinematic and Real Time Kinematic Systems

Kinematics positioning is the terminology utilized when the user vector R is changing dynamically (as in a moving vehicle), and therefore the user position vector becomes a function of time, $R(t)$. From Fig. 7 the equation for $R(t)$ is given by:

$$ R(t) = r(t) - \rho(t) $$

In the previous equation, when $R(t)$ changes slowly with respect to all elements then real time kinematics positioning takes place. Practical uses of kinematics positioning are found in GNSS receivers, which display velocity and acceleration of the platform where the receiver is placed.

Real time relative kinematics positioning can also be used when a reference receiver transmits the observed ranges for the very same satellites that are used by the receiver at the mobile platform. The transmission is done by a radio link (e.g., using Radio Technical Commission for Maritime Services, known as RTCM signals). If the message sent by the reference station is similar to those transmitted by the GNSS satellites, the reference station is called a pseudo-satellite or pseudolite.

GNSS Augmentation Systems

DGNSS differential positioning has evolved from local reference stations – which broadcast error correction – to the most actual and modern satellite systems to support GNSS. Such DGNSS satellite systems are named augmentation systems or satellite-based augmentation system (SBAS). The former was used mainly for Maritimes users (RTCM-104 1994) while the latest took the Radio Technical Commission for Maritime Services (RTCM) broadcasting correction techniques to implement pseudo-satellite systems (Wang 2002) which are mainly used in airports.

Nations have undertaken to develop and construct augmentation systems for their own specific territories. Such are the cases of those of the North America's Wide Area Augmentation System (WAAS), for Europe the European Geostationary Navigation Overlay Service (EGNOS), the Indian GPS Aided Geo Augmented Navigation (GAGAN), and the Japanese Multi-functional Satellite Augmentation System (MSAS). Using various kinds of differential positioning, commercial systems such as John Deere's StarFire $^{\text{TM}}$ for precision agriculture offer position accuracy levels ranging from +/− 13 to 1 in. Other services may use cellular phone or satellite telephony as in the case OmniSTAR which can provide up short-term positioning accuracies of 1–2 in. and long-term repeatability of 10 cm (Hofmann-Wellenhof et al. 2008).

International Atomic Time (TAI)

It is common to introduce time as the parameter to describe both the ephemeris and equations of movement of a user in a moving platform, but that also introduces a common mistake: to think time as being a unique quantity for all observers. However, after Albert Einstein's demonstration that time is a relative measurement, there is a need to define a reference time system. First there ought to be a time unit, under *Système international d'unités* (SI) concepts, and such a unit should be the same for any observer in an inertial frame of reference. Nowadays, such standard time unit is defined and measured by modern atomic clocks. The typical atomic clock realization is one that uses a cesium atom standard.

A "cesium atomic clock" is a device that uses as a reference the exact frequency of the microwave spectral line emitted by atoms of the metallic element cesium. A cesium clock operates by exposing cesium atoms to microwaves until they vibrate at one of their resonant frequencies and then counting the corresponding cycles as a measure of time. This frequency provides the fundamental unit of time, which may thus be measured by cesium clocks. Today, cesium clocks measure frequency with an accuracy of from 2 to 3 parts in 10^{14}, i.e., 0.00000000000002 Hz. This corresponds to a time measurement accuracy of 2 ns per day or 1 s in 1,400,000 years. It is the most accurate realization of a standard unit that mankind has yet achieved.

The SI second was defined in terms of the cesium atom in 1967, and in 1971 it was renamed International Atomic Time (TAI, from the French name Temps Atomique International). TAI is a high-precision atomic coordinate time standard based on the notional passage of proper time on Earth's geoid. It is the basis for Coordinated Universal Time (UTC), which is used for civil timekeeping all over the Earth's surface, and for Terrestrial Time, which is used for astronomical calculations. This time references are explained below.

TAI as a time scale is a weighted average of the time kept by over 200 atomic clocks in about 70 national laboratories worldwide. In the 1970s, it became clear that the clocks participating in TAI were ticking at different rates due to gravitational time dilation, and the combined TAI scale therefore corresponded to an average time of various clocks at different altitudes. Starting from 1 January 1977, corrections were applied to the output of all participating clocks, so that TAI would correspond to proper time at mean sea level (the geoid) (see chapter "▶ Introduction and History of Space Remote Sensing"). As of 1 January 2011, TAI was exactly 34 s ahead of UTC (this is the case since 1 January 2009): an initial difference of 10 s at the start of 1972 plus 24 leap seconds in UTC since 1972 (McCarthy and Seidelmann 2009).

Historically, time was measured using Earth's rotation with respect to the Sun. When the Sun is observed from a particular point on the surface of the geoid we get the *true solar time*; however, because days do not last the same all year round, it was more adequate to talk about *mean solar time*. For most civil uses, the solar mean time measured with respect to the meridian at Greenwich is known as Greenwich Mean Time (GMT) and its importance resides in that it helped to define internationally the Universal Time (UT).

It should be clear that measuring time with respect to the Sun is different from the time measured with respect to a more inertial reference frame as that defined by distance stars, and such time is called sidereal time (ST). ST is used by astronomers to locate celestial objects at astronomical ephemeris almanacs.

UT and ST, despite of having different origin, have very slight differences between their respective unit time, and therefore they both properly define the Terrestrial Time (TT).

Modern refined measurements established that the duration of days are affected by as many factors as those that change the moment of inertia of the entire Earth (dynamics of water, land, and wind), making the Earth's rotation to slow down; thus, UT is corrected by leap seconds and such time defines the Universal Time Coordinate (UTC). Such proper name was chosen as an international compromise to agree on a specific name to enclose all types of contemporary reference times coexisting when all sorts of corrections have been considered trying to adjust the TT.

Furthermore and in agreement with Einstein's theory of relativity, when the reference system is not an inertial one, then clocks measure time at different rate; thus, when a system is accelerated, the time unit shrinks or stretches, and, as a result, reference clocks tick differently when they move under acceleration. This effect is noticeable for GNSS because in general, all users' platforms are accelerated at different rates when they are at the surface of the Earth compared to when they are above it and the effect is accentuated when very different heights are involved.

As an example of such relativity effect, and because of both the high-precision time measurements involved and that the GNSS satellites and user's platforms are located at different heights moving at a particular relative velocity, there is the need

Table 2 Time reference frames

Reference	Duration of day			ITU[a] adjustments
	Solar		Sidereal	
Local meridian	Apparent solar time		UT0	
Greenwich meridian	Mean solar time (GMT)		Ephemeris time (ET)	Leap years
International celestial reference frame @ zero meridian	Universal time (UT)		Terrestrial time (TT)	
Geoid *temps atomique international* (**TAI**)	UT1 civil uses		**Universal time coordinate** (**UTC**) Maritime, aerial, computers, etc.	Leap seconds
Corrections				
Phenomena	Tides UT1R	Seasonal UT2	Terrestrial gravitation *Temps-coordonnée géocentrique* TCG	
Phenomena	Both UT2R		Solar system gravitation *Temps-coordonnée barycentrique* (TCB)	

[a] International Telecommunications Union

to consider a correction factor of about 4.45×10^{-10} Hz to their respective clocks (Audion 2001). GLONASS clock is at 5 MHz at ground but in sky is -2.18×10^{-3}.

Sometimes, for scientific purposes it could be convenient to avoid relativity effects due to gravitational field differences, thus measuring time at the center of the Earth which leads to a corrected time named *Temps Coordonnée Géocentrique* (TCG) or Geocentric Coordinate Time (Guinot 1986). Furthermore, when measuring time in other planets, or other space artifacts, then TCG has to be exchanged by another relativistic corrected time, the Barycentric Coordinate Time (TCB). The instant that the gravitational correction started to be applied serves as the epoch for Barycentric Coordinate Time, Geocentric Coordinate Time, and Terrestrial Time. All three of these timescales were defined to read 1 January 1977 00:00:32.184, exactly at that instant. (The offset is to provide continuity with the older Ephemeris Time.)

Table 2 summarizes the diversity of TT.

Nonetheless, the basis for all those time references is the TAI, and therefore there are equations that relate all those different times.

Signal Structure

GNSS satellites broadcast information that is used by GNSS receptors to calculate positions on Earth. Since the information is sent using radio transmission, it is important to consider some basic telecommunications techniques.

First of all, to be able to transmit data information in which long distances are involved, data are modulated in a carrier wave which, as in the GNSS case, is one of the electromagnetic type because the transmission is wireless in open space.

Second, to send digital information in a continuous wave carrier, the modulation technique depends on whether the information changes the amplitude, the frequency, or the phase of the carrier wave (Spilker 1977). These different modulation techniques respectively are: amplitude shift keying (ASK), frequency shift keying (FSK), or phase shift keying (PSK). A practical and simple implementation of a digital modulation is the BPSK when one binary symbol is transmitted using one bit at a time. Other implementations are: (1) QPSK, when two bits are used to transmit the information, and therefore 4 binary symbols are transmitted simultaneously and (2) QAM, a combination of QPSK and ASK when four bits transmit 16 binary symbols, etc.

Third, to access all different satellites in any specific GNSS constellation, there are techniques that use time, frequency, or codes assigned to each satellite which allows distinguishing among all multiple satellites that compose the constellation. Correspondingly, these multiple access techniques are (Lahi and Ding 2009): time division multiple access (TDMA), frequency division multiple access (FDMA), and code division multiple access (CDMA). Other distinguishing possibilities are: (1) the polarization of the traveling carrier wave; (2) orthogonal frequency division multiple access (OFDM), which stresses the orthogonality frequency feature among waves; and (3) wave division multiple access WDMA, which exploits wavelength as a distinction among carriers.

Table 3 summarizes the signal structure related to actual GNSS.

Finally, to recuperate any message that is sent by satellite, there is a format under which a particular full message is to be sent. A practical example can be taken from GPS in which such format (Borre et al. 2007) is formed in a top-down fashion as: a master frame, frames, subframes, and digital words. Such format is shown in Fig. 10.

Carrier Frequencies

Since the carrier wave is an essential part in GNSS, let us consider some useful carrier electromagnetic wave characteristics. With respect to selecting a specific satellite operating communications frequency, when using global communication environments it is essential to respect international telecommunication regulations as well as physical laws for electromagnetic waves. For example, GNSS use the L and C bands (Wither 2000) of the international radio-electrical spectrum since those bands contain the frequencies for mobile communications. And with respect to physical environments, the carrier frequency should be selected taking into account that such frequency is going to travel large distances across the ionosphere, atmosphere, the biosphere, and within cities' concrete canyons, etc. (Bensky 2008).

Table 3 GNSS signal structure

GNSS				GPS			GLONASS			GALILEO			BEIDOU	
Multiple access method				CDMA			FDMA			CDMA			CDMA	
International L band 1–2 GHz subdivisions E, G and B			**Carrier Frequency** [GHz]	Ranging code	Length [chips]	Code-data modulation /speed [MHz]	Ranging code	Length [chips]	Code-data modulation /speed [MHz]	Service	Length 10230 [chips]	Code-data modulation / speed [MHz] 1.023 MHz	Service	Modulation
L1	G1		1602.00				CA/P		BPSK/ 0.511					
	E1	B1-2	1589.74										OS	QPSK/ MBOC
L1c	E1 L1 E2		1575.42	CA/M SPS/PPS	1023	BPSK / 1.023/5.115				OS/ SoL/ PRS/CS	4092/ 8184	BOCc(1,1) data BOC (15,2.5) pilot 1.023 Mcps		
	E2	B1-1	1561.10										OS	QPSK/ MBOC
L2c			1381.05	CA/M		BPSK- BOC /1.023	CDMA modernization							
	E6		1278.75							CS/PRS	✓	BPSK pilot Data 5.115Mcps BOC(10,5)		

	E3	B3	1268.52										OS	MOBC
L2	G2		1246.00				p		BPSK/5.11					
L2			1227.60	C/P(Y)/ M	10230	BPSK/ 0.5115 / 10.23/5.115								
L3	E5 b		1207.14							OS/CS/ S oL		BPSK data Pilot		
	E5		1196.91							OS/CS/ S oL	✓	QPSK/ BPSK-BOC (15,10)d ataI cos pilotQ sin		
L5	E5 a		1176.45	CS/M	10230	BPSK- BOC /10.23				OS/CS/ S oL	✓	Bpsk 10.23		
	E5 a+	B2	1027.14							OS/SoL /SAR	✓	QPSK pilotQ DataI	OS	QPSK
POWER [dBW]				−158 to −150			−167 to −161			−155			NA	

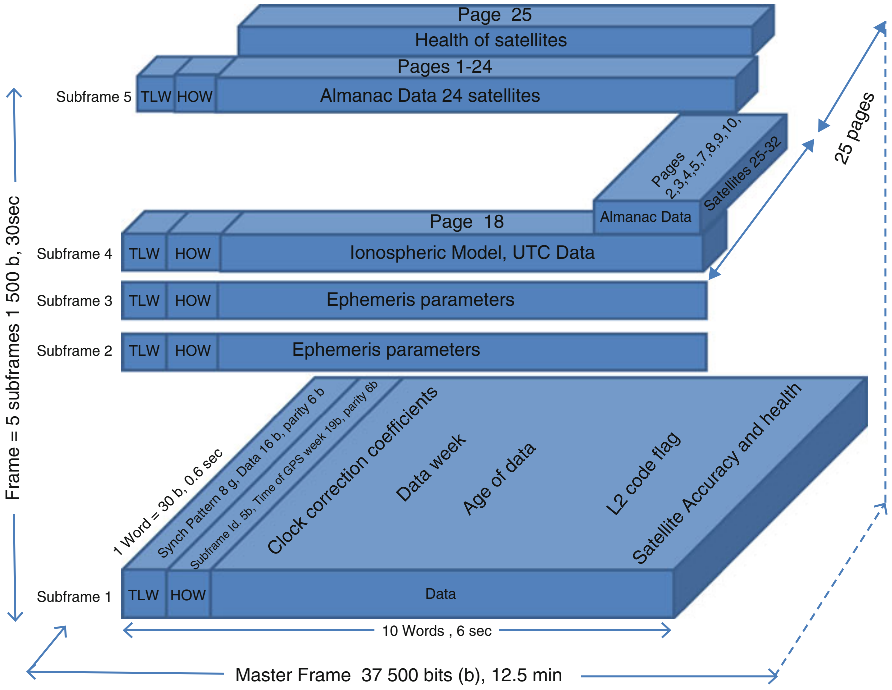


Fig. 10 GPS message format

From the electrical point of view when selecting the operating frequency the design engineer works backward from the given technical specifications, such as speed of information transmission, selected multiple access and modulation techniques, transmission power, and efficient use of the satellite's energy, all this of course within the allocated transmission frequencies for GNSS. It is worth to note that all necessary frequencies employed by satellites are derived from a unique atomic clock which runs in a very well-determined frequency in agreement with all previous described technological communication considerations.

Since high-precision positioning is pursued, there are other frequency design considerations such as relativistic effects, ionosphere effects, multipath trajectories, and even encryption of the information for security reasons. On one hand, two frequencies are normally transmitted by the GNSS to allow corrections for ionospheric effects, and on the other, as the satellite atomic clock runs slower with respect to the master control clock on Earth, a correction is needed to compensate for the different gravitational potential between the satellite and the platform where the user is as well as their relative velocities.

Figure 11 shows the two sinusoidal waveform carriers in the L band in both time and frequency domains.

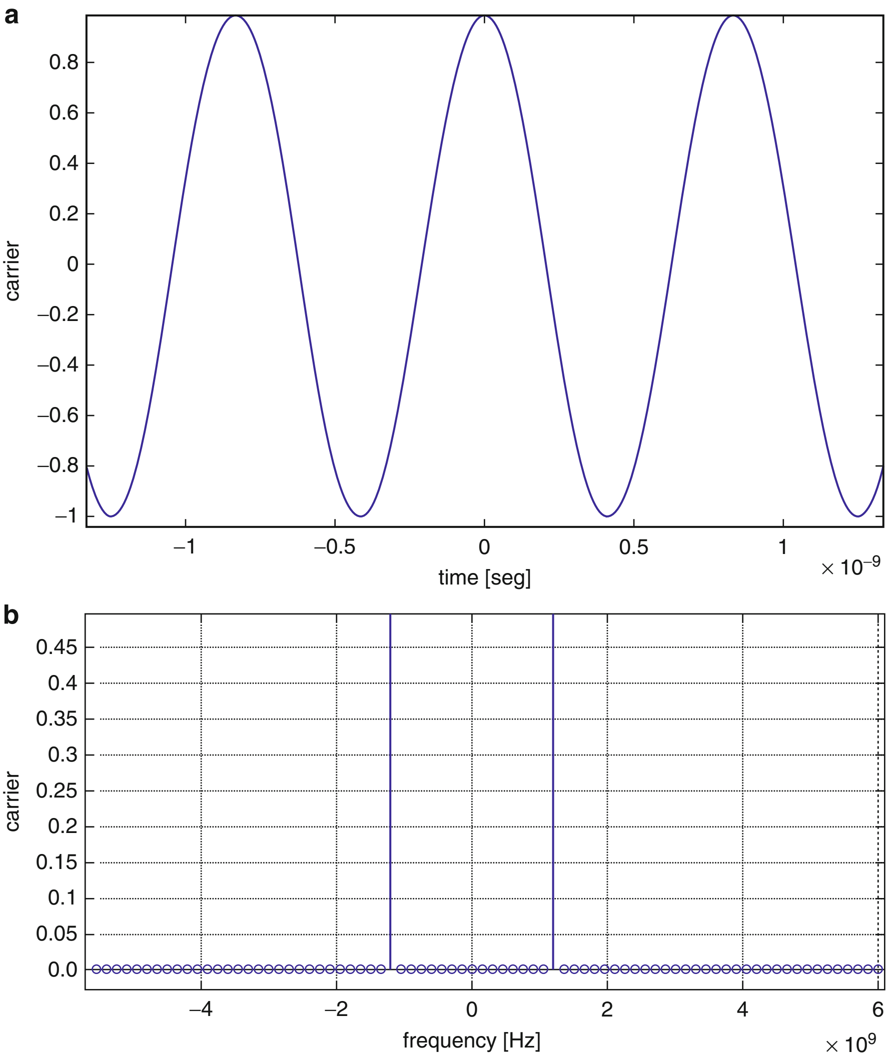


Fig. 11 Carriers in L band: (**a**) time domain and (**b**) frequency domain

Data Stream and Messages

Once the carrier is selected, as described before, then the data stream and the message format should be defined. Data stream refers to both transmission speed and the way data is organized, while the message refers to all data needed for a GNSS receiver to calculate the position of a user.

The speed of data transmission affects directly the time at which anyone receptor performs a position fix for the first time. Remembering that at least four satellites are needed for GNSS positioning, the specific multiple access technique selected affects directly the transmission speed of data which forces the designer to make a compromise between the specific radio communication techniques and the amount of information needed to calculate the user's position.

In Table 4 we can find data transmission speeds for the main GNSS.

At the average data speed of 50 bps, it is common, for a receiver with the current processing technology, to take 12 min to calculate the user's position for the very first time.

With respect to all existing different GNSS messages, each GNSS has organized its message in a specific structure which basically contains information that, beside making position fix, timing, and prediction possible, also allows the user to know the satellite constellation ephemeris status and diagnosis, quality and availability of data for a given specific receiver location and data necessary to initiate acquisition of GNSS signals and for correction of errors due to clock and ionosphere.

For example the GPS message follows the format described in Fig. 10

Multiple Access and Pseudorandom Codes

It was mentioned that to distinguish and access each and all satellites in a specific GNSS constellation, a multiple access communication technique is needed. GPS uses CDMA to access all different satellites in its constellation and GLONASS uses FDMA. Thus, GLONASS only needs to use a single modulation to transmit all useful information but, if GLONASS wants to encrypt its message, an encryption code should be defined.

Encryption codes or CDMA codes are generally binary and generated using different amounts of bits which defines automatically the length of the code. Bits used to generate codes in fact do not transmit information and that is why those bits are rather called chips. Either by using a code to create a CDMA or by using a code for encryption, both produce the same effect on a binary modulated signal, that is, both spread the total power radiated by the satellite (Dixon 1976). Despite of this spreading of power across the spectrum density of the signal still, spread spectrum systems (SS) keep the spectrum constrained to the length and speed at which the code is transmitted. CDMA codes or encryption codes receive the generic name of direct sequences (DS).

For an observer without the DS, the spread GNSS signal will look just like a sinusoidal wave embedded in noise and such apparent noise would have specific statistics. Since the DS code is digitally generated with a given amount of chips, the code will repeat itself, and therefore the apparent noise receives the name of pseudorandom noise (PRN). Pursuing a low auto correlation code sequence is the purpose of the gold codes random processes (Proakis 2008).

Different codes can be used for a given GNSS and named after the purpose for which they were defined. For example, for the GPS, a code for CDMA is called C/A

Table 4 GNSS messages

GNSS				GPS		GLONASS		GALILEO		BEIDOU	
Multiple access, method				CDMA		FDMA		CDMA		CDMA	
International L band 1–2 GHz Subdivisions E, G, and B			**Carrier Frequency** [GHz]	Service	Data speed [sps]	Service	Data speed [sps]	Service	Data speed [sps]	Service	Modulation
L1	G1		1602.00			CA/P	50				
	E1	B1-2	1589.74							OS	QPSK/ MBOC
L1c	E1 L1 E2		1575.42	CA/M SPS/PPS	50			OS/SoL/ PRS/ CS	250		
	E2	B1-1	1561.10							OS	QPSK/ MBOC
L2c			1381.05	CA/M	100	CDMA modernization					
	E6		1278.75					CS/PRS	1000		
	E3	B3	1268.52							OS	MOBC
L2	G2		1246.00			p	50				
L2			1227.60	C/P(Y)/M	50						
L3	E5 b		1207.14					OS/CS/SoL	250		
	E5		1196.91					OS/CS/SoL	250/50		
L5	E5 a		1176.45	CS/M	100			OS/CS/SoL	50		
	E5 a+	B2	1027.14					OS/SoL / SAR	250/50	OS	QPSK

code or civil applications code while, a proper code for encryption is called Y (formerly P) code. Y code was originally conceived to give the precise positioning service (PPS) which allowed positioning with the full precision available from GPS. While C/A codes are well known and open for all public, Y codes are not.

It is interesting to note that, on one hand, a 1,024 chip length code transmitted at 1.024 Mbps per second will take 1 ms to be transmitted and thus one can expect to acquire the specific satellite transmitting such code in about that time, while a code that is 16,777,216 chips long or more will take at least 1 week to be detected. The latter is the case for military or restricted codes; furthermore, Y codes are changed every week making it almost impossible to decode.

There is another advantage of using DS coding, and that is it makes it possible to calculate the range between two communications points (Zigangirov 2004). To do so, by knowing the encrypted or the CDMA codes, all that GNSS receivers have to do is to correlate a locally generated code with the arriving spread signal and the resulting time for correlation will give the pseudo-range information (see Pseudo-Range below). The correlation action basically unspreads the signal as well and, therefore, reconstructs the original information signal.

GALILEO has defined other spread spectrum technique which is intended to reduce interference between the BPSK modulated based CDMA signals. This new technique receives the name of binary offset code (BOC) and, because it uses a single carrier but with a difference in phase of 90°, it produces two subcarriers the BOCsine and BOCcosine which simultaneously makes it possible to permit two different types of services. All codes and their respective services for GNSS can also be found in Table 3.

Modulation Techniques

With respect to modulation, BPSK is the most common realization in GNSS; however, Galileo uses BOC modulation which is a modified version technique. Figure 12 depicts both modulations. Specific modulation techniques for each GNSS are also given in Table 3.

True information bits are sent in the carrier wave as well of the corresponding chips for the DS code which are mixed under the XOR (bitwise eXclusive OR) binary operation with the purpose of recuperating easily the message at the receiver end. Thus, the signal received at the users end is:

$$r(t) = d(t)\sqrt{2P}\cos\left(2\pi f t + \vartheta_0\right)c(t-\tau) + \nu(t)$$

where $d(t)$ is the modulating data, P is the carrier power, f is the instantaneous frequency at the receiver antenna including the Doppler shift, $c(t)$ is the DS ranging code, τ is the time for correlation, $\nu(t)$ is the noise added to the original signal, and ϑ_0 is the initial carrier phase.

Figure 13 shows a GNSS signal modulated under the information and DS code in the time domain and its typical power.

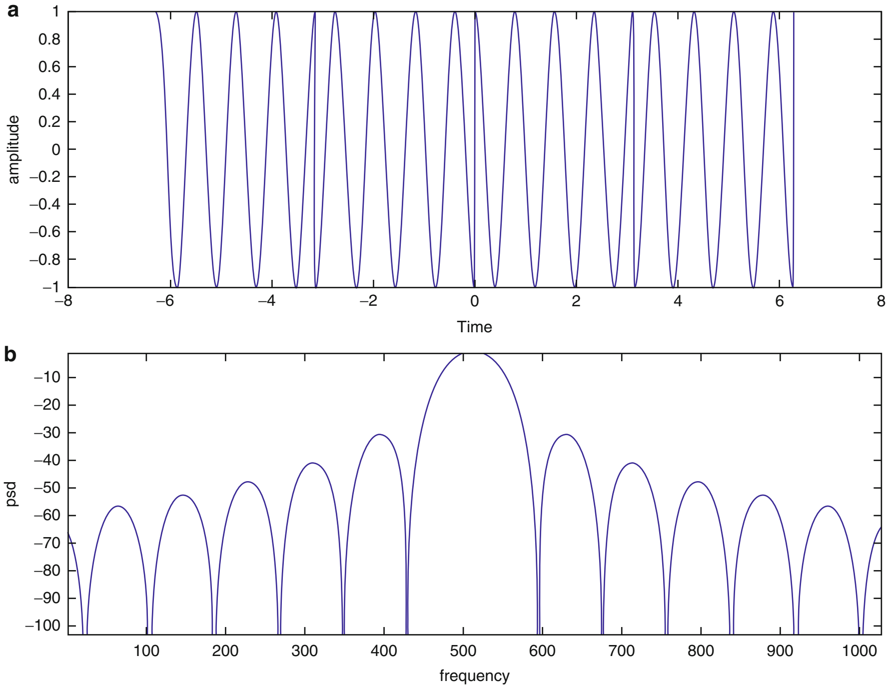


Fig. 12 BPSK and the BOC modulation: (**a**) time domain; (**b**) power spectrum density

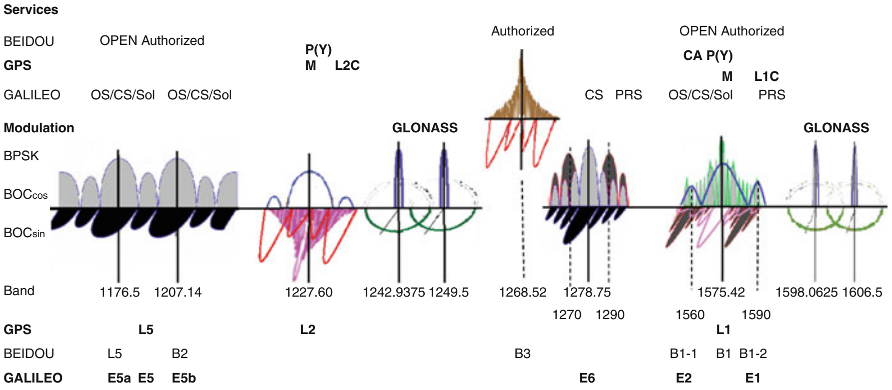


Fig. 13 GNSS signals

Pseudo-Range

In the origins of GPS, to have access to the PPS, it was necessary to have special receivers and the corresponding military permit. Such a restriction challenged some researchers who wanted to benefit science by using the most advanced and high-precision positioning GNSS system available in contemporary times. Their challenging task was to calculate precise ranges to GNSS satellites without knowing any of the GNSS message at all.

Remember that because GNSS receivers perform a correlation to spread out GNSS signals, a by-product of such operation exists: the correlation time. In fact correlation time is exactly the time that it takes the signal to travel from the satellite to the receiver. Correlation time multiplied by the velocity of the carrier wave gives the range to the emitting satellite. However, since satellite and receiver operate in different time references, such difference introduces a bias into the measurements. When using this biased correlation time to calculate the range then biased range is named pseudo-range (Samama 2008).

Pseudo-range provides the possibility to calculate the position of the receiver without the navigation message because by knowing the correlation time and either the period of the carrier or the period of the PRN code, it is easy to calculate the number of cycles that would fit in the pseudo-range. In either case, carrier or code tracking are the corresponding observed variables commonly known as observations.

On one hand, observations may have cycle slips during the correlation process producing an error in their respective pseudo-range creating an ambiguity term (Strang and Borre 1997). Ambiguity adds to the total error budget in GNSS. On the other hand, carrier phase measurements are affected by the Doppler shift effect which can be estimated by calculating how the pseudo-range carrier observation varies with time, and this rate of change gives the so-called accumulated delta range (Draganov 2006). Range measurement is improved when calculated from this observation.

Ranges can be estimated using stationary observations by the Weiner estimator, but Kalman estimation is more adequate when the delta ranges observations are used (Mendel 1995).

Sources of Error and Error Budgets

We have seen that errors are introduced in the observations. There are two kinds of errors, the systematic and the properly stochastic errors. Systematic errors just produce bias; however, stochastic errors make the positioning a process that can only give estimates of the exact user position.

Bias errors come from different sources: from the space segment (Clock bias, ephemeris errors), from the propagation channel (atmospheric refraction, both by the ionosphere and troposphere, multipath due to reflection from fixed obstacles also called imaging), or from the user segment (clock bias and phase center variation) (Leick 2004).

Systematic or bias errors can be eliminated or greatly reduced by using one or a combination of the differential positioning techniques (Muñoz et al. 2009). When observations are made simultaneously, electronic signals can be considered to have the same epoch, and thus clock bias can be assumed to be uncorrelated. By doing differential positioning among two receivers and one satellite, it is possible to cancel these specific space segment biases. Meanwhile, doing the counterpart differential positioning, that is, among two satellites and one receiver cancels the specific user segment bias. A combination of both differential modes cancels most bias errors.

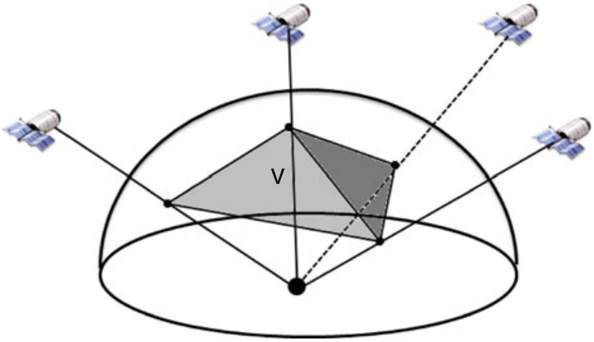
Bias error can be modeled and introduced into the observations equations. As an example, when several observations are carried out for the same point and made at different times, which implies that the corresponding signals have different epochs, then errors can be assumed correlated and easily modeled (Seeber 2003). Because advanced research in GNSS allowed clock, ephemeris, and atmosphere bias to be modeled, another way to reduce bias effects is the use of augmentation systems and broadcasting correction information.

It is very important to notice that since ranging is practically a geometry calculation, the actual and specific geometric arrangement between the satellite segment and the user segment impacts directly in the calculations. The error introduced by this geometry is given by the dilution of precision (DOP) (Bar-Shalom et al. 2001). The meaning of DOP can be easily understood based on Fig. 14.

Despite the fact that GNSS signals may be modulated with the specific purpose of reducing interference or/and jamming, there still a great dependence on the total power delivered onto the user receiver. Nowadays due to the presence of many electromagnetic fields in practical real life, undesired noise is present during observations affecting the signal to noise ratio, the carrier to noise power density ratio, and may favors jamming or interference.

Thus, true stochastic errors may include: residual biases, randomly multipath imaging due to moving obstacles or to the kinematics of the receiver platform, cycle slips due to the stochasticity of the receiver's way of detection, and random observation errors.

Fig. 14 DOP



DOP is proportional to the volume V

position accuracy = *GDOP* × measurement accuracy

GDOP is the total *DOP* both in time and space

poor GDOP when V is small ⇔ satellites are very close together
good GDOP when V is largest ⇔ 3 satellites on horizon
and the other overhead

Remember, however, that accumulated delta range is only affected by bias errors, and therefore it is the technique that gives the most precise positioning calculation.

At the end, sources of error may vary depending on the specific observation and the technique used to reduce noise effects. When using a given modulation technique with the purpose of reducing interference and jamming, there are two advantages: errors become uncorrelated and independent of the observation point. Thus, both advantages simplify the calculation of the total error because now the total squared error is just the sum of the squared errors. Nevertheless, such calculations ought to be done for the specific user segment and receiver and the differential technique used.

Conclusion

GNSS have been presented as space based systems that cover most of the space-time around the Earth. Such systems have also extended capabilities through augmentation systems to send electronic signals that allow position fixing for a single user or a network of users with an extraordinary precision.

Errors were described so one can easily calculate what will be the total error involved in a particular observation or surveying campaign and have a figure of positioning quality.

Cross-References

- ▶ Current and Future GNSS and Their Augmentation Systems
- ▶ International Committee on GNSS
- ▶ Introduction to Satellite Navigation Systems

References

C. Audion, *The Measurement of Time: Time, Frequency and The Atomic Clock* (Cambridge University Press, Cambridge, 2001)

Y. Bar-Shalom, X. Rong-Li, T. Kirubarajan, *Estimation with Applications to Tracking and Navigation* (Wiley, New York, 2001)

A. Bensky, *Wireless Positioning Technologies and Applications* (Artech House, Boston, 2008)

G. Beutler, *Methods of Celestial Mechanics*, vol. I (Springer, Berlin, 2005)

K. Borre et al., *A Software Defined GPS and Galileo Receiver* (Birkhauser, Boston, 2007)

J. Carnes, *UTM Using your GPS with the Universal Transverse Mercator Map Coordinate System* (MapTools, Woodside, 2007)

H.C. Chen et al., The performance comparison between GPS and BeiDou-2/COMPASS: a perspective from Asia. J. Chin. Inst. Eng. **32**(5), 679–689 (2009)

China Satellite Navigation Project Centre, COMPASS/Beidou navigation satellite system development. 3rd meeting of the international committee on GNSS, Pasadena, 8–12 Dec 2008

R.C. Dixon, *Spread Spectrum Systems* (Wiley, New York, 1976)

A. Draganov, *GPS Accumulated Delta Range Processing for Navigation Applications.* United States Patent Application Publication. U.S. 2006/0195262 A1, 2006

Global Navigation Satellite System: GLONASS, *Interface Control Document* (Coordination Scientific Information Centre, Moscow, 1998)

B. Guinot, Is the international atomic time TAI a coordinate time or a proper time? Celest. Mech. **28**, 155–161 (1986)

B. Hofmann-Wellenhof, H. Lichtenegger, E. Wasle, *GNSS Global Navigation Satellite Systems* (Springer, Wien/New York, 2008)

B.P. Lahi, Z. Ding, *Modern Digital and Analog Communication Systems*. Oxford Series in Electrical and Computer Engineering (Oxford University Press, Oxford, 2009)

A. Leick, *GPS Satellite Surveying*, 3rd edn. (Wiley, Hoboken, 2004)

D.D. McCarthy, P.K. Seidelmann, *Time: From Earth Rotation to Atomic Physics* (Wiley, Weinheim, 2009)

J. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications, and Control* (Prentice-Hall, Englewood Cliffs, 1995)

G. Mohinder, W. Lawrence, A. Angus, *Global Positioning Systems, Inertial Navigation and Integration* (Wiley, New York, 2001)

D. Muñoz et al., *Position Location Techniques and Applications* (Elsevier, Amsterdam, 2009)

J.G. Proakis, *Digital Communications*, 4th edn. (McGraw-Hill, New York, 2008)

S. Revnivykh, GLONASS sattus and progress. 3rd meeting of the international committee on GNSS (ICG)/UNOOSA, Pasadena, 8–12 Dec 2008

R. Rogers, *Applied Mathematics in Integrated Navigation Systems* (AIAA, Reston, 2003)

RTCM-104, *Recommended Standards for Differential NAVSTAR-GPS Services. V 2.1* (Radio Technical Commission for Maritime Services, Washington, DC, 1994)

N. Samama, *Global Positioning* (Wiley, Hoboken, 2008)

G. Seeber, *Satellite Geodesy* (Humbert, New York, 2003)

M. Shaw et al., U.S. space-based PNT, policy and program review. 3rd meeting of the international committee on GNSS (ICG)/UNOOSA, Pasadena, 8–12 Dec 2008

J.J. Spilker, *Digital Communications by Satellite* (Prentice-Hall, Englewood Cliffs, 1977)

G. Strang, K. Borre, *Linear Algebra, Geodesy, and GPS* (Wellesley-Cambridge, Wellesley, 1997)

B. Triggs et al., Boundle adjustment, a modern synthesis. Calculations, in *Vision Algorithms: Theory and Practice*. Lecture Notes in Computer Science (LNCS), vol. 1883 (Springer, Berlin, 2000), pp. 298–372

P. Verhoef, European GNSS programmes Galileo and EGNOS. 3rd meeting of the international committee on GNSS (ICG)/UNOOSA, Pasadena, 8–12 Dec 2008

J. Wang, Pseudolite applications in positioning and navigation: progress and problems. J. Glob. Position. Syst. **1**(1), 48–56 (2002)

D. Wells, *Guide to GPS Positioning* (Canadian GPS Associates, Fredericton/New Brunswick, 1987)

A. Wilson, *The First Galileo Satellites. Galileo in-orbit Validation Element: GIOVE* (ESA publications Division, Noordwijk, 2006)

D.J. Wither, *Radio Spectrum Management*. IEE telecommunications Series, vol. 45 (The Institution of Electrical Engineers, London, 2000)

G. Xu, *GPS Theory, Algorithms and Applications* (Springer, Berlin, 2003)

K.S.H. Zigangirov, *Theory of Code Division Multiple Access Communication* (Wiley, Hoboken, 2004)

International Committee on GNSS

Sergio Camacho-Lara and Joseph N. Pelton

Contents

S. Camacho-Lara (✉)
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Santa María Tonantzintla, Puebla, México
e-mail: sergio.camacho@inaoep.mx; sergiocamacho99@yahoo.com

J.N. Pelton
International Space University Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_12

Abstract

Global Navigation Satellite Systems (GNSS), with their extremely high accuracy, global coverage, all-weather operation, and usefulness at high velocities, are a dual-use technology that are becoming a new global utility that increasingly improve people's daily lives. GNSS applications are growing, and their quality is improving in such areas as aviation, maritime and land transportation, mapping and surveying, agriculture, power and telecommunication networks, disaster warning and emergency response, and a host of commercial and social applications.

At the turn of the millennium, it became apparent that the two Global Navigation Satellite Systems that had existed, the Global Positioning Systems (GPS) of the United States and the Global Navigation Satellite System (GLONASS) of the Russian Federation, would soon be joined by the Galileo system of Europe and the Compass/BeiDou of China, as well as by the regional Quasi-Zenith Satellite System (QZZS) of Japan and the Indian Regional Navigation Satellite System (IRNSS) of India. The emergence of new GNSS and regional augmentations focused attention on the need for the coordination of program plans among current and future operators in order to enhance the utility of GNSS services. It also made clear that the providers of GNSS services should pursue greater compatibility and interoperability among all current and future systems in terms of spectrum, signal structures, time, and geodetic reference standards to the maximum extent possible.

Although coordination between the providers of the GNSS was already taking place on a bilateral basis, the desirability of having a forum in which all GNSS providers participated became an attractive idea. Such a forum would allow discussion and coordination on issues of common interest such as protection of the radio navigation spectrum from disruption and interference, global compatibility, and interoperability of space-based position, navigation, and timing services (PNT) that could be used separately or together without interfering with each other. After 1999, and following several years of discussing terms of reference, objectives, and work plan, the International Committee on GNSS (ICG) became such a forum.

Keywords

Argos • Chinese navigation satellite system (Compass/BeiDou) • Compatibility and interoperability of GNSS • European Geostationary Navigational Overlay System (EGNOS) • European satellite navigation system (Galileo) • Global Navigation Satellite System (GNSS) • Global Positioning System (GPS) Satellite • Global Navigation Satellite System (GLONASS) • GPS-Aided GEO-Augmented Navigation System (GAGAN) • Indian Regional Navigation Satellite System (IRNSS) • International committee on GNSS (ICG) • Japanese Multi-functional Satellite Augmentation System (MSAS) • Japanese regional navigation satellite system; Quasi-Zenith Satellite System (QZSS) • Position, navigation, and timing services (PNT) • UNISPACE III United Nations Committee on the Peaceful Uses of Outer Space (COPUOS) • Wide Area Augmentation System (WAAS)

Introduction

Satellite navigation builds upon terrestrial-based radio navigation that has been used by aviation and shipping over the past 100 years. For a number of years, navigation satellite systems were limited to the Joint United States-France Argos Satellite System, the US Navstar GPS Satellite System, and the Union of Soviet Socialist Republics/Russian Federation GLONASS systems. Today the number, capability, and complexity of the Global Navigation Satellite Systems have doubled and continue to expand.

Navigation satellites broadcast signals are used by a receiver to determine exactly the receiver's position, velocity, and precise time worldwide. User receivers of satellite navigation signals measure the distance of the receiver equipment to the satellite using a technique called "passive ranging." In this technique, the distance to each satellite is derived from the measurement of the time the navigation signal needs to travel from the satellite to the receiver. The three-dimensional position of the receiver can be calculated if signals from at least three satellites are available. The signal from a fourth satellite is used to avoid the need for a precise atomic clock at the receiver.

Standard GNSS signal processing provides around 100-m accuracy at the location of the receiver with four satellites in view, while precision signal processing provides around 20–10-m accuracy. Reference stations make differential GNSS (DGNSS) services possible providing higher time and position accuracies. If, in addition to the signals from the satellites, a user receiver also receives the signal of a ground-based reference station, the accuracy at the location of the user receiver is around 1 m. If, in addition, a space-based augmentation system is used as a reference station, the position accuracy increases to the order of centimeters or less.

For the everyday user of the GNSS signal, without interoperability of the systems, the accuracies that can be obtained are limited by the number of satellites of one GNSS that the receiver equipment can view. For mountainous terrain and for urban settings, the number of satellites that can be viewed is likely to be limited as there might not be a direct line of view of the GNSS receiver to one or more of the satellites. This is often the case in canyons or among tall buildings. When full interoperability is achieved between the four Global Navigation Satellite Systems and the two regional systems, receivers in mountainous or urban areas will be able to view a large number of satellites at the same time leading to very high accuracies in position and timing measurements. This will be a spectacular result of the work being carried out by the ICG.

Establishment of the ICG

Following the Third United Nations Conference on the Exploration and Peaceful Uses of Outer Space (UNISPACE III), held in 1999, in its resolution 54/68, the United Nations General Assembly endorsed the "Vienna Declaration: Space Millennium for Human Development" (Report of the Third United Nations Conference 1999). The Vienna Declaration called for action, among other matters, to improve the efficiency and security of transport, search and rescue, geodesy, and other activities by promoting the enhancement of, universal access to, and compatibility of space-based navigation and positioning systems. In response to that call, in 2001 the United Nations Committee on the Peaceful Uses of Outer Space (COPUOS) established the Action Team on Global Navigation Satellite Systems (GNSS) to carry out those actions under the chairmanship of Italy and the United States.

The Action Team on GNSS, consisting of 38 member states and 15 intergovernmental and nongovernmental organizations, recommended, among other things, that an International Committee on GNSS (ICG) should be established to promote the use of GNSS infrastructure on a global basis and to facilitate exchange of information. The United Nations Committee on the Peaceful Uses of Outer Space included this recommendation in the Plan of Action proposed in its report to the General Assembly on the review of the implementation of the recommendations of UNISPACE III (Report of the Action Team 2004).

In 2004, in its resolution 59/2, the General Assembly of the United Nations endorsed the Plan of Action. In the same resolution, the General Assembly invited GNSS and augmentation system providers to consider establishing an ICG in order to maximize the benefits of the use and applications of GNSS to support sustainable development.

At the "United Nations International Meeting for the Establishment of the International Committee on Global Navigation Satellite Systems (ICG)" held on December 1–2 2005, in Vienna, Austria, the ICG was established on a voluntary basis as an informal body for the purpose of promoting cooperation, as appropriate, on matters of mutual interest related to civil satellite-based positioning, navigation, timing, and value-added services, as well as compatibility and interoperability among the GNSS systems, while increasing their use to support sustainable development, particularly in the developing countries. This fact was noted with appreciation by the General Assembly in its resolution 61/111, of December 14, 2006.

In establishing the ICG, in 2006, the representatives of GNSS core system providers, GNSS augmentation providers (see chapter ▶ "Current and Future GNSS and Their Augmentation Systems" for a description of the satellite systems), and the international organizations primarily associated with the use of GNSS identified the overlap of GNSS mission objectives and the interdisciplinary nature of applications of GNSS services. Aware that the complexity and cost of user equipment should be reduced whenever possible, the founders of the ICG considered the need to protect the investment of the current user base of GNSS services through the continuation of existing services, particularly thorough greater compatibility and interoperability among all current and future GNSS systems in terms of spectrum, signal structures, time, and geodetic reference standards.

With the above understandings, the representatives of GNSS core system providers, GNSS augmentation providers, and the international organizations primarily associated with the use of GNSS developed terms of reference for the International Committee on GNSS in which they agreed on the objectives of the ICG, its participants (members, associate members, and observers), procedures of work, structure, and organization. The terms of reference that were adopted at its meeting

Fig. 1 The United Nations Office for Outer Space Affairs, Executive Secretariat of the International Committee on GNSS, is located at the Vienna International Centre, Vienna, Austria (Courtesy of the United Nations Office at Vienna)

in 2008 also provided for their revision on the basis of proposals made by members or associate members and adopted by consensus of the members.

The structure of the ICG consists of a chairperson, a plenary session of the Committee, an executive secretariat, and working groups. The chair will rotate on an annual basis among the members and associate members. The Office for Outer Space Affairs, part of United Nations Office at Vienna, is the Executive Secretariat for the ICG and its Provider's Forum (Fig. 1).

Membership of the ICG

The participants in the ICG are the following governments, organizations, and associations:

(a) Members:
Current and future core system providers, including China (Compass/BeiDou Navigation Satellite System (CNSS)), the European Union (European Satellite Navigation System (Galileo)), the Russian Federation (Global Navigation Satellite System (GLONASS)), and the United States (GPS Satellite); member state of the United Nations with an active program in implementing or promoting a wide range of GNSS services and applications (Italy, Malaysia, United Arab

Emirates); current and future space-based regional or augmentation system providers including, for example, the European Space Agency (European Geostationary Navigation Overlay Service (EGNOS)), India (GPS and Geostationary Augmented Navigation System (GAGAN) and Indian Regional Navigation Satellite System (INRSS)), Japan (Multi-functional Transport Satellite (MTSAT) Satellite-Based Augmentation System (MSAS) and Quasi-Zenith Satellite System (QZSS)), Nigeria (Nigerian Communication Satellite Space-Based Augmentation System (NigComsat-1 SBAS)), the Russian Federation (System of Differential Corrections and Monitoring (SDCM)), and the United States (Wide Area Augmentation System (WAAS))

(b) Associate members:
International and regional organizations and associations dealing with GNSS services and applications, including the Office for Outer Space Affairs of the United Nations Secretariat, the Civil GPS Service Interface Committee (CGSIC), the International Association of Geodesy (IAG), the International Association of Geodesy Reference Frame Sub-Commission for Europe (EUREF), the International Cartographic Association (ICA), the International GNSS Service (IGS, formerly International GPS Service), the International Society for Photogrammetry and Remote Sensing (ISPRS), the International Earth Rotation and Reference Systems Service (IERS), the Fédération Internationale des Géomètres (FIG), the European Position Determination System (EUPOS), and the International Astronautical Federation

(c) Observers:
The Committee on Space Research (COSPAR), the Bureau International des Poids et Mesures (BIPM), the International Association of Institutes of Navigation (IAIN), the Union Radio-Scientifique Internationale (URSI), the International Telecommunication Union (ITU), and the Interagency Operations Advisory Group (IOAG)

Objectives of the ICG

The objectives of the ICG are to:

(a) Benefit users of GNSS services through consultations among members of the ICG
(b) Encourage coordination among providers of GNSS core systems and augmentations in order to ensure greater compatibility and interoperability
(c) Encourage and promote the introduction and utilization of satellite positioning, navigation, and timing services, particularly in the developing countries through assistance with the integration of GNSS services into their infrastructures
(d) Assist both the members of the ICG and the international user community by, inter alia, serving as the focal point for international information exchange related to GNSS activities, respecting the roles and functions of GNSS service providers and intergovernmental bodies such as the International

Telecommunication Union (ITU), the International Civil Aviation Organization (ICAO), and the International Maritime Organization (IMO)

(e) Better address future user needs in the GNSS development plans and applications
(f) Report periodically on its activities to the Committee on the Peaceful Uses of Outer Space

The participants in the ICG agreed that these objectives will be accomplished by an indicative work plan that would be reviewed at meetings of the ICG for accomplishments and updating as might be desirable or necessary.

Work of the ICG

The ICG meets least once every year in plenary session. Meetings of the ICG are organized by the designated host, assisted by the United Nations Office for Outer Space Affairs which acts as the Executive Secretariat of the ICG.

The ICG may establish, as mutually agreed and on an ad hoc basis, working groups to investigate specific areas of interest, cooperation, and coordination. The chairpersons of such working groups report at each plenary session on accomplishments and future plans.

All recommendations of the ICG or its working groups are decided on the basis of consensus of its members, do not create legal obligations, and will be acted upon at the discretion of each member, associate member, or observer.

The ICG's indicative work plan contains the following broad scope elements:

(a) Compatibility and interoperability
(b) Enhancement of performance of GNSS services
(c) Information dissemination
(d) Reference frames, timing, and applications
(e) Coordination

Working Groups of the ICG

In order to carry out the work associated with items (a) to (d) of the above listed elements throughout the period between meetings of the ICG, the Committee decided to establish Working Groups A–D, assigning to them specific tasks. The Working Groups report and make recommendations to the ICG on a yearly basis. Recommendations accepted by consensus in the ICG can lead to additional items on the work plan of the Working Groups. The initial work plans of the Working Groups and the highlights and modifications to their work plans as agreed at the sixth meeting of the ICG, held in Tokyo, Japan, in 2011, are presented below.

ICG Working Group A: Compatibility and Interoperability

As compatibility and interoperability are highly dependent on the establishment of standards for service provision and user equipment, the ICG decided to address the topic of the adoption of common guidelines. However, the ICG would not itself set guidelines; instead, it will identify applications where no guidelines currently exist and recommend possible organizations that could appropriately set new guidelines. As required, consultation with existing standard-setting bodies, such as the International Civil Aviation Organization (ICAO), the International Maritime Organization (IMO), the International Telecommunication Union (ITU), and the International Organization for Standardization (ISO), is carried out.

Among the first actions assigned to Working Group A was the establishment of a Providers Forum to enhance compatibility and interoperability among current and future global and regional space-based systems.

In 2011, the Working Group A on compatibility and interoperability addressed all four areas of its current work plan. Interference detection and mitigation and open service provision and performance monitoring by multi-GNSS networks were the major areas of focus, leading to three of the working group's four recommendations (Report of Working Group A 2011).

The session on multi-GNSS monitoring was held jointly with Working Groups B and D, as was the session on interoperability. This resulted in constructive dialogue with these working groups and an agreed plan of practical steps including establishment of a subgroup to collectively investigate international GNSS monitoring and assessment.

ICG Working Group B: Enhancement of Performance of GNSS Services

As a unique combination of GNSS service providers and major user groups, the ICG is exceptionally placed to promote and coordinate activities aimed at enhancing GNSS performance, recommending system enhancements, and meeting future user needs.

Among the first actions assigned to Working Group B were the development of a reference document on models and algorithms for ionospheric and tropospheric corrections, the examination of the problem of multipath and related mitigation actions affecting both GNSS systems and user receivers, especially for mobile receivers, and the extension of GNSS service to indoor applications.

In 2011, the Working Group B discussed, among other aspects, the dissemination of disaster information. The Working Group concluded that satellite navigation systems may provide essential contributions, but the service concept still needs further elaboration. Due to the importance of this issue, a new work item was introduced in the work plan of the Group on the basis of seven recommendations that were approved by the ICG (Report of Working Group B 2011).

In addition, at the Sixth Meeting of the ICG, the existing actions in the work plan were confirmed, and good progress was shown in various areas of the Working Group's work plan, including indoor positioning, signal authentication, precise positioning, transportation, maritime, and space applications. Since more and more application-related issues are introduced and discussed within Working Group B, it was agreed to form a dedicated subgroup on applications.

ICG Working Group C: Information Dissemination

A great many people, in their work environments and everyday life, already rely on GNSS products. The awareness of the improvements that are resulting both from better technology in receiver equipment and in the signal that is being provided, coupled with greater knowledge on the use the GNSS signal, will improve the benefits that can be derived in a large range of every day work-life areas. To support greater awareness and knowledge of GNSS and of the use of their signals, the ICG is promoting the establishment of user information centers by GNSS providers.

Among the first action assigned to Working Group C was the establishment of a GNSS web information portal drawing on contributions from members, associate members, and observers of the ICG (A GNSS web portal). Working Group C was also asked to consider the use of the Regional Centres for Space Science and Technology Education, affiliated to the United Nations located in Africa (Nigeria and Morocco), Asia and the Pacific (India), and Latin America and the Caribbean (Brazil and Mexico), to promote GNSS use and applications as well as the development of a GNSS curricula that could be introduced at these centers and other institutions of higher education. At the Sixth Meeting of the ICG, held in September 2011 in Tokyo, Japan, the Provider's Forum invited its members to make proposals for an updated design of the interim ICG Information Portal.

In 2011, the Working Group C on information dissemination and capacity building addressed further aspects of its work plan, including training for capacity building in developing countries, distance learning programs, web-based courses and tutorials, interactive programs for middle/high schools, multimedia softwares and demonstration data sets to enrich the training and research, programs promoting the use of GNSS technologies as tools for scientific applications, the International Space Weather Initiative, and regional workshops on applications of GNSS. A new item on education and training programs on GNSS was added to the work plan (Report of Working Group C 2011).

Working Group D: Reference Frames, Timing, and Applications

The ICG is establishing links with national and regional authorities and relevant international organizations, particularly in developing countries. On the basis of these links, the ICG organizes and sponsors regional workshops and other types of activities in order to fulfill its objectives.

Among the first actions assigned to Working Group D was the definition of minimum operational performance standards for GNSS performance monitoring networks; the establishment of working groups focused on (1) Site Quality, Integrity and Interference Monitoring (SQII); (2) developing a strategy for support by the International Committee of regional reference systems (e.g., the African Geodetic Reference Framework (AFREF), the European Position Determination System (EUPOS), the IAG Reference Frame Sub-Commission for Europe (EUREF), and the Geocentric Reference System for the Americas (SIRGAS)); and (3) developing a strategy for support by the ICG of mechanisms to detect and mitigate sources of electromagnetic interference, taking existing regulatory mechanisms into consideration.

In 2011, the Working Group D completed the development of templates describing the geodetic and timing references for the navigation satellite systems currently represented in the ICG. The Working Group proposed that the templates be published on the ICG Information Portal. ICG also welcomed progress by the Bureau International des Poids et Mesures (BIPM) toward the production of the "Rapid Coordinated Universal Time (UTC)" as a more immediately accessible time reference that could be used to better harmonize the UTC broadcast by each GNSS. The Working Group recommended that interested system providers supply data from their respective monitor stations for inclusion in regular processing with the IGS network of reference stations. Such inclusion is aimed at improving the alignment of the various GNSS reference frames with each other and with the International Terrestrial Reference Frame. Working Group D also recommended that the ICG support and endorse the IGS Multi-GNSS Global Experiment (IGS M-GEX) and actively encourage participation and/or contributions from, among others, GNSS providers, international organizations related to GNSS, and entities involved in timing, navigation, aviation, transportation, GIS, and relevant fields, including national mapping agencies, space agencies, universities and research institutions, as well as industry receiver manufacturers and service providers (Report of Working Group D 2011).

The Future of the GNSS International Coordination Process

Because of its membership composition, the ICG brings together the providers of the global and regional navigation satellite systems, the large professional associations of users, and the international organizations that have a mandate to regulate the use of the GNSS spectrum. These are the key actors that need to discuss issues of compatibility and interoperability among the systems and the protection of the GNSS frequency spectrum in order to provide better and more cost-effective services to all.

In the future, the ICG will consider and make recommendations and agree on actions to promote appropriate coordination across GNSS programs. Furthermore, the ICG encourages its members, associate members, and observers to maintain communication, as appropriate, with other groups and organizations involved in

GNSS activities and applications through the relevant channels within their respective governments and organizations.

The ICG could also support the establishment of national and/or regional planning groups for GNSS that would address regulations associated with the use of GNSS services and suggest organizational models to use at the national level for cocoordinating and governing GNSS use.

In addition to the coordination provided by the ICG plenary, high-level coordination is carried out by the Provider's Forum, established at the recommendation of the ICG's Working Group A.

Providers Forum

The Providers Forum was established at the second meeting of the ICG in 2007, in Bangalore, India, with the aim to promote greater compatibility and interoperability among current and future providers of the Global Navigation Satellite Systems (GNSS). The current members of the Providers Forum, including China, India, Japan, the European Union, the Russian Federation, and the United States, addressed key issues such as ensuring protection of GNSS spectrum and matters related to orbital debris/orbit de-confliction.

Global and regional system providers, members of the Providers Forum, at the third meeting of the ICG held in 2008, in Pasadena, United States, agreed that at a minimum, all Global Navigation Satellite Systems (GNSS) signals and services must be compatible. To the maximum extent possible, open signals and services should also be interoperable, in order to maximize benefit to all GNSS users. For many applications, common carrier frequencies are essential to interoperability, and commonality of other signal characteristics is desirable. In some cases, carrier frequency diversity may be preferable to improve performance. The Providers Forum will continue to investigate the benefits of carrier frequency commonality and diversity, as well as of compatibility and interoperability, as these latter terms and desired outcomes are defined below:

Interoperability refers to the ability of global and regional navigation satellite systems and augmentations and the services they provide to be used together to provide better capabilities at the user level than would be achieved by relying solely on the open signals of one system.

1. Interoperability allows navigation with signals from different systems with minimal additional receiver cost or complexity.
2. Multiple constellations broadcasting interoperable open signals will result in improved observed geometry, increase end-user accuracy everywhere, and improve service availability in environments where satellite visibility is often obscured.
3. Geodetic reference frames' realization and system time standards should adhere to existing international standards to the maximum extent practical.
4. Any additional solutions to improve interoperability should be encouraged.

Compatibility refers to the ability of global and regional navigation satellite systems and augmentations to be used separately or together without causing unacceptable interference and/or other harm to an individual system and/or service.

1. The International Telecommunication Union provides a framework for discussions on radiofrequency compatibility. Radiofrequency compatibility should involve thorough consideration of detailed technical factors, including effects on receiver noise floor and cross correlation between interfering and desired signals.
2. Compatibility should also respect spectral separation between each system's authorized service signals and other systems' signals. Recognizing that some signal overlap may be unavoidable, discussions among providers concerned will establish the framework for determining a mutually acceptable solution.
3. Any additional solutions to improve compatibility should be encouraged.

Objectives of the Providers Forum

To achieve the desired compatibility and interoperability of the GNSS, the providers of the systems agreed on their own terms of reference, which have been updated to adapt to evolving circumstances, including a statement of objectives of the Providers Forum, its membership and work procedures, structure, and organization. The objectives of the Providers Forum are to:

(a) Promote compatibility and interoperability among current and future global and regional space-based systems by exchanging detailed information about planned or operating systems and the policies and procedures that govern their service provision, consistent with the template for information sharing among providers that was circulated prior to the first meeting
(b) Act as a mechanism to continue discussions on important issues addressed by the ICG that require focused inputs from system providers

The Providers Forum is not a policy-making body, but it provides a means to promote discussion among system providers based on agreed guidelines for provision of open services, including transparency, cooperation, performance monitoring, and spectrum protection and agreed principles for ensuring compatibility and interoperability among systems.

Results of the Latest Meeting of the ICG

The Tenth Meeting of the International Committee on Global Navigation Satellite Systems known as the (ICG) was organized by the United States Department of State and the University Corporation for Atmospheric Research (UCAR) in Boulder, Colorado, on behalf of the Government of the United States during the period

November 1–6, 2015. The purpose of this meeting was to continue reviewing and discussing developments in Global Navigation Satellite Systems (GNSS) and to allow ICG members, associate members, and observers to address recent developments in their organizations and associations with regard to GNSS services and applications. ICG also addressed relevant challenging issues associated with observing earth processes using GNSS. In association with the Tenth Meeting, there was an associated meeting of providers of GNSS services whose function was described above (ICG Providers' Forum 2015).

The Tenth Meeting was attended by representatives of China, India, Italy, Japan, Malaysia, the Russian Federation, the United Arab Emirates, the United States, and the European Union, as well as the following intergovernmental and nongovernmental organizations: Arab Institute of Navigation (AIN), Asia-Pacific Space Cooperation Organization (APSCO), Civil GPS Satellite Service Interface Committee (CGSIC), European Space Agency (ESA), International Aeronautical Federation (FAI), International Association of Geodesy (IAG) and IAG Reference Frame Sub-Commission for Europe (EUREF), International Association of Institutes of Navigation (IAIN), International Bureau of Weights and Measures (BIPM), International Federation of Surveyors (FIG), and International GNSS Service (IGS). Representatives of the Office for Outer Space Affairs of the United Nations Secretariat also participated. Australia and Canada were invited to attend as observers. The representatives of the Regional Centres for Space Science and Technology Education, affiliated to the United Nations, located in China, Mexico, and Morocco, and the Space Generation Advisory Council attended the meeting (Tenth Meeting of the ICG 2015).

The ICG recalled that the United Nations General Assembly, in its resolution 69/85 of December 16, 2014, had noted with satisfaction the continuous progress made by the ICG toward achieving compatibility and interoperability among global and regional space-based positioning, navigation, and timing systems and in the promotion of the use of GNSS and their integration into national infrastructure, particularly in developing countries.

The ICG also noted the significant results from its four working groups that specifically focus on the following issues: compatibility and interoperability, enhancement of the performance of GNSS services, information dissemination and capacity building, and reference frames, timing, and applications. These results, as noted below, included the following:

The Working Group on Compatibility and Interoperability decided to continue addressing the need for worldwide GNSS spectrum protection through a recommendation to providers and user community member states to promote the implementation of protection measures for GNSS operations in their nations and/or regions as well as other parts of the world. The efforts of the working group led to a recommendation to the United Nations Committee on the Peaceful Uses of Outer Space (UNCOPUOS) to establish a multiyear agenda item focused on national efforts to protect the Radio Navigation Satellite Services (RNSS) spectrum and pursue GNSS Interference Detection and Mitigation (IDM) in member states. The International GNSS Monitoring and Assessment (IGMA) Task Force intends to initiate a joint trial project with IGS that will demonstrate a global GNSS Monitoring and Assessment capability in its ongoing efforts. Finally, the interoperability task force will also engage in ongoing work on open service performance standards. The ICG noted that the Compatibility and Performance Standards subgroup has been renamed the Compatibility and Spectrum Subgroup, which will also have responsibility for the IDM Task Force. The approved new work plan for the Working Group on Compatibility and Interoperability now includes work focused on system-of-systems operations, pending tasking from the Providers Forum that also met in Boulder in conjunction with the ICG.

This will lead to new architecture that will be developed within the newly named Systems, Signals, and Services Working Group of the ICG.

Working Group on the Enhancement of GNSS Service Performance made important progress in establishing an interoperable GNSS Space Service Volume (SSV). Characteristics key to establishing an interoperable GNSS SSV were given by all six providers. Members of the Working Group will now focus on developing an official booklet on interoperable GNSS Space Service Volume for ultimate ICG approval. Work continued on assessing the benefits.

The group also reviewed the progress in analyzing the benefits of the NeQuick Galileo ionospheric model for single-frequency users in Low Earth Orbit (LEO). The Working Group members acknowledged the benefits of ranging signals broadcast from Galileo satellites in eccentric, non-nominal Medium Earth Orbit (MEO) for position, velocity, and time (PVT) applications and scientific demonstrations. Progress was also reported on the use of GLONASS for geodetic applications showing similar performance to other GNSS systems. Finally agreement was reached on the effectiveness of the use of wide band signals to minimize multipath error and significantly improve the accuracy for users.

The Working Group on Information Dissemination and Capacity Building proposed to expand knowledge sharing, by engaging in faculty/student exchange programs and also providing textbooks/teaching materials. There are also new efforts underway to increase cooperation and support among Providers' Service Centers and the United Nations-affiliated Regional Centres for Space Science and Technology Education. The concept of National and Regional Positioning, Navigation, and Timing (PNT) Advisory Committees was also considered.

The Working Group on Reference Frames, Timing, and Applications called to the attention of the ICG the United Nations General Assembly resolution on the Global Geodetic Reference Frame (GGRF) passed in February 2015. The Committee of Experts for the United Nations Global Geospatial Information Management (UN-GGIM) endorsed the establishment of a working group on the GGRF, whose task is to develop a "roadmap" for the realization of the GGRF. Specific areas of progress were noted with regard to global geodetic and timing references for GNSS services. This particularly focuses on the computations of the new International Terrestrial Reference Frame (ITRF2014). ITRF2014 is foreseen as a significant improvement over the current ITRF2008. Meetings of the ICG are scheduled for Russia (2016), Japan (2017), China (2018), and India (2019) (Tenth Meeting of the ICG 2015).

Conclusion

With new systems and increasing of new frequency bands and applications of new services and more and more space systems, there are increasing demands on the use of the frequency spectrum around the bands utilized by the GNSS. One should expect this trend to continue. An uncoordinated approach in which the use of the frequency bands are awarded to industry or government institutions based only on considerations of the use of one or two space systems or on the benefits to a specific application could result in an adverse impact to the signal reception of all GNSS. Since GNSS in its civil application is a service-oriented system, which is aimed at serving all humankind, this issue warrants significant attention across the world.

It is precisely these types of issues that are being dealt with by the International Committee on GNSS. There are many GNSS events around the world on a yearly basis, each contributing to one or more aspects that promote and strengthen the use of GNSS in a myriad of applications. However, the meetings of the ICG and its Provider Forum and the intersessional work carried out by their Working Groups are the unique mechanisms where the operators of GPS, GLONASS, Galileo, Compass/BeiDou, IRNSS, and QZSS can identify issues that could adversely affect the use of the GNSS.

The issues addressed at the ICG include whether frequency coordination related to satellite navigation satellites should be conducted only on a bilateral basis, as has been the case, or whether frequency coordination should also be a multilateral process. Another key area dealt with by the ICG are the practical and technical issues of how to separate civil- and defense-related services in the context of signaling, frequency assignments, and transmission requirements. With a membership that includes all GNSS operators and the large user associations, the ICG is the best and perhaps the only place to discuss and resolve complex situations that could prevent the achievement of compatibility and interoperability among the GNSS to the detriment of current and future users of GNSS.

Cross-References

- ► Current and Future GNSS and Their Augmentation Systems
- ► Global Navigation Satellite Systems: Orbital Parameters, Time and Space Reference Systems and Signal Structures
- ► Introduction to Satellite Navigation Systems

References

A GNSS web portal with information on ICG matters has been established by the United Nations Office for Outer Space Affairs at http://unoosa.org/oosa/en/SAP/gnss/icg.html. Last accessed 14 Jan 2016

International Committee on Global Navigation Satellite Systems (ICG): Providers' Forum, (2015), http://unoosa.org/oosa/en/SAP/gnss/icg/providersforum.html. Last accessed 14 Jan 2016

Report of the Third United Nations Conference on the Exploration and Peaceful Uses of Outer Space, Vienna, 19–30 July 1999 (United Nations publication, Sales No. E.00.I.3), chap. I, resolution 1

Report of the 9th meeting of Action Team 10 to COPUOS on the Establishment of the an International Committee on Global Navigation Satellite System (ICG): 27 February 2004, Vienna, Austria

Report of Working Group A, Recommendation 2.1 for Committee Action (2011), http://www.unoosa.org/pdf/icg/2011/icg-6/wgA/ICG_WGA_2011.pdf. Last accessed 14 Jan 2016

Report of Working Group B, Enhancement of Global Navigation Satellite Systems Services Performance (2011), http://www.unoosa.org/pdf/icg/2011/icg-6/wgB/ICG_WGB_2011.pdf. Last accessed 14 Jan 2016

Report of Working Group C, (2011), http://www.unoosa.org/pdf/icg/2011/icg-6/wgC/ICG_WGC_2011.pdf. Last accessed 14 Jan 2016

Report of Working Group D, (2011), http://www.unoosa.org/pdf/icg/2011/icg-6/WgD/ICG_WGD_2011.pdf. Last accessed 14 Jan 2016

Tenth Meeting of the International Committee on Global Navigation Satellite Systems (ICG), Boulder, 1–6 Nov 2015, http://www.unoosa.org/pdf/icg/2015/icg10/icg10joint-statement.pdf. Last accessed 2 Jan 2016

Current and Future GNSS and Their Augmentation Systems

Sergio Camacho-Lara

Contents

S. Camacho-Lara (✉)
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Santa María Tonantzintla, Puebla, Mexico
e-mail: sergio.camacho@inaoep.mx

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_25

Abstract

Global Navigation Satellite System (GNSSs) is the standard generic term for satellite navigation systems that provide autonomous geospatial positioning with global coverage. GNSS allows small electronic receivers to determine their location (longitude, latitude, and altitude) to within a few meters using time signals transmitted along a line-of-sight by radio from satellites. Receivers on the ground, air, or water calculate the precise time as well as position, which can be used as a reference for scientific experiments and numerous everyday applications.

Currently, the Navstar Global Positioning System (GPS) of the United States, the Global Navigation Satellite System (GLONASS) of the Russian Federation, and the People's Republic of China BeiDou/Compass navigation system are the three fully operational global GNSS. The European Union's Galileo positioning system is a GNSS in the initial deployment phase, scheduled to be operational in 2014. The global coverage for each system is generally achieved by a constellation of 24–30 Medium Earth Orbit (MEO) satellites distributed between several orbital planes. The actual systems vary but use orbit inclinations greater than 50°

and orbital periods of roughly 12 h (height 20,000 km/12,500 miles). These global systems are being joined by the regional Quasi-Zenith Satellite System (QZZS) of Japan and the Indian Regional Navigation Satellite System (IRNSS) of India. These regional systems utilize satellites at smaller inclinations in elliptical orbits with apogees around 24,000 and 39,000 km or in inclined geostationary orbits at around 36,000 km. As accuracy in position, time, or speed measurements increases with the number of satellites that can be observed by a receiver, the signals received from the global GNSS satellites are complemented by signals provided by satellite-based augmentations systems (SBAs). Such is the motivation for the Wide-Area Augmentation System of the United States, the System for Differential Correction and Monitoring (SDCM) of the Russian Federation, the European Geostationary Navigation Overlay Service (EGNOS), the GPS and Geo-Augmented Navigation system (GAGAN) of India, and the Multifunctional Transport Satellite (MTSAT) Satellite-based Augmentation System (MSAS) of Japan. Altogether, by 2020 there will be around 120 navigation and positioning satellites in orbit at any given moment. It is possible that a user could receive signals from as many as ten satellites, leading to accuracies only available at the research level today. This chapter presents the characteristics of all current and future generations of navigation and positioning satellites.

Keywords

GNSS • GPS • GLONASS • Galileo • Compass/BeiDou • IRNSS • QZSS • WAAS • SDCM • EGNOS • GAGAN • MSAS • Positioning • Navigation and Timing (PNT) system • Location-Based Service • Satellite-Based Augmentation System (SBA) • Compatibility and interoperability • GNSS user services and policies • Standard Positioning Service (SPS)

Introduction

The various Global Navigation Satellite Systems (GNSSs) around the world have grown into a global utility whose multiuse services are integral to national security of the providers of the systems as well to global security, economic growth, transportation safety, search and rescue activities, and scientific research. As such, the GNSS capabilities are an essential element of the worldwide economic and social infrastructure.

The existing and the soon to be built GNSS will transmit both open service signals and restricted signals (encrypted) for military or restricted commercial services. With four global position, navigation, and timing (PNT) systems that will be fully operational by 2020 and at least five operational satellite-based augmentation systems, the issues of compatibility and interoperability among the systems become of paramount importance both for the providers and the users of the systems. This means that the signals of the various systems should not interfere with each other and that, to the extent possible, a receiver could use the signal of more than one GNSS.

Compatibility refers to the ability of multiple satellite navigation systems to be used separately or together, without interfering with the navigation performance of any of the various systems. Interoperability refers to the ability of the open services of multiple satellite navigation systems to be used together to provide better capabilities at the user level than would be achieved by relying solely on one service, without significantly increasing the complexity and cost of the receivers.

This chapter presents descriptions of the satellite constellations of the currently operating GNSS and of those under construction as well as of their augmentation systems. The current and future signal frequency bands of the open and restricted service and other characteristics of the GNSS and their augmentation systems are provided. The services that each GNSS will provide and their policy for users are also indicated. The reader should note that through their participation in the "International Committee on GNSS" (ICG), all providers have agreed to the definition of compatibility and interoperability as stated in the previous paragraph and have stated their commitment to attempt to achieve compatibility and interoperability with other systems through bilateral and multilateral coordination. This is an excellent goal but as can be seen in the description of the current frequency bands of the BeiDou/Compass and Galileo systems, it is difficult to achieve.

The Global Positioning System (GPS) of the United States

The Navstar Global Positioning System, hereafter referred to as GPS, is a space-based radio navigation system owned by the United States Government (USG) and operated by the United States Air Force. GPS has provided positioning, navigation, and timing services to military and civilian users on a continuous worldwide basis since first launch in 1978. An unlimited number of users with a civil or military GPS receiver can determine accurate time and location, in any weather, day or night, anywhere in the world.

In an effort to ensure beneficial services are available to the greatest number of users without degrading security interests, two GPS services are provided. The precise positioning service (PPS) is available primarily to the military of the United States and its allies for users properly equipped with PPS receivers. The standard positioning service (SPS), as initially described in the SPS signal specification (see below), was originally designed to provide civil users with a less accurate positioning capability than PPS, through a feature known as selective availability (SA). In view of technological advancements, the USG has committed to maintain the discontinuance of the SA feature to degrade globally the SPS. The US President announced in 2007 that selective availability will not be built into modernized GPS III satellites. The United States National Space Policy released in June 2010 reaffirms the long-standing and stable US policy on space-based navigation.

Global Positioning System Satellite Constellation

The global positioning system (GPS) standard positioning service (SPS) consists of space-based positioning, navigation, and timing (PNT) signals delivered free of direct user fees for peaceful civil, commercial, and scientific uses worldwide.

The GPS baseline constellation consists of 24 slots in six orbital planes, with four slots per plane. Three of the slots are expandable and can hold no more than two satellites. Satellites that are not occupying a defined slot in the GPS constellation occupy other locations in the six orbital planes. This 24-slot arrangement ensures there are at least four satellites in view from virtually any point on the planet. The air force normally flies more than 24 GPS satellites to maintain coverage whenever the baseline satellites are serviced or decommissioned. The extra satellites may increase GPS performance but are not considered part of the core constellation. Figure 1 shows a schematic of the GPS constellation of satellites.

The constellation is being reconfigured to an expanded 24 baseline operational satellites and spare, also operational, to provide better coverage and availability around the world. Currently, GPS had 31 operational satellites in orbit to ensure a baseline constellation of 24 satellites plus three spares. Four additional satellites are in residual status and could be reactivated if one of the currently operational satellites experienced a sudden breakdown. Eight of the Block IIR-M satellites and one new

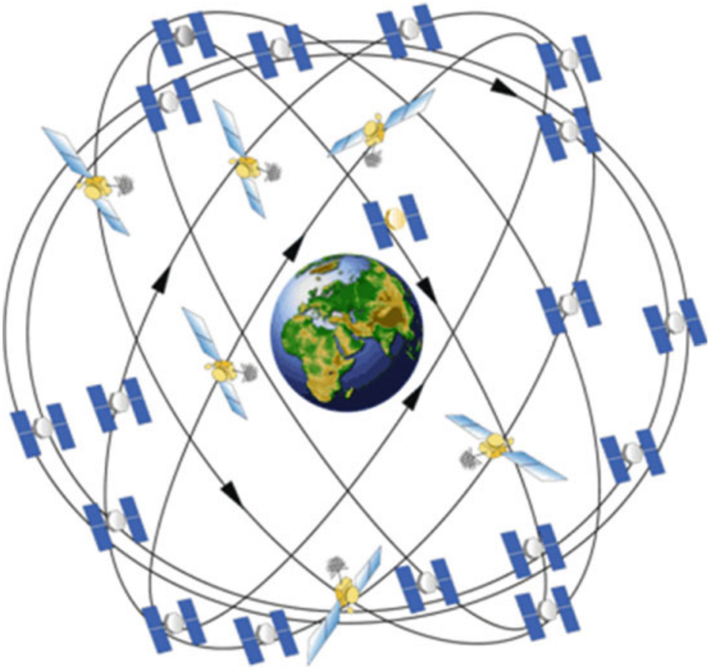


Fig. 1 Schematic of the GPS constellation (Courtesy National Coordination Office for Space-based PNT)

Block IIF satellite are broadcasting a second civil signal called L2C. Two IIF satellite are also broadcasting a new civil signal at L5, which is being used for safety-of-life applications (see below).

Current and Future Satellite Generations

The GPS constellation is a mix of new and legacy satellites. The following text describes the various generations, or blocks, of GPS satellites that are currently flying. It also describes the satellites under development as part of the GPS modernization program.

Block IIA is an upgraded version of the GPS Block II satellites launched in 1989–1990. The "II" refers to the second generation of GPS satellites, although Block II was actually the first series of operational GPS satellites. The "A" stands for advanced (Fig. 2).

The IIR series were produced to replace the II/IIA series as the II/IIA satellites gradually degraded or exceeded their intended design life. The "R" in Block IIR stands for replenishment.

The IIR(M) series of satellites are an upgraded version of the IIR series, completing the backbone of today's GPS constellation. The "M" in IIR(M) stands for modernized, referring to the new civil and military GPS signals added with this generation of spacecraft. There are 12 IIR satellites in the GPS constellation, forming the backbone of today's GPS along with seven healthy IIR(M) satellites in the constellation.

Fig. 2 A Block IIA satellite (Courtesy National Coordination Office for Space-based PNT)

Fig. 3 A Block IIF satellite (Courtesy National Coordination Office for Space-based PNT)

The Block IIF series expand on the capabilities of the IIR(M) series with the addition of a third civil signal in a frequency protected for safety-of-life transportation. The "F" in IIF stands for follow-on. Compared to previous generations, GPS IIF satellites have a longer life expectancy and a higher accuracy requirement. Each spacecraft uses a mix of rubidium and cesium atomic clocks to keep time within eight billionths of a second per day (Fig. 3).

The GPS III series is the newest block of GPS satellites. GPS III will provide more powerful signals in addition to enhanced signal reliability, accuracy, and integrity – all of which will support precision, navigation, and timing services. Based on the current contracts and funding, four GPS III satellites will be produced with options to purchase an additional eight satellites. Future versions will feature increased capabilities to meet demands of military and civilian users alike.

Current and Planned GPS Signals

The modernization program of the US global positioning system (GPS) continued with the first new Block IIF satellite launched in 2010 and now operating normally, broadcasting a new civil signal, L5, in addition to other civil signals: L2C and L1 C/A. These signals, which are transmitted at specific frequencies, are described below.

GPS L1

The L1 frequency, transmitted by all GPS satellites, contains a course/acquisition (C/A) code ranging signal with a navigation data message that is available for peaceful civilian, commercial, and scientific use, and a precision P(Y) code ranging signal with a navigation data message available to users with valid cryptographic keys. GPS satellites also transmit a second P(Y) code ranging signal with a navigation data message on the L2 frequency.

GPS L2C

The second civil signal, known as “L2C,” has been designed specifically to meet commercial needs. When combined with L1 C/A in a dual-frequency receiver, the L2C signal enables ionospheric correction, improving accuracy. For professional users with existing dual-frequency operations, L2C signals deliver faster signal acquisition, enhanced reliability, and greater operating range for differential applications. The L2C modulation also results in a signal that is easier to receive under trees and even indoors. This also supports the further miniaturization of low-power GPS chipsets for mobile applications.

The first GPS IIR-M satellite featuring L2C capabilities was launched in 2005. Every GPS satellite fielded since then has included an L2C transmitter. As of January 2010, there were seven GPS satellites broadcasting L2C signals. In June 2011, the Air Force successfully completed a GPS constellation expansion known as the “Expandable 24” configuration. Three of the 24 slots were expanded, and six satellites were repositioned, so that three of the extra satellites became part of the constellation baseline. As a result, GPS now effectively operates as a 27-slot constellation with improved coverage in most parts of the world. Interface specification information for the L2C signal can be found on the website of the Los Angeles Air Force Base (NAVSTAR Specification).

GPS L5

The third civil signal, known as “L5,” is broadcast in a radio band reserved exclusively for aviation safety services and radio navigation satellite services. With a protected spectrum, higher power, greater bandwidth, and other features, the L5 signal is designed to support safety-of-life transportation and other high-performance applications. Future aircraft will use L5 signals in combination with L1 C/A (also in a protected band) to improve accuracy via ionospheric correction and robustness via signal redundancy. The use of L5 signals will increase capacity, fuel efficiency, and safety in airspace, railroads, waterways, and highways. When used in combination with L1 C/A and L2C, L5 will provide a very robust service that may enable submeter accuracy without augmentations and very long-range operations with augmentations. The operational L5 signal is available with the follow-on series of GPS satellites, Block IIF, beginning in 2010. Interface specification information on the L5 signal can also be found on the website of the Los Angeles Air Force Base (Global Positioning).

GPS L1C

The fourth civil signal, known as “L1C,” has been designed to enhance interoperability between GPS and international satellite navigation systems. The United States and the European Union originally developed L1C as a common civil signal for GPS and the European Satellite Navigation System (Galileo). It features a multiplexed binary offset carrier (MBOC) waveform designed to improve mobile reception in cities and other challenging environments. Other satellite navigation systems, such as Japan’s Quasi-Zenith Satellite System (QZSS) and China’s Compass/BeiDou system, also plan to broadcast signals similar to L1C, to enhance interoperability with GPS. The United States launched its first L1C broadcasts at the same frequency as the original L1 C? A signal that is retained for backward compatibility. Interface specification information for the L1C signal can be found on the website of the Los Angeles Air Force Base (Performance Standards and Specifications, 3).

Although GPS will provide operationally three new modernized civil signals in the future: L2C, L5, and L1C, the performance specifications in the signal performance standards (SPS) apply only to users of the L1 (1,575.42 MHz) coarse/acquisition (C/A) signal, as this is the only civil GPS signal that has reached full operational capability at this time.

The modernized GPS constellation has been performing at very high accuracy levels – the 1 year performance as of July 2010 provided a user range error of one-half meter, the best ever. The reliability of the constellation has been enhanced by solid performance from the Block IIR and IIR-M satellites which have solar array and power capacity that far exceeds the specified mean mission duration, and there have been no clock failures in these satellites to date. The first Block IIF satellite was declared operational in August 2010 and the second one launched in January 2012. Ten more Block IIFs are in the pipeline with satellites three through five already in production.

Signal-in-Space Health

For accurate and trustworthy measurements of position, velocity, and timing, it is important for the user to know that the satellite sending the signal is functioning properly. The SPS signal-in-space (SPS SIS) health is the status given by the real-time health-related information broadcast by each satellite as an integral part of the SPS SIS. The SPS SIS health is also sometimes referred to as “satellite health” or “space vehicle health” or “SV health.” For this standard, there are three possible SPS SIS health conditions: “healthy,” “marginal,” and “unhealthy.” The mapping of the real-time health-related information broadcast by the satellite to these three conditions is given as follows.

The SPS SIS accuracy is described in two statistical ways: one way is as the 95th percentile (95 %) SPS SIS user range error (URE) at a specified age of data (AOD), the other is as the 95 % SPS SIS URE over all AODs. With either statistical expression, the SPS SIS accuracy is also known as the SPS SIS pseudo-range accuracy. Other accuracy-related SPS SIS performance parameters include the SPS SIS pseudo-range rate (velocity) accuracy defined as the 95 % SPS SIS pseudo-range rate error over all AODs and the SPS SIS pseudo-range acceleration (rate rate) accuracy defined as the 95 % SPS SIS pseudo-range acceleration error over all AODs.

The SPS SIS integrity is defined to be the trust which can be placed in the correctness of the information provided by the SPS SIS. SPS SIS integrity includes the ability of the SPS SIS to provide timely alerts to receivers when the SPS SIS should not be used for positioning or timing. The SPS SIS should not be used when it is providing misleading signal-in-space information (MSI), where the threshold for "misleading" is a not-to-exceed (NTE) tolerance on the SIS URE. For this SPS PS, the four components of integrity are the probability of a major service failure, the time to alert, the SIS URE NTE tolerance, and the alert (either one or the other of two types of alerts).

It is recognized that GPS receivers cannot always monitor the broadcast NAV message data since interruptions may be caused by temporary signal blockages, abnormal receiving antenna orientation, radio frequency interference (particularly jamming), and intermittent environmental effects. Although the GPS receiver is responsible for taking appropriate action when it cannot monitor, process, or apply the current real-time health-related information in the NAV message data, it is possible for the control segment to aid some GPS receivers by giving them some advance warning of impending SPS SIS health changes. This action will only be beneficial for SPS SIS integrity if the SPS SIS health changes from healthy to marginal or from healthy to unhealthy (Performance Standards and Specifications).

Services Provided and Provision Policies

GPS provides two levels of service: a standard positioning service, which uses the C/A code on the L1 frequency, and a precise positioning service, which uses the C/A code on the L1 frequency and the P(Y) code on both the L1 and L2 frequencies. Authorized access to the precise positioning service is restricted to the United States Armed Forces, federal agencies, and selected allied armed forces and governments. The standard positioning service is available to all users worldwide on a continuous basis and without any direct user charge. The specific capabilities provided by the GPS open service are published in the *GPS Standard Positioning Service Performance Standards*. The United States Department of Defense, as the operator of GPS, will continue enabling codeless/semi-codeless GPS access until 31 December 2020, by which time the L2C and L5 signals will be available on at least 24 modernized GPS satellites.

United States Space-Based Positioning, Navigation, and Timing Policy

The current United States space-based positioning, navigation, and timing policy, signed by the president in July of 2010, states that the United States must maintain its leadership in the service, provision, and use of global navigation satellite systems (GLONASS). To this end, the United States shall take the following steps:

- Provide continuous worldwide access, for peaceful civil uses, to the global positioning system (GPS) and its government-provided augmentations, free of direct user charges.
- Engage with foreign GNSS providers to encourage compatibility and interoperability, promote transparency in civil service provision, and enable market access for US industry.
- Operate and maintain the GPS constellation to satisfy civil and national security needs, consistent with published performance standards and interface specifications. Foreign positioning, navigation, and timing (PNT) services may be used to augment and strengthen the resiliency of GPS.
- Invest in domestic capabilities and support international activities to detect, mitigate, and increase resiliency to harmful interference to GPS, and identify and implement, as necessary and appropriate, redundant and back-up systems or approaches for critical infrastructure, key resources, and mission-essential functions.

The policy promotes the global use of GPS technology through the following key provisions:

- No direct user fees for civil GPS services
- Open and free access to the information necessary to develop and build equipment
- Performance improvements for United States space-based positioning, navigation, and timing services
- Promotion of GPS standards
- International compatibility and interoperability for the benefit of end users
- Protection of the radio navigation spectrum from disruption and interference

Wide-Area Augmentation System

A satellite-based augmentation system (SBAS) is a system that supports wide-area or regional augmentation through the use of additional satellite-broadcast messages. The Wide-Area Augmentation System (WAAS) of the United States is an SBAS developed by the Federal Aviation Administration to augment the global positioning system (GPS), with the goal of improving its accuracy, integrity, and availability. Essentially, WAAS is intended to enable aircraft to rely on GPS for all phases of flight, including precision approaches to any airport within its coverage area. Such systems include multiple ground stations, located at accurately surveyed points. The ground stations take position measurements of one or more of the GNSS satellites, the satellite signals, or other environmental factors which may affect the signal received by the users. Using these measurements, information messages are created and sent to one or more satellites for broadcast to the end users.

WAAS currently relies on the service of two leased geostationary satellites positioned at 107° W latitude and 133° W longitude. On 3 April 2010, the telemetry tracking and control system on the Intelsat satellite (positioned at 133° W longitude) failed, but service was restored after a short hiatus. Mitigation efforts have now been taken to ensure that dual coverage requirements are met over the long term. The objective of this system is to provide a user receiver with at least two geostationary satellites in view during localizer performance vertical operations.

In addition, to achieve increased accuracy for the arctic region, more satellite-based augmentation (SBA) reference systems have been installed in this part of the world that is not covered by GEO satellites. The Iridium low earth orbit satellite constellation with high inclination service to the polar region is now used to provide WAAS service in the polar regions by connecting with these augmented SBAs. This additional connectivity in the arctic region now increases accuracy. Precision is increased from 2.1 m down to 1.6 m (Breaking the Ice).

There are 38 wide-area reference stations throughout North America (in Canada, Mexico, and the United States, including Alaska and Hawaii) and Puerto Rico. The Federal Aviation Administration of the United States plans to upgrade the wide-area reference stations with receivers capable of processing the new GPS L5 signal.

International Cooperation to Ensure Compatibility and Pursue Interoperability

In addition to participating in the International Committee on GNSS (ICG), the United States pursues its international GNSS coordination objectives in many other ways. These include working through and with the Asia-Pacific Economic Cooperation forum, as well as standard-setting bodies such as relevant United Nations specialized agencies, the International Telecommunication Union (ITU), the International Civil Aviation Organization (ICAO), and the International Maritime Organization. Finally, the United States also pursues bilateral cooperation with other system providers.

The Global Navigation Satellite System (GLONASS) of the Russian Federation

The Global Navigation Satellite System, hereafter referred to as GLONASS, is a space-based radio navigation system owned by the government of the Russian Federation and operated for the Russian government by the Russian Aerospace Defense Forces. It both complements and provides an alternative to the United States' global positioning system (GPS) and is currently the only alternative navigational system in operation with global coverage and similar precision. It is possible to purchase GNSS receivers that can acquire both the GPS and the GLONASS frequencies.

The Soviet Union military identified, in the late 1960s, a need for a satellite radio navigation system (SRNS) for use in precision guidance of a new generation of ballistic missiles. The existing Tsiklon satellite navigation system required several minutes of observation by the receiving station to fix a position making them unusable for navigation positioning purposes. In 1968–1969, research institutes of the Ministry of Defense, Academy of Sciences, and Soviet Navy joined together to establish a single solution for air, land, sea, and space forces. This resulted in a 1970 requirements document that established the requirements for such a space-based system. After further basic research, in 1976 a decree was issued by the Soviet Union establishing the Global Navigation Satellite System (GLONASS).

The GLONASS: System Description

The nominal baseline constellation of GLONASS comprises 24 GLONASS-M satellites that are uniformly deployed in three roughly circular orbital planes at an inclination of 64.8° to the equator. The altitude of the orbit is 19,100 km. The orbit period of each satellite is 11 h, 15 min, 45 s. The orbital planes are separated by 120° right ascension of the ascending node. Eight satellites are equally spaced in each plane with a 45° argument of latitude. Moreover, the orbital planes have an argument of latitude displacement of 15° relative to each other. This constellation configuration provides for continuous, global coverage of the Earth's surface and near-Earth space and for minimizing the effect of disturbances on deformation of the orbital constellation.

Current and Future Satellite Generations

Development of GLONASS began in the Soviet Union in 1976. Beginning on 12 October 1982, numerous rocket launches added satellites to the system until the constellation was completed in 1995. However, due to the time that it took to orbit all the satellites of the system, the earlier satellites became nonfunctional and the system quickly became incomplete.

In the 2000s the restoration of the system was made a top government priority and funding was substantially increased. This resulted in GLONASS becoming the most expensive program of the Russian Federal Space Agency (ROSCOSMOS). By 2010, GLONASS had achieved 100 % coverage of Russia's territory, and in October 2011, the full orbital constellation of 24 satellites was restored, enabling full global coverage. Information on the official GLONASS website indicates that currently the Russian Federation has 31 GLONASS satellites in orbit, with 24 operating to provide global coverage and the rest are spares. The complete GLONASS constellation needs only 24 satellites to be fully functional (GLONASS).

The main contractor of the GLONASS program is Joint Stock Company Reshetnev Information Satellite Systems (formerly called NPO-PM). The company, located in Zheleznogorsk, is the designer of all GLONASS satellites, in cooperation with the Institute for Space Device Engineering and the Russian Institute of Radio Navigation and Time. Serial production of the satellites is accomplished by the company PC Polyot in Omsk.

Over the three decades of development, the satellite designs have gone through numerous improvements, and can be divided into three generations: the original GLONASS (since 1982), GLONASS-M (since 2003), and GLONASS-K (since 2011).

GLONASS First Generation Satellites

The true first generation of GLONASS (also called Uragan) satellites were all three-axis stabilized vehicles, generally weighing 1,250 kg and were equipped with a modest propulsion system to permit relocation within the constellation. Over time they were upgraded to Block IIa, IIb, and IIv spacecraft, with each block containing evolutionary improvements.

Six Block IIa satellites were launched in 1985–1986 with improved time and frequency standards over the prototypes and increased frequency stability. These spacecraft demonstrated a 16-month average operational lifetime. Block IIb spacecraft, with a 2-year design lifetimes, appeared in 1987, of which a total of 12 were launched, but half were lost in launch vehicle accidents. The six spacecraft that made it to orbit worked well, operating for an average of nearly 22 months.

Block IIv was the most prolific of the first generation. Used exclusively from 1988 to 2000, and continued to be included in launches through 2005, a total of 25 satellites were launched. The design life was 3 years; however, a number of these spacecraft exceeded this by a number of years.

Block II satellites were typically launched three at a time from the Baikonur Cosmodrome using Proton-K Blok-DM-2 or Proton-K Briz-M boosters. The only exception was when, on two launches, an Etalon geodetic reflector satellite was substituted for a GLONASS satellite.

GLONASS Second Generation Satellites

The second generation of satellites, known as GLONASS-M, was developed beginning in 1990 and first launched in 2003. These satellites possess a substantially increased lifetime of 7 years and weigh slightly more at 1,480 kg. They are approximately 2.4 m in diameter and 3.7 m high, with a solar array span of 7.2 m for an electrical power generation capability of 1,600 W at launch. The aft payload structure houses 12 primary antennas for L-band transmissions. Laser corner-cube reflectors are also carried on board to aid in precise orbit determination and geodetic research. Onboard cesium clocks provide the local clock source. Figure 4 below shows a GLONASS-M satellite.

A total of 14 second generation satellites were launched through the end of 2007. As with the previous generation, the second generation spacecraft were launched in triplets using Proton-K Blok-DM-2 or Proton-K Briz-M boosters.

Fig. 4 A GLONASS-M satellite (Courtesy of Roscosmos and Information Satellite Systems Reshetnev Company)

GLONASS Third Generation Satellites

GLONASS-K is a substantial improvement over the previous generation. It is the first unpressurized GLONASS satellite with a much reduced mass (750 kg versus 1,480 kg of GLONASS-M). It has an operational lifetime of 10 years, compared to the 7-year lifetime of the second generation GLONASS-M. It will transmit more navigation signals to improve the system's accuracy, including new CDMA signals in the L3 and L5 bands which will use modulation similar to modernized GPS, Galileo, and Compass/Beidou satellites. The new satellite's advanced equipment – made solely from Russian components – will allow the doubling of GLONASS' accuracy. As with the previous satellites, these are three-axis stabilized, nadir-pointing satellites with dual solar arrays for power sources. The first GLONASS-K satellite was successfully launched on 26 February 2011, and this series is now fully deployed. Figure 5 shows a GLONASS-K satellite.

Due to their weight reduction, GLONASS-K spacecraft can be launched in pairs from the Plesetsk Cosmodrome launch site using the substantially lower cost Soyuz-21b boosters or in six-at-once from the Baikonur Cosmodrome using Proton-K Briz-M launch vehicles.

The present constellation officially includes two reserve satellites, GLONASS 714 and 726, but neither of these satellites will ever be brought back to active service. Rather than being possible replacement satellites, these vehicles are being used to train the ground team to operate spare satellites in a full or nearly full constellation. GLONASS 727, in orbital slot 3, which was taken out of service on

Fig. 5 A GLONASS-K satellite (Courtesy of Roscosmos and Information Satellite Systems Reshetnev Company)

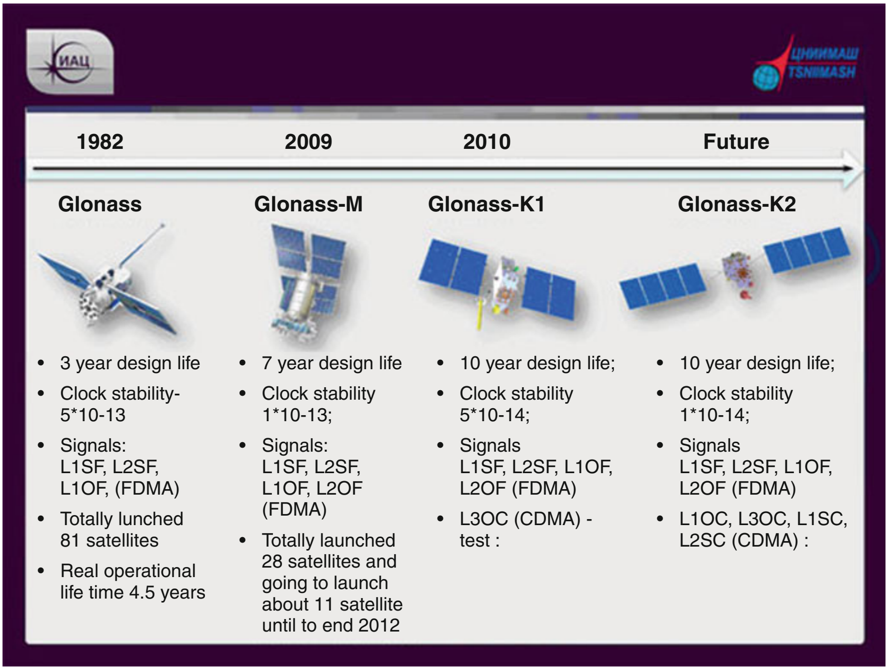
8 September 2010. GLONASS-M satellites now compose the major components of the GLONASS system (GLONASS).

Current and Planned GLONASS Signals

Each GLONASS satellite transmits two types of navigation signals in two sub-bands of L-band: a standard accuracy signal and a high accuracy signal. L1 carrier frequencies are in the band 1,598.06–1,604.40 MHz, and L2 carrier frequencies are in the band 1,242.94–1,248.63 MHz with right-hand circular polarization. GLONASS uses the frequency division multiple access (FDMA) in both L1 and L2 sub-bands.

ROSCOSMOS has deployed the GLONASS K1 spacecraft and is now working on the deployment of 9 K-2 satellites starting in 2017. The K2 satellites represent a new generation with expanded capabilities that will have a 10 year lifetime with spacecraft optimized to operate in a vacuum. Besides transmitting CDMA signals on L3, GLONASS-K2 will also transmit CDMA signals on L1. Alll the GLONASS-K satellites will transmit the legacy FDMA satellites in addition to the CDMA signals.

A modernized GLONASS-K satellite, GLONASS-KM is now under study. In addition to transmitting legacy FDMA signals on L1 and L2 and CDMA signals on L1, L2, and L3, CDMA signals may also be transmitted on the GPS L5 frequency at 1,176.45 MHz. This transmission coding will facilitate interoperability with the other GNSS. Also being studied is an alternative to the present three-plane, equally spaced satellite constellation. A different constellation design would be possible using CDMA signals. Such a move would require that the legacy FDMA signals be switched off. However, any such move would require at least 10-years' advance



1982	2009	2010	Future
Glonass	**Glonass-M**	**Glonass-K1**	**Glonass-K2**
• 3 year design life • Clock stability-5*10-13 • Signals: L1SF, L2SF, L1OF, (FDMA) • Totally lunched 81 satellites • Real operational life time 4.5 years	• 7 year design life • Clock stability 1*10-13; • Signals: L1SF, L2SF, L1OF, L2OF (FDMA) • Totally launched 28 satellites and going to launch about 11 satellite until to end 2012	• 10 year design life; • Clock stability 5*10-14; • Signals L1SF, L2SF, L1OF, L2OF (FDMA) • L3OC (CDMA) - test :	• 10 year design life; • Clock stability 1*10-14; • Signals L1SF, L2SF, L1OF, L2OF (FDMA) • L1OC, L3OC, L1SC, L2SC (CDMA) :

Fig. 6 Design characteristics of the GLONASS generations of satellites (Courtesy ROSCOSMOS)

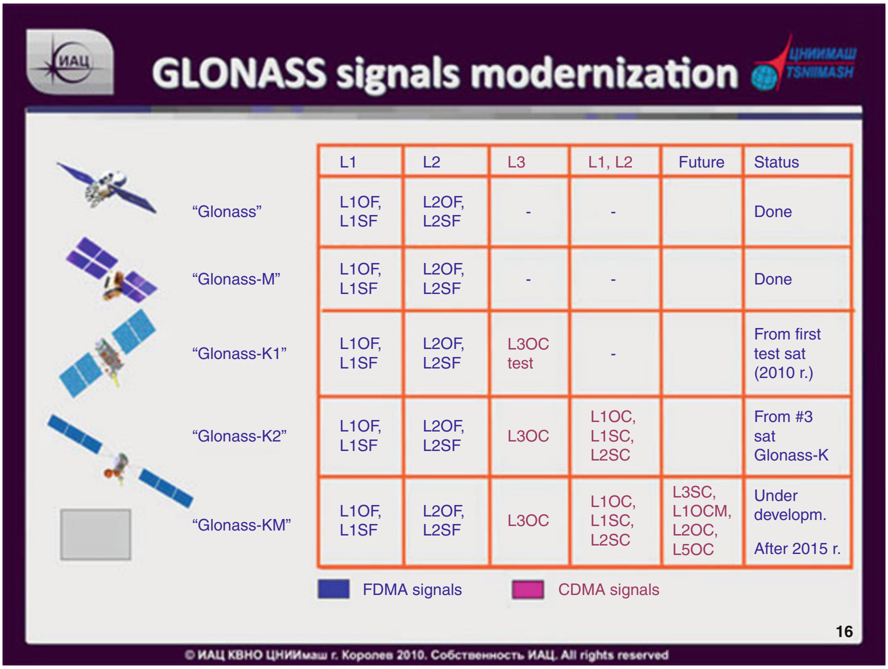
notice (R.B. Langley, GLONASS Update Delves into Constellation Details, http://www.gpsworld.com/gnss-system/glonass/news/glonass-update-delves-constellation-details-10499). In 2020, Russia plans to have 30 satellites in orbit, including six in reserve. To support the orbital grouping, Russia plans to launch 22 new-generation GLONASS-KM satellites in the period 2012–2020 to replace the outdated ones.

The design characteristics of the various generations of GLONASS satellites are shown in Fig. 6, and the modernization of the GLONASS signals are shown in Fig. 7 below.

In Figs. 6 and 7, the types of signals transmitted are identified with respect to access techniques as: OF = open-access FDMA, SF = special (military) FDMA, OC = open-access CDMA, OCM = open-access CDMA modernized.

Performance Standards Versus Actual Performance

The document that defines requirements related to the interface between the space segment and the navigation user segment is the interface control document (version 5.1, 2008). An update to this document is being prepared and should be available in the 2012–2013 time frame. At present, the main performance characteristics for GLONASS civil service are defined by the GLONASS Standard Positioning Service



GLONASS signals modernization

	L1	L2	L3	L1, L2	Future	Status
"Glonass"	L1OF, L1SF	L2OF, L2SF	-	-		Done
"Glonass-M"	L1OF, L1SF	L2OF, L2SF	-	-		Done
"Glonass-K1"	L1OF, L1SF	L2OF, L2SF	L3OC test	-		From first test sat (2010 r.)
"Glonass-K2"	L1OF, L1SF	L2OF, L2SF	L3OC	L1OC, L1SC, L2SC		From #3 sat Glonass-K
"Glonass-KM"	L1OF, L1SF	L2OF, L2SF	L3OC	L1OC, L1SC, L2SC	L3SC, L1OCM, L2OC, L5OC	Under developm. After 2015 r.

🟦 FDMA signals 🟪 CDMA signals

16

© ИАЦ КВНО ЦНИИмаш г. Королев 2010. Собственность ИАЦ. All rights reserved

Fig. 7 Modernization of the GLONASS (Courtesy ROSCOSMOS)

Performance Requirements (U.N. Office for Outer Space Affairs). According to this document, for "GLONASS-M" satellite constellation:

- The signal-in-space user range error value over any 24-h interval for all healthy satellites should be less than or equal to 6.2 m, with a 0.95 probability when using open service signals containing ephemeris and clock data transmitted by the operational constellation.
- The position dilution of precision availability (the percentage of time over any 24-h interval that position dilution of precision availability is less than or equal to 6 for the constellation of operational satellites) should be equal to or better than 98 % for the full 24-satellite constellation.
- The corresponding real-time and absolute mode positioning accuracy in the state reference frame using signal-in-space only (neglecting user clock bias and errors due to propagation environment and receiver) and assuming that position dilution of precision availability is equal to 2, should be 12.4 m over any 24-h interval for any point within the service volume with 0.95 probability.

Additional information on GLONASS performance monitoring is contained in the website of the Information and Analysis Center (IAC) of ROSCOSMOS (GLONASS).

Services Provided and Provision Policies

GLONASS is a part of the critical state space-based position, navigation, and timing (PNT) infrastructure providing for national security and economic development. Thus, creating, developing, and sustaining the PNT infrastructure is a state responsibility.

GLONASS satellites broadcast two types of navigation signals in L1 and L2 frequency bands: the standard positioning signal and the high accuracy positioning signal. The Russian Federation maintains a policy of no direct user fees for the standard positioning GLONASS signal. The high accuracy positioning signal is modulated by a special code and is reserved for military and selected users and applications. The operator also provides open, free access to GLONASS information necessary to develop and build user equipment. The Russian Federation participates in international cooperation with the providers of other GNSS to ensure compatibility and pursue interoperability.

System for Differential Correction and Monitoring

On 11 December 2011, Roscosmos launched the Luch-5A geostationary relay satellite. Roscosmos reported that the satellite's antennas and solar panels deployed successfully and that the satellite would be positioned in a geostationary orbit at 16° west longitude.

Luch-5A carries a transponder to enable the System for Differential Correction and Monitoring (SDCM). SDCM is a satellite-based augmentation system (SBAS) for GLONASS and GPS, compatible with the US Wide-Area Augmentation System and other SBASs. Luch-5A is the first in a series of new data relay satellites designed to rebuild the Luch multifunctional space relay system that was used to transmit to Earth live TV images, communications, and telemetry from Mir, the Soviet/Russian space station; the Russian segment of the International Space Station; and other low-Earth-orbiting (LEO) spacecraft.

The SBAS transponder will transmit correction and integrity data for GLONASS and GPS on the GPS L1 frequency with a C/A pseudorandom noise code to be assigned by the GPS Directorate. The data will be provided by the SDCM, which uses a ground network of monitoring stations on Russian territory as well as some overseas stations.

As the SDCM primary service area is Russian territory, the main lobe of the SBAS antenna beam will be directed to the north with an angle of 7° relative to the equator. The transmitted power will be 60 W and will give a signal power level at the Earth's surface roughly equal to that of GLONASS and GPS signals, about -158 dBW.

The current international SBAS data format has a limited capability for broadcasting corrections for both GLONASS and GPS satellites combined. There is space for only 51 satellites, insufficient for the current number of satellites on orbit. Studies are being carried out in an attempt to resolve this problem. One option is to use a dynamic satellite mask, where an SDCM satellite would only broadcast corrections and integrity data for those GLONASS and GPS satellites in view of users in the territory of the Russian Federation.

International Cooperation to Ensure Compatibility and Pursue Interoperability

The Russian Federation recognizes that spectral separation of authorized service signals and other systems' signals is not, in practice, always feasible and that such overlap exists now and might continue to do so in the future and believes that providers of the GNSS should try to resolve those issues through consultations and negotiations. In that spirit, the Russian Federation has coordinated with the United States, on GLONASS-GPS compatibility and with the European Union/European Space Agency (ESA) on GLONASS-Galileo compatibility and interoperability and participates in the ICG and its Providers Forum.

The Compass/Beidou Global Navigation Satellite System of China

The BeiDou Navigation System or BeiDou (Compass) Navigation Satellite System is a project by the People's Republic of China to develop an independent satellite navigation system. The name may refer to either one or both generations of the Chinese satellite navigation system.

Space Segment System Description

The first BeiDou system, officially called BeiDou Satellite Navigation Experimental System, and known as BeiDou-1, consists of three satellites and has limited coverage and applications. This system was a demonstration Phase 1 and has been offering navigation services mainly for customers in China and from neighboring regions since 2000.

Phase II is BeiDou Navigation Satellite (regional) System which is aimed at providing service for areas in China and its surrounding areas from 2012 but which in fact can provide services to many areas around the world. Phase III will see the BeiDou Navigation Satellite System established completely and will provide a fully global service by 2020.

The second generation of the system, known as Compass or BeiDou-2, approved by the government in 2004, will be a global satellite navigation system consisting of 35 satellites and is still under construction. The Compass or BeiDou-2 Navigation Satellite System will consist of five geostationary satellites and 30 non-geostationary satellites. BeiDou-2 became operational with coverage of China in December 2011 with ten satellites in use. BeiDou-2 will offer services to users in Asia-Pacific region from 2012 and to the global community by 2020.

On 2 November 2006, China announced that from 2008 BeiDou would offer an open service with an accuracy of 10 m, timing of 0.2 ns, and a speed of 0.2 m/s (Marks).

On 25 February 2012, China successfully launched a BeiDou-2 satellite from the Xichang Satellite Launch Center in the southwestern Sichuan province. The satellite was carried by a Long March 3C carrier rocket into a geosynchronous orbit and became the fifth geostationary and eleventh overall spacecraft in the current Beidou-2 constellation. The geostationary (GEO) satellites are located at 58.75° E, 80° E, 110.5° E, 140° E, and 160° E in orbits inclined at 55°.

The system is initially capable of providing high accuracy continuous, real-time passive 3-D geospatial positioning and speed measurement services for users in China and its neighboring regions, covering an area of about 120° longitude in the Northern Hemisphere.

On 30 April 2012, a Chinese Long March 3B rocket successfully lifted off to carry the Compass-M3 and Compass-M4 satellites that became China's 12th and 13th global navigation satellites in orbit. Today, China has fully assembled its own GNSS with global coverage. The nongeostationary satellites are placed in three 55° inclined orbits at an altitude of 21,500 km, in medium earth orbit (MEO), and in inclined geosynchronous orbit.

Compass Second Generation Satellites

The Compass-M satellites were developed from the DFH-3B satellite platform and are deployed in near circular orbits with a 55° inclination, at altitudes ranging between 21,500 and 24,100 km. The first Compass-M satellite was launched in April 2007. China has developed two models for Compass-M satellites. The two satellites that have been launched into MEO are based on the DFH-3 bus and are equipped with an apogee propulsion system for final orbit insertion. The second model is not equipped with an apogee propulsion system, and its platform is completely different from the DFH-3B bus. Still under development, the latter model will be flown in the final construction phase of the BeiDou-2 (Compass) constellation. Figure 8 shows a satellite of the Compass-M series.

Current and Planned Beidou-2 Signals

The frequency bands of the Compass/BeiDou Navigation Satellite System, as indicated in 2010 by the China National Administration of GNSS and Applications, included:

B1: 1,559.052–1,591.788 MHz
B2: 1,166.22–1,217.37 MHz
B3: 1,250.618–1,286.423 MHz

Fig. 8 A Compass-M series satellite (Courtesy NASA Spaceflight.com)

The frequencies bands for Compass, identified as B1, B2, and B3 above, are in fact allocated in four bands: E1, E2, E5B, and E6 and overlap with Galileo signals (see figure below). The fact of overlapping could be convenient from the point of view of the receiver design, but on the other hand raises the issues of intersystem interference, especially within E1 and E2 bands, which are allocated for Galileo's publicly regulated service. Compass proposes to use frequencies planned for Galileo's Public Regulated Service (PRS) – and for the GPS military code – meaning that in an emergency, Europe could not jam the Chinese signal without also jamming its own encrypted, security-related signals as well. The same problem holds for the US military (C. Alan).

However, under the International Telecommunication Union (ITU) policies, the first nation to start broadcasting in a specific frequency will have priority to that frequency, and any subsequent users will be required to obtain permission prior to using that frequency, and otherwise ensure that their broadcasts do not interfere with the original nation's broadcasts. It now appears that Chinese Compass satellites will start transmitting in the E1, E2, E5B, and E6 bands before Europe's Galileo satellites and thus have primary rights to these frequency ranges (Levin 2009). This is an issue that is being discussed bilaterally and in multilateral fora such as the ITU and the ICG (Fig. 9).

In July 2009, China presented one possible solution to the impasse on frequency overlay between that country's Compass system and the European Galileo program. At a meeting of Working Group A, on issues related to compatibility and interoperability, of the International Committee on GNSS held in Vienna, China showed plans to move the Compass signal modulation to binary offset carrier (BOC), with an alt BOC (15,10) open service (OS) signal in the

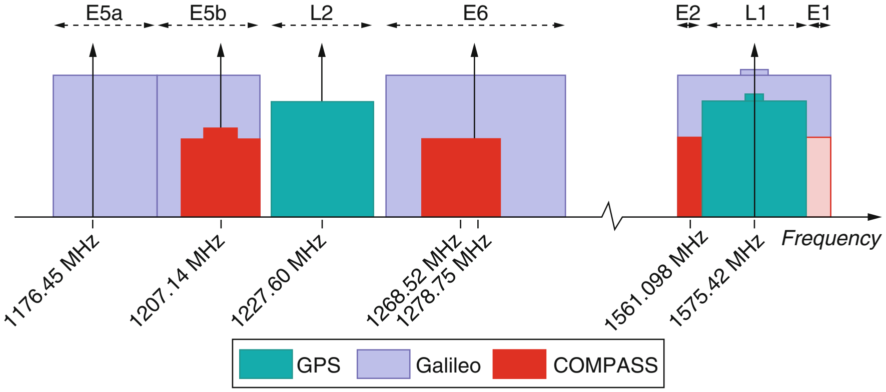
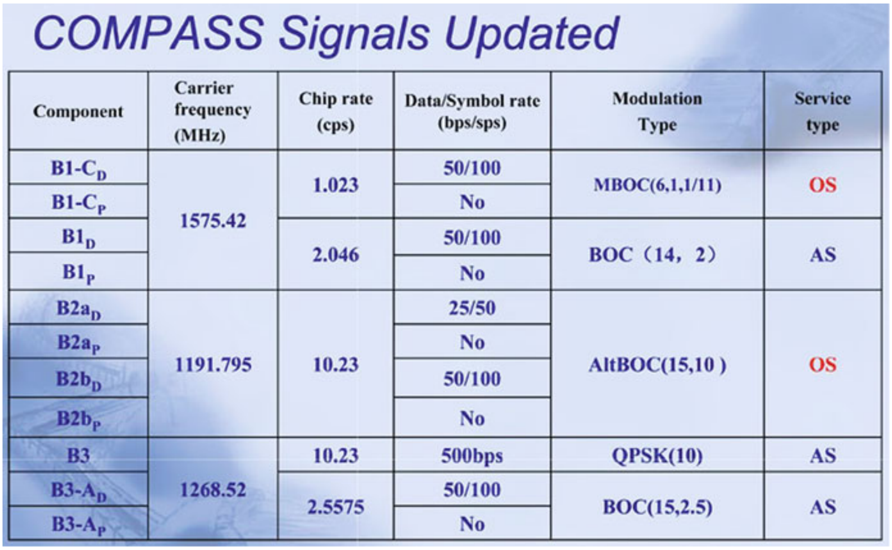


Fig. 9 Proposed frequency transmission bands for Compass, Galileo, and GPS (Credit: Wikipedia)



COMPASS Signals Updated

Component	Carrier frequency (MHz)	Chip rate (cps)	Data/Symbol rate (bps/sps)	Modulation Type	Service type
B1-C_D	1575.42	1.023	50/100	MBOC(6,1,1/11)	OS
B1-C_P			No		
B1$_\text{D}$		2.046	50/100	BOC（14，2）	AS
B1$_\text{P}$			No		
B2a$_\text{D}$	1191.795	10.23	25/50	AltBOC(15,10)	OS
B2a$_\text{P}$			No		
B2b$_\text{D}$			50/100		
B2b$_\text{P}$			No		
B3	1268.52	10.23	500bps	QPSK(10)	AS
B3-A$_\text{D}$		2.5575	50/100	BOC(15,2.5)	AS
B3-A$_\text{P}$			No		

Fig. 10 Technical information on the updated Compass signals (Courtesy: China National Administration of GNSS and Applications)

aeronautical radio navigation band at E5b (centered at 1,191.8 MHz) (see Fig. 10) (Working Group).

Performance Standards Versus Actual Performance

The performance standards indicated for Compass/Beidou-2 are: coverage area, global positioning accuracy 10 m (95 %), velocity accuracy 0.2 m/s, and timing accuracy 20 ns. With the 13 GNSS satellites that are in place in the constellation, Compass can already offer limited global coverage and has been offering the open service signal to users in China and neighboring areas.

Compass features a multiplexed binary offset carrier (MBOC) waveform designed to improve mobile reception in cities and other challenging environments and to increase interoperability with GPS, Galileo, and other satellite navigation systems, such as Japan's Quasi-Zenith Satellite System (QZSS). The technical characteristics of the signals for users and manufacturers of receiver equipment are available at the BeiDou website in the "BeiDou Navigation Satellite System Signal-In-Space Interface Control Document (Test Version)" (Beidou).

Services Provided and Provision Policies

The Compass/BeiDou Navigation Satellite System can provide two types of service at the global level: open service and authorized service. Through its open service, it provides cost-free positioning, velocity, and timing services. Through its authorized service, it provides higher accuracy positioning, velocity, and timing services, as well as system integrity information, for authorized users. The Compass/BeiDou Navigation Satellite System can provide two kinds of authorized services, including a wide-area differential service (with a positioning accuracy of 1 m) and a short-message communication service in China and nearby areas.

Perspective on Compatibility and Interoperability

The Compass/BeiDou Navigation Satellite System will achieve frequency compatibility with other satellite navigation systems under the ITU framework through bilateral coordination and taking advantage of multilateral discussions such as those carried out in Working Group A of the International Committee on GNSS. Presently, the COMPASS/BeiDou Navigation Satellite System has held coordination meetings with the operators of the GPS, Galileo, GLONASS, and QZSS systems.

The Compass/BeiDou Navigation Satellite System will achieve interoperability with other satellite navigation systems by coordinating through bilateral or multilateral platforms, including within the framework of the International Committee on GNSS (ICG). On the bilateral level, the system has held coordination meetings with the operators of the GPS and Galileo systems concerning compatibility and interoperability.

The European Satellite Navigation System (Galileo)

The European global navigation satellite system Galileo is a space-based radio navigation system owned by the European Union (EU) and, once functional, will be operated by the dedicated facilities of the European Space Agency (ESA) and several European national space agencies. Galileo is Europe's GNSS, providing a highly accurate, guaranteed global positioning service under civilian control. Galileo provides enhanced distress localization and call features for the provision of a search and rescue service interoperable with the COSPAS-SARSAT system.

As far back as the 1990s, the European Union saw the need for Europe to have its own global satellite navigation system. The conclusion to build one was taken in similar spirit to decisions in the 1970s to embark on other well-known European endeavors, such as the Ariane launcher and the Airbus. The European Commission and European Space Agency joined forces to build Galileo as an independent European system under civilian control.

European independence is the chief reason for taking this major step. However, other subsidiary reasons included that by being interoperable with GPS and GLONASS, Galileo will allow positions to be determined accurately for most places on Earth, even in high-rise cities where buildings obscure signals from satellites low on the horizon. This is because the overall number of satellites available from which to take a position is more than doubled. Another important reason building Galileo was that by placing satellites in orbits at a greater inclination to the equatorial plane than GPS, Galileo will achieve better coverage at the high European latitudes.

Space Segment

The fully deployed Galileo system will consist of 30 satellites. Initially this was to be 27 operational and 3 spares but was later changed to 24 operational and 6 spares. The satellites will have an approximate revolution period of 14 h. There will be eight operational satellites per orbital plane, occupying evenly distributed orbital slots. The six additional spare satellites (two per orbital plane) complement the nominal constellation configuration. Once this is achieved, the Galileo navigation signals will provide a good coverage even at latitudes up to 75° north and 75° south (Galileo Navigation). Figure 11 shows a schematic of the Galileo constellation.

By offering dual frequencies as standard, Galileo will deliver real-time positioning accuracy down to the meter range. It will guarantee availability of the service under all but the most extreme circumstances and will inform users within seconds of any satellite failure, making it suitable for safety-critical applications such as guiding cars, running trains, and landing aircraft.

Current and Future Satellite Generations

As was the case with other GNSS, the Galileo satellite constellation will be built through several generations of satellites that will have evolving capabilities and carry out needed functions in a step-by-step manner. The two main phases of the Galileo program are described below.

Fig. 11 The Galileo satellite constellation showing its high latitude coverage (Courtesy European Space Agency)

In-Orbit Validation (IOV) Phase and Full Operational Capacity (FOC) Phase

During this phase, the system is assessed through tests, the operation of two experimental satellites and a reduced constellation of four operational satellites and their ground infrastructure. The two experimental satellites were launched in, respectively, December 2005 and April 2008. Their purpose was to characterize the medium earth orbit (MEO) environment (radiations, magnetic field, etc.) and to test in such environment the performance of critical payload technology (atomic clocks and radiation hardened digital technology). They also provide an early experimental signal-in-space making it possible to secure the frequency spectrum required for Galileo in accordance with World Radio Conference RNSS allocations.

The first two Galileo satellites were launched into the first orbital plane on 21 October 2011. To date four pairs of FOC satellites have so far been launched by Soyuz from French Guiana. These launches were on 22 August 2014, 27 March 2015, 11 September 2015, and 17 December 2015. These satellites each weigh approximately 700 kg and have a design lifetime of 12 years.

As with any satellite, separation marks the start of the critical Launch and Early Orbit Phase (LEOP), when the spacecraft must carry out a series of automated and, later, commanded actions, including deployment of the solar arrays to obtain power, the switching on of the satellite's systems and the setting of initial configurations. This short but crucial LEOP period was overseen by a tightly integrated team of mission operations specialists from ESA's European Space Operations Center (ESOC) and the Center National d'Etudes Spatiales (CNES). The LEOP phase was formally completed on 3 November 2011 and control of the satellites was passed to the Galileo Control Center (GCC) at the Forschungszentrum der Bundesrepublik Deutschland für Luft-und Raumfahrt, the German Aerospace Center (DLR) in Oberpfaffenhofen, Germany.

When the navigation payload is switched on, it will mark the start of Galileo's in-orbit test campaign. This rigorous check of the navigation signals is being conducted from ESA's ground station in Redu, Belgium.

The fully deployed Galileo system will consist of 24 operational satellites plus six in-orbit spares, positioned in three circular medium earth orbit (MEO) planes at 23,222 km altitude above the Earth and at an inclination of the orbital planes of 56° to the equator. The full system will consist of 30 satellites, control centers in Europe, and a network of sensor stations and uplink stations installed around the globe. The FOC satellite launches will continue through 2020. Initial preliminary service began as of the end of 2016.

Under an agreement with the European Union, ESA acts as the technical lead for Europe's GNSS program and is responsible for acquiring satellites for the program.

Figure 12 shows the first operational Galileo satellite, FOC-FM1 as it was tested by OHB System AG.

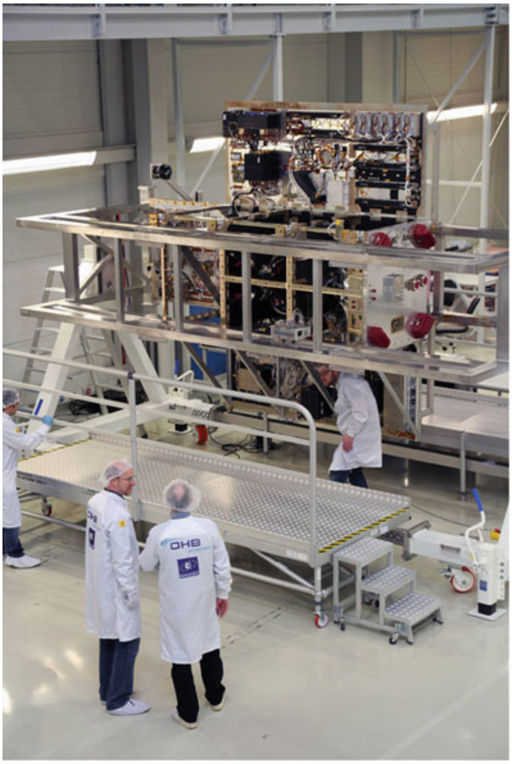


Fig. 12 The first operational Galileo satellite, FOC-FM1 being tested by OHB (Courtesy OHM System AG)

As prime contractor for the FOC satellites, OHB is responsible for developing the satellite platform and integrating the satellite with its payload – the part of the satellite that provides Galileo's precision positioning measurements and services to users worldwide. The SV1 payload was developed at Surrey Satellite Technology Ltd in Guildford, United Kingdom. In April 2012, SSTL delivered the first of 14 full operational capability (FOC) payloads for Galileo to prime contractor OHB System AG, and both companies have set up series production facilities to prepare the remaining satellites. In addition to these first 14, the OHB-SSTL consortium won a contract from the European Space Agency (ESA) in February 2012 to build a further eight Galileo FOC satellites.

As payload prime contractor, SSTL is tasked with the development, assembly, integration, and test (AIT) of the navigation payloads at their technical facility in Guildford, Surrey. This first payload has been shipped to prime contractor OHB in Bremen, Germany, for mechanical integration of the payload with the satellite platform and the beginning of the overall space vehicle AIT. SSTL's payload design is based on European-sourced atomic clocks, navigation signal generators, high-power traveling wave tube amplifiers, and antennas.

The definition phase and the development of the in-orbit validation phase of the Galileo program are carried out by the European Space Agency and cofunded by ESA and the EU. The FOC phase is funded by the EU and managed by the European Commission. The Commission and ESA have signed a delegation agreement by which ESA acts as design and procurement agent on behalf of the Commission (Galileo Navigation).

Current and Planned Galileo Signals

Galileo will transmit radio navigation signals in four different operating frequency bands: E1 (1,559 ~ 1,594 MHz), E6 (1,260 ~ 1,300 MHz), E5a (1,164 ~ 1,188 MHz), and E5b (1,195 ~ 1,219 MHz). These signals, which are transmitted at specific frequencies, are described below.

Galileo E1

The Galileo E1 band is centered at 1,575.42 MHz. It comprises two signals that can be used alone or in combination with signals in other frequency bands, depending on the performance demanded by the application. The signals are provided for the open service and the public regulated service, both of which include a navigation message. Moreover, an integrity message for the safety-of-life service is included in the open service signal. The E1 carrier is modulated with a CBOC (6,1,1/11) (following the MBOC spectrum) code for the open source and a BOCcos (15,2,5) code for the public regulated service.

Galileo E6

The Galileo E6 signal is transmitted on a center frequency of 1,278.75 MHz and comprises commercial service and public regulated service signals, which are modulated with a binary phase shift keying (BPSK)(5) and BOCcos(10,5) code, respectively. Both signals include a navigation message and encrypted ranging codes.

Galileo E5

The wideband Galileo E5 signal is centered on a frequency of 1,191.795 MHz. This signal is generated with an AltBOC modulation of sideband subcarrier rate of 15.345 MHz. This scheme provides two side lobes. The lower side lobe of E5 is called the Galileo E5a signal, which is centered on a frequency of 1,176.45 MHz and provides a second signal (dual-frequency reception) for the open service and safety-of-life services, both of which include navigation data messages. The upper side lobe of E5 is called the Galileo E5b signal, which is centered on a frequency of 1,207.14 MHz and provides a safety-of-life service, including a navigation message with an integrity information message.

Search and Rescue Signal

The search-and-rescue downlink signal is transmitted by the Galileo satellites in the frequency range of between 1,544 and 1,545 MHz.

Performance Standards Versus Actual Performance

The Galileo open service aims at making positioning, navigation, and timing services widely available, free of charge. The target Galileo open service positioning, navigation, and timing accuracy performances are specified as the 95th percentile of the positioning, navigation, and timing error distribution for different user types and take into account any type of error, including those not under the responsibility of the Galileo system. Hence, the target positioning, navigation, and timing performance specifications are subject to several assumptions on the user terminal and local environment: clear sky visibility, absence of radio frequency interference, reduced multipath environment, mild local ionospheric conditions, absence of scintillations, and fault-free user receiver.

The nominal performance specification for single frequency open service user (E1) and dual-frequency open service user (E1–E5b) are, respectively, 15 m and 4 m in horizontal accuracy and 35 m and 8 m, with a timing accuracy of 30 ns with respect to UTC.

Detailed information on the interface between the Galileo space segment and the Galileo user segment and on performance standards is being added to the Galileo Open Service Signal-In-Space Interface Control Document (OS SIS ICD) as Galileo is being developed (Galileo, 13).

Services Provided and Provision Policies

The Galileo mission and services have been elaborated during the initial definition phase in consultation with user communities and the European Member States.

The services that are planned to be provided by Galileo are the following:

- Open service: Basic signal provided free of charge.
- Safety-of-life service: Enhanced signal including an integrity function that will warn the user within a few seconds in case of a malfunction. This service will be offered to the safety-critical transport community, e.g., aviation. It will be certified according to the applicable standards, e.g., those of the International Civil Aviation Organization (ICAO).
- Commercial service: Combination of two encrypted signals for higher data throughput rate and higher accuracy authenticated data.
- Public regulated service: Two encrypted signals with controlled access for specific users like governmental bodies.
- Search and rescue service: Galileo will contribute to the international COSPAS-SARSAT cooperative system for humanitarian search and rescue activities. Each satellite will be equipped with a transponder transferring the distress signal from the user to the Rescue Coordination Center and informing him/her that the situation has been detected for distress beacons fitted with Galileo open service receivers.

European Geostationary Navigation Overlay Service (EGNOS)

The European Geostationary Navigation Overlay Service (EGNOS) provides an augmentation signal to the GPS standard positioning service. The EGNOS signal is transmitted on the same signal frequency band and modulation as the GPS L1 (1,575.42 MHz) C/A civilian signal function. While the GPS consists of positioning and timing signals generated from spacecraft orbiting the Earth, thus providing a global service, EGNOS provides correction and integrity information intended to improve positioning navigation services over Europe.

The EGNOS space segment consists of three navigation transponders on board three geostationary satellites and broadcasting corrections and integrity information for GPS satellites in the L1 frequency band (1,575.42 MHz). EGNOS uses the following three geostationary satellites: INMARSAT AOR-E, positioned at 15.5 W; INMARSAT IOR-W (F5), positioned at 25.0 E; and ARTEMIS, positioned at 21.5 E.

The EGNOS ground segment is composed of a network of ranging integrity monitoring stations, four mission control centers, six navigation land Earth stations, and the EGNOS wide-area network, which provides the communication network for all the components of the ground segment. Two additional facilities, the performance assessment and system checkout facility and the application specific qualification facility, are also part of the ground segment to support system operations and service provision.

The main objective of the EGNOS open service is to improve the achievable positioning accuracy by correcting several sources of errors affecting GPS signals. The accuracy achievable with the EGNOS open service is specified as the 95th percentile of the error distribution and is 3 m in horizontal accuracy and 4 in vertical accuracy. However, the typical measured positioning accuracy in the middle of the EGNOS open service area is significantly better than that specification (around 1 m (95 %) in vertical accuracy).

Perspective on Compatibility and Interoperability

Galileo coordinates with other space-based positioning, navigation, and timing systems to ensure compatibility. Achieving compatibility is essential when coordinating, and it involves both radio frequency compatibility and national security compatibility.

Galileo also coordinates at the bilateral and multilateral levels, including the ICG, to achieve interoperability among the GNSS and believes that in order to achieve interoperability:

- Common center frequency, common modulation, and common maximum power levels, based on the same link budget assumptions, are necessary.
- Highest minimum power level is desirable.
- The availability of information on open signals characteristics (such as a public signal-in-space interface control document) is necessary.
- Geodetic reference frames and system time references steered to international standards are necessary.
- Performance standards and system architecture descriptions must be published.

As indicated previously, the issue of overlapping of the planned frequency bands for transmission of some of its signals with the planned Compass frequency bands is still to be resolved. The agreements reached, or lack thereof, will be notified publicly in due course and reflected in their respective “Signal-In-Space Interface Control Documents.”

The Multifunctional Transport Satellite (MTSAT) Satellite-Based Augmentation System (MSAS) and the Quasi-Zenith Satellite System (QZSS) of Japan

Description of MSAS

The Multifunctional Transport Satellite (MTSAT) Satellite-Based Augmentation System (MSAS) provides GPS augmentation information for civil aircraft onboard satellite navigation systems under the Fukuoka Flight Information Region; it is one of the satellite-based augmentation systems that complies with ICAO standards and recommended practices.

MSAS provides navigation services for all aircraft within Japanese airspace via two geostationary satellites: MTSAT-1R, located at 140.0E, and MTSAT-2, located at 145.0E.

The initial satellite in the series, MTSAT-1, was built by Space Systems/Loral (SS/L) as a Multifunctional Transport Satellite. This satellite was intended to be also an advanced geostationary satellite for air traffic control and for weather observation. MTSAT-1 was manufactured under a contract with the Japanese Ministry of Transport Civil Aviation Bureau & Meteorological Agency. This multifunctional satellite was to provide communications and navigational services for aircraft and provide weather data to users throughout the entire Asia-Pacific region. The integration and ground testing of the satellite was conducted at SS/L's facilities in Palo Alto, California. The spacecraft was delivered in March 1999 and launched on an H-2S rocket in 2000 but due to a rocket failure never reached orbit (MTSat 1, 14).

MTSAT-1R was built by SS/L as a replacement for MTSAT-1 and was successfully launched into geostationary orbit by the launcher H-2A-2022 on 26 February 2005. It was positioned at an orbital slot of approximately 140° East and provides high-quality digital voice and data communications in the L, Ku, and Ka bands. After launch, MTSAT-1R was renamed Himawari-6.

The MTSAT-2 is a multifunctional satellite with a dual purpose. It is an integral part of a next-generation global-scale air traffic safety system comprised of communications, navigation, tracking, and air traffic control for improving traffic congestion conditions and safety in the Asia-Pacific region. In addition, the MTSAT-2 is designed to obtain, collect, and deliver meteorological images and/or data (MTSat 2, 15).

Figure 13 shows a satellite of the MTSAT series.

MTSAT-2 was built by Mitsubishi Electric as prime contractor and by the Boeing Company as a subcontractor. MTSAT-2 was launched on 26 February 2006 by the Japanese H-2A-2024 launch vehicle and placed in geostationary orbit at 145.0 E. MTSAT-2 was renamed Himawari 7 after launch.

The master control stations generate augmentation information based on the GPS and MTSAT signals received at the ground-monitoring stations and at the monitoring and ranging stations. The ground-monitoring stations monitor GPS satellite signals and transfer the information to the monitoring and ranging stations which monitor the MTSAT orbits.

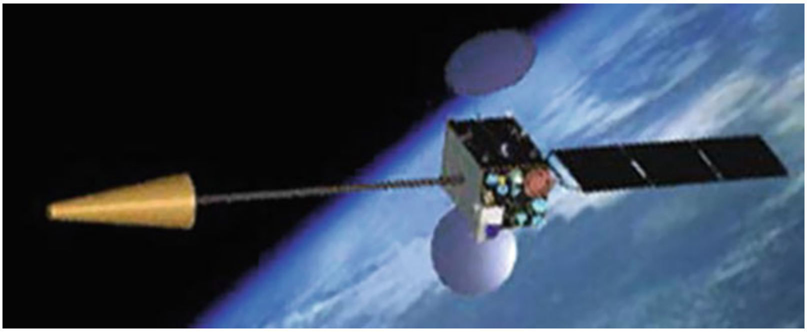


Fig. 13 An artist representation of an MTSAT satellite (Courtesy: JAXA)

Current and Planned Signals

MSAS navigation signals transmit from the L1 C/A GPS satellites at a center frequency of 1,575.42 MHz. The signal is modulated by a BPSK (Bi-Phase Shift Keying) technique with pseudorandom noise (PRN) spreading codes having a clock rate of 1.023 MHz, which is contained in the 250 bps/500 sps binary navigation data stream. MSAS is planning to expand bandwidths for L 1 and L 5. This implementation is under study, in accordance with the improvement schedule for the Wide-Area Augmentation System of the United States.

Performance Standards Versus Actual Performance

MSAS provides horizontal guidance for navigation, which is used in nonprecision approaches. According to ICAO standards and recommended practices, in order to satisfy these requirements: horizontal accuracy is less than 20 m, the observed value is less than 2.2 m (95 %), integrity (probability of hazardous, misleading information) is less than $1 \times 10 - 7$/h, fault tree analysis leads to $0.903 \times 10 - 7$/h, and availability is more than 99.9 %.

Services Provided and Provision Policies

MSAS is used for aircraft navigation and for that purpose offers three advanced functions. In the event of a GPS failure, the health status of GPS is transmitted via the integrity function of MSAS, while the differential correction function provides ranging error data. MSAS also employs a ranging function to generate GPS-like signals and enable aircraft to use MTSAT as an additional GPS satellite. In order to ensure the reliability of this function, MSAS monitors MTSAT/GPS signals, carries out ranging for the MTSAT satellite orbit, and estimates of ionospheric delay on a 24-h-a-day, 7-days-a-week basis.

Perspective on Compatibility and Interoperability

MSAS is compatible and can interoperate with other satellite-based augmentation systems.

Description of QZSS

QZSS is a regional space-based, all-weather, continuous positioning, navigation, and timing system that provides interoperable signals for GPS (L1, L2, and L5), a wide-area differential GPS augmentation signal called “L1-SAIF” and an experimental signal, “LEX,” having a message that contains more data, at a shorter time of transmission. QZSS provides navigation services for East Asia, including Japan and Oceania.

Fig. 14 An artist representation of a QZSS satellite (Courtesy: JAXA)

Space Segment

The space segment comprises the QZSS satellites, which function as celestial reference points, emitting precisely time-encoded navigation signals from space. The operational constellation of three satellites operates in 24-h orbits with an altitude of apogee of less than 39,581 km and a perigee of more than 31,911 km. Each of the three satellites is placed in its own separate orbital plane inclined 39–47° relative to the equator. The orbital planes are equally separated (i.e., phased 120° apart), and the satellites are phased so that there is always one satellite visible from Japan at a high elevation angle.

The satellite is a three-axis stabilized vehicle whose mass, without propellant, is approximately 1,800 kg, including a 320 kg-navigation payload. The major elements of its principal navigation payload are the atomic frequency standard for accurate timing; the onboard navigation computer to store navigation data, generate the ranging code, and stream navigation messages; and the 1.2/1.6 GHz band transmitting antenna whose shaped-beam gain pattern radiates near-uniform power of signals at the four 1.2/1.6 GHz band frequencies to users on or near the surface of the Earth. Figure 14 shows a QZSS satellite watching Japan from above.

Current and Planned Signals

The QZSS navigation signals transmitted from the satellites consist of five modulated carriers: two L1 carriers at center frequency 1,575.42 MHz (154f0), L2 at center frequency 1,227.6 MHz (120f0), L5 at center frequency 1,176.45 MHz (115f0), and LEX at center frequency 1,278.75 MHz (125f0) where f0 = 10.23 MHz. It should be noted that f0 is the output of the onboard frequency reference unit to which all signals generated are coherently related.

The L1 signal consists of four BPSK modulation signals. Two of them, the L1 C/A and the L1-SAIF, are modulated with two different pseudorandom noise spreading codes that are modulo-2 add sequences of the outputs of two 10-bit-linear-feedback-shift registers (10-bit-LFSRs) having a clock rate of 1.023 MHz and a period of 1 ms. Each of them is modulo-2 added to a 50 bps/50 sps or 250 bps/500 sps (i.e., standard positioning service) binary navigation data stream prior to BPSK.

The other two signals, L1Cp and L1Cd, are modulated with two different spreading codes having a clock rate of 1.023 MHz and with two same square waves having a clock rate of 0.5115 MHz. Data stream is modulo-2 added to L1Cd. Only the L1-SAIF signal is transmitted through a separate horn antenna using a different L1 carrier wave.

The L2 signal is BPSK with an L2C spreading code. The L2C code has a clock rate of 1.023 MHz with alternating spreading codes having a clock rate of 0.5115 MHz: L2CM with a period of 20 ms and L2CL with a period of 1.5 s. A 25 bps/50 sps data stream is modulo-2 added to the code prior to phase modulation.

The L5 signal consists of two BPSK signals (I and Q) multiplexed in quadrature. The signals in both I and Q channels are modulated with two different L5 spreading codes. Both of the L5 spreading codes have a clock rate of 10.23 MHz and a period of 1 ms. A 50 bps/100 sps binary navigation data stream is transmitted on the I channel and no data (i.e., a data-less "pilot" signal) on the Q channel.

The LEX signal is also BPSK modulated. A set of small Kasami code sequences is employed for the spreading code having a clock rate of 5.115 MHz.

Performance Standards Versus Actual Performance

The specification of signal-in-space user range error is less than 1.6 m (95 %), including time and coordination offset error to GPS. User positioning accuracy for QZSS is defined as positioning accuracy of the combined GPS L1 C/A and QZSS L1 C/A for a single frequency user and L1-L2 for a dual-frequency user. The figures of specification are 21.9 m (95 %) and 7.5 m (95 %), respectively. These specifications have already been verified by simulation using actual system design and parameters measured in an engineering model test. Simulation results for signal-in-space have shown user range error of 1.5 m (95 %), a positioning accuracy of 7.02 m (95 %) for a single frequency user and of 6.11 m (95 %) for a dual-frequency user.

The L1-SAIF signal provides wide-area differential GPS correction data, and its positioning accuracy is to be estimated 1 m (1 sigma rms) without large multipath error and ionospheric disturbance (Quasi-Zenith Satellites System).

Services Provided and Provision Policies

GPS interoperable signals like L1 C/A, L2C, L5, and L1C are to be provided free of charge to direct users. Regarding GPS performance enhancement signals, such as L1-SAIF and LEX, a charging policy is under examination. In order to support the design of the QZSS receiver by the receiver manufacturers and its application by a positioning, navigation, and timing service provider, interface specifications for QZSS users were released at an early stage of system development. The document describes not only radio frequency properties, message structure, and definition but also system characteristics, service performance properties, and the concept of operation. Both Japanese and English versions of the document can be downloaded from the JAXA website (Interface Specifications for QZSS, 16).

The Indian Regional Navigation Satellite System (IRNSS) and the Global Positioning System-Aided GEO-Augmented Navigation (GAGAN) System

The Indian Regional Navigation Satellite System, hereafter referred to as IRNSS, is an autonomous regional satellite navigation system being developed by the Indian Space Research Organization (ISRO). The IRNSS would be operated by ISRO under control of the Indian government. The requirement of such a navigation system is driven by the fact that access to global navigation satellite systems such as GPS is not guaranteed in hostile situations. The IRNSS would provide two services, the standard positioning service (SPS) open for civilian use and the restricted service (RS), which will be encrypted, available only for authorized users (military).

System Description

The proposed system would consist of a constellation of seven satellites and a support ground segment. Three of the satellites in the constellation will be placed in geostationary orbit. These geostationary orbit satellites (GSO) will be located 34° E, 83° E, and 131.5° E longitude. The GSO satellites will be in orbits with a 24,000 km apogee and 250 km perigee inclined at 29°. Two of the satellites will cross the equator at 55° E and 111.5° E. Such an arrangement would mean all seven satellites would have continuous radio visibility with Indian control stations. The satellite payloads would consist of atomic clocks and electronic equipment to generate the navigation signals.

IRNSS signals will consist of a special positioning service and a precision service. Both will be carried on L5 (1,176.45 MHz) and S band (2,492.08 MHz). The SPS signal will be modulated by a 1-MHz BPSK signal. The precision service will use binary offset coding, BOC (5,2).

The navigation signals themselves would be transmitted in the S-band frequency (2–4 GHz) and broadcast through a phased array antenna to maintain required coverage and signal strength. The satellites would weigh approximately 1,330 kg and their solar panels generate 1,400 W.

The system is intended to provide an absolute position accuracy of better than 20 m throughout India and within a region extending approximately 2,000 km around it.

Current and Planned Signals

The IRNSS constellation transmits navigation signals in L5 and S bands. Standard position services (SPS) and authorized/restricted services (RS) that use encryption technologies are the basic services offered by IRNSS. The IRNSS standard position and restricted services are transmitted on L5 (1,164–1,215 MHz) and S (2,483.5–2,500 MHz) bands. The carrier frequencies and bandwidths of these signals are the following:

- SPS – L5 has a carrier frequency of 1,176.45 MHz and a bandwidth of 24 MHz
- RS – L5 has a carrier frequency of 1,176.45 MHz and a bandwidth of 24 MHz
- SPS – S has a carrier frequency of S 2,492.028 MHz and a bandwidth of 16.5 MHz
- RS – S has a carrier frequency of 2,492.028 MHz and a bandwidth of 16.5 MHz

The standard position service (SPS) signal is BPSK(1) modulated on both the L5 and S bands. The navigation data is at a data rate of 25 bps and is modulo 2 added to a pseudorandom noise code chipped at 1.023 Mcps identified for the standard position service. The CDMA-modulated code modulates the L5 and S carriers at 1,176.45 and 2,492.028 MHz, respectively.

The restricted service is only available for authorized users. The restricted service signal is transmitted on L5 and S bands using binary offset coding (BOC). It has two channels: a “data” channel and a “pilot,” or “data-less,” channel. The navigation data at 25 bps is modulo 2 added with designated PRN code chipped at 2.046 Mcps in the data channel. The CDMA bit stream modulates the L5 and S carriers using BOC (5,2). The pilot channel is transmitted using primary and secondary codes without data modulation. The primary codes are chipped at 2.046 Mcps. The pilot carrier is in phase quadrature with the data channel.

GPS-Aided GEO-Augmented Navigation System (GAGAN)

The GPS-aided GEO-Augmented Navigation System (GAGAN) is a satellite-based augmentation system to enhance the use of the GPS signal. As an operational system, it is planned that the space segment will consist of two geostationary satellites, located at 82° E and 55° E, respectively. An additional on-orbit spare (located at 83° E) will also be added.

Services Provided and Provision Policies

Regarding the Indian Regional Navigation Satellite System, standard position services and authorized service/restricted service are the basic services offered by

IRNSS. The standard position service is free for all users. The restricted service is encrypted and, as such, available only to authorized users.

GAGAN will provide a safety-of-life service that meets all the requirements of accuracy, integrity, continuity, and availability required by ICAO for the utilization by civil aviation for en route, nonprecision, and precision approaches.

Conclusion

This chapter has sought to provide a detailed summary of the various satellite navigation and timing systems that are deployed or are being deployed around the world. Clearly, this a very dynamic period with the Russian GLONASS system now restored to a very high level of worldwide performance and a fully global system now deployed by the Chinese. In addition European, Japanese, and Indian initiatives are also moving rapidly ahead. Although the US Navstar/GPS system along with the Wide-Area Augmentation System (WAAS) is the most widely used system – partly because of the global availability of low cost GPS receivers – the other systems described in this chapter are going to be more and more widely used. The compatibility among the GNSS will be key to the interoperability. Since this is a rapidly changing area with the deployment of many new satellites and system capability, it is prudent to go to the various websites cited in this chapter to check on the very latest information that becomes available on a month to month basis.

Cross-References

- ▶ Global Navigation Satellite Systems: Orbital Parameters, Time and Space Reference Systems and Signal Structures
- ▶ International Committee on GNSS
- ▶ Introduction to Satellite Navigation Systems

References

C. Alan, The system: a healthy constellation. GPS World (Apr 2008), http://www.gpsworld.com/gnss-system/compass/the-system-a-healthy-constellation-4226

Beidou (11 Dec 2011), http://www.beidou.gov.cn/attach/2011/12/27/201112273f3be6124f7d4c7bac428a36cc1d1363.pdf. Last accessed 4 Apr 2011, 6

Breaking the Ice, Inside GNSS: Engineering Solutions from the Global Navigation Satellite System Community (Sept/Oct 2011), http://www.insidegnss.com/node/2748. Last accessed 4 Apr 2016

Galileo, European Commission (2016), http://ec.europa.eu/growth/sectors/space/galileo/. Last accessed 4 Apr 2016

Galileo Navigation, European Space Agency (2016), http://www.esa.int/Our_Activities/Navigation/The_future_-_Galileo/What_is_Galileo. Last accessed 4 Apr 2016

Global Positioning System Wing (GPSW) Systems Engineering & Integration IS-GPS-800A (8 June 2010), http://www.gps.gov/technical/icwg/IS-GPS-800A.pdf. Last accessed 4 Apr 2016

GLONASS Constellation Status (4 Apr 2016), http://www.glonass-ianc.rsa.ru/en/. Last accessed 4 Apr 2016

GLONASS Future and Evolution. Navipedia (12 Apr 2015), http://www.navipedia.net/index.php/GLONASS_Future_and_Evolutions. Last accessed 4 Apr 2016

Interface Specifications for QZSS Quasi Zenith Satellite System Navigation Service (Nov 2014), http://qz-vision.jaxa.jp/USE/is-qzss/index_e.html. Last accessed 4 Apr 2016

D. Levin, Chinese square off with Europe in space. The New York Times (23 Mar 2009), http://www.nytimes.com/2009/03/23/technology/23iht-galileo23.html?_r=2%26scp=1%26sq=chinese%20europe%20galileo%26st=cse

P. Marks, China's satellite navigation plans threaten Galileo. New Scientist (8 Nov 2006), http://www.newscientist.com/article/dn10472-chinas-satellite-navigation-plans-threaten-galileo.html. Last accessed 4 Apr 2016

MTSat 1, 1R (Himawari 6) – Gunter's Space Page (5 June 2016), http://space.skyrocket.de/doc_sdat/mtsat-1.htm. Last accessed 4 Apr 2016

MTSat 2 (Himawari 7) – Gunter's Space Page (21 Jan 2015), http://space.skyrocket.de/doc_sdat/mtsat-2.htm. Last accessed 4 Apr 2016

NAVSTAR Global Positioning System Interface Specification (Mar 2006), http://www.losangeles.af.mil/shared/media/document/AFD-081021-035.pdf. Last accessed 4 Apr 2016

Performance Standards and Specifications for Global Positioning Satellite System, http://www.gps.gov/technical/ps/. Last accessed 4 Apr 2016

Quasi-Zenith Satellites System (QZSS) JAXA (2006), http://qzss.jaxa.jp/index_e.html. Last accessed 4 Apr 2016

U.N. Office for Outer Space Affairs, ST/SPACE/50, *Current and Planned Global and Regional Navigation Satellite Systems and Satellite-Based Augmentations Systems*, http://unoosa.org/pdf/publications/icg_ebook.pdf. Last accessed 4 Apr 2016

Working Group Workshop on GNSS Interoperability, Vienna, 30–31 July 2009, http://www.unoosa.org/oosa/en/SAP/gnss/icg/wg/wga02/pres.html. Last accessed 4 Apr 2016

Part III

Space Remote Sensing

Introduction and History of Space Remote Sensing

Scott Madry

Contents

Abstract

This chapter introduces the subject of remote sensing both in terms of its technology and its many applications. Remote sensing via satellite has become a key service that is used in many civil applications such as agriculture, forestry, mining (and prospecting for many types of resources), map making, research in geosciences, urban planning, and even land speculation. Perhaps, one of the most vital uses of remote sensing today is related to disaster warning and recovery. The first use of remote sensing was essentially for military purposes and this remains the case today, and, thus, this chapter addresses these applications as well. Remote sensing, Earth observation, related geographical information systems (GIS), plus the interpretation and use of this type of data are today often referred to today as Geomatics.

This section starts with a history of remote sensing and then continues with a discussion of the technology and its applications. In a number of ways meteorological or weather satellites are essentially a specialized form of remote sensing satellites. Thus the history presented here covers not only what are considered

---

S. Madry (✉)
Global Space Institute, Chapel Hill, NC, USA
e-mail: Scottmadry@mindspring.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_37

remote sensing satellites but meteorological satellites as well. The meteorological satellites are discussed in much greater detail later in this handbook.

Keywords
Corona Cosmos earth resources technology satellite (ERTS) • Geographical information systems (GIS) • Geomatics geostationary orbiting environmental satellites (GOES) • Hyperspectral sensing • Infrared sensing • Landsat Lidar • Multispectral sensing • National oceanic and atmospheric administration (NOAA) • Optical sensing • Pixel • Polar-orbiting environmental satellites (POES) • Polar sun-synchronous orbit • Radar sensing • Remote sensing • Resolution • Sensor SPOT • Television and infrared observation satellite (TIROS) • Zenit satellites of the Soviet Union

Introduction

Humans have always sought the high ground looking for food, exploring new lands, and watching for danger. Seeking the heights is an activity as old as the human species itself, as old as our urge to explore and discover. As technology has continually developed, we have found new ways to provide this important capability with greater precision and with greater coverage. As our capabilities have increased, we have also have found many new uses for that view from on high. Remote sensing can be defined as the art and science of acquiring information from sensors or systems that are not in direct contact with the objects being studied. The first practical remote sensing systems used photographic cameras, or simply the viewer's eyes, from balloons, kites, and even small cameras attached to pigeons. The development of aircraft and the rapid technological innovations of the First and Second World Wars created practical aerial remote sensing capabilities that have helped to map and understand our world.

The dawn of the space age was the next logical extension of this, by now quite important, aerial view of the Earth. Indeed, both the Soviet Union and the USA were quick to see the tremendous potential of space remote sensing for military reconnaissance, weather prediction, environmental analysis, and other applications. The original photographic systems used film and photographic prints to capture data. Over time more and more sophisticated sensors were developed to capture data across an ever broader range of spectra – both above and below visible light and the visual range available to the human eye. Over time an interpreter's skill in analyzing remote sensing data has given way to modern electro-optical systems and digital computers that help us to acquire, store, and analyze data about not only our Earth but our neighboring planets as well.

Over the decades since the first military remote sensing satellite launches in the 1950s, the technology has developed tremendously, and we are now in a very different situation from the Cold-War era. Today, many nations are developing, launching, and operating their own remote sensing satellite systems for a wide variety of civil, commercial, and strategic applications.

Satellite remote sensing has changed the way that we think about our planet and has served as important stabilizing influences in the Cold-War era and perhaps even more so in the post Cold-War era. This versatile technology has great technology transfer potential and has served as the basis for our initial robotic exploration of the moon and other planets as well.

The Purpose of Remote Sensing and Geomatic Systems

This chapter of the handbook will cover what remote sensing is, how it has developed, how it is used, and where the technology and its many applications are going. It is particularly important to note that there is a new term of art called Geomatics. Geomatics refers to all forms of remote sensing that includes the sensing technology, the capturing and display of data on geographical information systems (GIS), and the capturing and interpretation of all forms of remote sensing data.

History of Satellite Remote Sensing Services

The development of space remote sensing is an evolution of mapping and cartography, which has its origins in many cultures around the world. Accurate maps were created and used in ancient Mesopotamia and Egypt over 3,500 years ago. Eratosthenes (276–195 BC) measured the circumference of the Earth and divided the Earth into 60 even grids, and Hipparchus of Nicea (165–127 BC) defined our 360° system that is still in use around the world today. It was in Ptolemy's *Geography* that the world was first defined within an even more precise system of degrees, minutes, and seconds of longitude and latitude. In China, the Zhou emperors (1,100 BC) had royal geographers and Phei Hsiu (267 AD) had a detailed set of maps created that covered all of the lands within the Chinese domain. The era of European exploration in the sixteenth through eighteenth centuries created a flourishing of cartographic techniques and methods that laid the foundation for accurate mapping and analysis of our world.

Even before the development of aircraft, people were finding ways to look from above, and the first recorded aerial photo was taken of Boston, Massachusetts, in 1860 by Samuel Archer King from a balloon. In the American Civil War, balloons were used for military observation, and Dr. Julius Neubronner patented a miniature pigeon camera system in 1903 that worked quite well. Gaspard-Felix Tournachon, known as "Nadar," photographed Paris from his balloon in the 1860s, and Alfred Nobel developed a photo-taking solid rocket in 1897. The US Army Signal Corps used a series of large box kites to photograph the damage of the April 1906 great earthquake and fire in San Francisco (see Fig. 1), California, and the pictures were prominently used by newspapers throughout the USA to rally relief support, the first

Fig. 1 Kite-based photographic image of San Francisco in 1906 (Image courtesy of the US Geological Survey)

known use of remote sensing for disaster response, which is now a very common application (Madry and Pelton 2010).

The development of the airplane and the military needs of the First World War caused a huge improvement in aerial reconnaissance, which was quickly followed by a growing commercial aerial mapping industry. The Second World War produced the next great leap in the development of advanced cameras, films (including the first infrared film to defeat camouflage), image interpretation, and other advanced systems. After the war there were thousands of surplus aircraft and aerial cameras, trained pilots, and landing fields throughout the world that spurred the next leap in aerial mapping systems, and the development of the German A4/V2 rockets laid the foundation for space remote sensing systems.

Modern remote sensing systems all derive directly from Cold-War military needs and capabilities. The Soviet Union and the USA shared an intense desire to acquire reconnaissance capabilities capable of covering large areas deep inside the other's territory. The Soviets authorized their Zenit spy satellites in 1958, only months after the flight of Yuri Gagarin, which used the same Vostok capsule. A total of 10 of the first 20 Cosmos series launches were Zenit systems, carrying four cameras that could acquire 1,500 individual photographs of the Earth per mission. The entire capsule was then deorbited and the cameras refurbished and reused.[1]

At the same time, the USA approved their Corona system in 1958, with the first launch and film recovery in 1960. The Corona was one of the most secret of US military programs and suffered 13 failures in a row before the first useable film was recovered. The exposed film was deorbited in a reentry capsule that was snagged in midair, while the entire satellite was deorbited and burned up upon reentry. The Corona program ran through 1972 and mapped over 5.5 million square miles in over 100 missions, an amazing technological feat at that time. This satellite, in its time, was by far the highest-resolution imaging satellite available in Earth orbit. While it

---

[1] Zenit, *Encyclopedia Astronautica*, http://www.astronautix.com/craft/zenit.htm

Fig. 2 Image from Corona satellite of the US Pentagon (Image from US Defense Department Archives)

was one of the most secret military programs of its day, the Corona images are now available online and are available for use for long-term environmental analysis (see Fig. 2). The direct descendants of these early systems are flying today, continuing to push the development of the technology and providing an important stabilizing factor in global geopolitics.[2]

The first explicitly civil satellite remote sensing system was the US TIROS, the Television and Infrared Observation Satellite. While it started as a defense department initiative, it was transferred to NASA in 1959 and launched in April of 1960. It was not in a polar orbit, as is standard now, and had very poor spatial resolution by our standards today, but it quickly demonstrated the tremendous potential that satellite imaging could have for a variety of applications including weather, ice monitoring, ocean studies, and more. Nine more TIROS satellites followed, and the program was later moved to NASA, and its descendants continue as the National Oceanic and Atmospheric Administration's (NOAA's) Geostationary Operational Environmental Satellite (GOES) and Polar Operational Environmental Satellite (POES) systems.

The Landsat program was the first systematic moderate-resolution civil remote sensing system. It was started in 1970 and the first satellite was launched on July 23, 1972, only 2 years later. The Landsat program was first conceived in 1966 as a direct outgrowth of the successful Mercury, Gemini, and Apollo astronaut

---

[2] Corona Program Summary, FAS Space Policy Project, Military Satellite Programs, http://www.fas.org/spp/military/program/imint/corona.htm

photographs and was originally named ERTS, for Earth Resources Technology Satellite. The Landsat system continues in operation and it has generated years of broad synoptic coverage of our planet and originated hundreds of scientific applications and commercial uses.[3]

France launched their very successful SPOT program in 1986 that became the first system to produce remote sensing data for commercial consumption. [4] Once this process of collecting remote sensing data by satellite for commercial distribution became established, the idea caught on quickly. Today, there are over 30 nations around the world operating a wide range of satellite systems that fall into several general categories. The geostationary weather satellites provide a broad, synoptic coverage of our planet and are primarily used for continental scale weather forecasting and severe weather monitoring. These are operated as governmental organizations as a public good, but many of the other remote sensing satellite systems now offer commercial services to a global customer base although some restrictions often apply.

The optical systems in polar orbit can be grouped into low (1,000–250 m), medium (100–30 m), high (30–5 m), and very high (less than 5 m) spatial resolution. Currently available commercial satellite data of as little as 35–50 cm spatial resolution can be purchased today. This is a truly amazing transition from the first highly classified military satellites to a global commercial market for high-resolution images of the Earth. Active microwave systems using radar are also operating as civil systems, providing a very different and useful view of the Earth. All of these systems share common parameters and provide us with a continuous stream of data about our planet.

The latest evolution is reflected in the deployment of remote sensing satellites with hyperspectral imaging sensors. These satellites instead of sensing a broad range of the spectrum provide the ability to break down imaging into very narrow bands. The data output from these sensors is deployed in so-called data cubes with the “x” and “y” axis being able to depict a spatial area and the “z” axis in the data cube showing the results across narrow spectral bands.

Hyperspectral imaging presents a very difficult issue in terms of processing the data since it requires a broadband downlink to send the data back to Earth, and it presents a very difficult issue as to how to analyze and interpret the mountain of data generated by such imaging. The amount of data provided in hyperspectral sensing (since it may be collecting data over 200 different narrow spectral ranges) is so large, one can really only process by computers. Hyperspectral sensing can produce almost two orders of magnitude greater data than is the case with multispectral sensing. This is because data is collected in very narrow spectral slots (i.e., up to 200) rather than across five to ten broadbands.[5]

---

[3] The Landsat program, http://landsat.gsfc.nasa.gov/

[4] J. Rao, Spot remote Sensing system, http://www.sphttp://www.space.com/6870-spot-satellites.html

[5] Hyperspectral imaging, http://en.wikipedia.org/wiki/Hyperspectral_imaging

The other new development in the field is a new approach to remote sensing that uses small satellites in low Earth constellations to deploy networks at lower costs and to achieve new levels of temporal coverage. The advent of 3-D printing and other advanced manufacturing capabilities allows the lower-cost manufacture of smaller satellites that still remain quite capable. New innovation in launch vehicle design and manufacture by new space manufacturers also allow these networks to be deployed at lower cost. New systems that are now deployed by Skybox and Planet Labs have pioneered this innovative approach to remote sensing satellite networks and have allowed new applications, driven particularly by more rapid coverage of the same location. In addition very high-esolution sensors in GEO-based systems may also redefine the longer-term future of remote system design and operation.

Overview of Satellite Remote Sensing Services

Remote sensing satellites now provide the world with an amazing variety of information about our spaceship Earth. Two primary orbits are used; a Sun-synchronous polar orbit, often between 500 and 800 km, provides high- and moderate-resolution coverage of the entire planet on a regular basis. This special orbit allows a satellite to pass over the same location on the Earth periodically at the same time of day and with the same solar illumination. The satellite has to shift its orbit by approximately 1° per day as the Earth orbits around the Sun to achieve this periodicity. These types of satellites provide data for uses as varied as land-use planning, forestry, military reconnaissance, and natural resources monitoring. Weather satellites in the geosynchronous arc provide a constant view of a portion of the globe at a lower resolution but with a constant view. A ring of these, operated by several nations, now provide a constant worldwide watch over our planet for severe weather and broad climate applications.

By themselves, each of these capabilities is useful, but the integration of different types of remote sensing data, along with other types of information for various sources, is driving many new scientific investigations, practical applications, and commercial markets. Geographic information systems, or GIS, allows the integration of remote sensing data with population and other demographic data, in situ environmental telemetry, GPS positioning, and other relevant data. GIS has now become such an important way of depicting information in a useful and integrated way that it now represents a larger commercial market than remote sensing.[6] There are important and developing political, legal, social, and dual-use issues with all of these technologies, and these will continue to evolve as newer and more powerful systems and applications become available. The fact that over 300 million people now have access to Google Earth is a powerful demonstration of the impact and utility of geospatial data combined with high-resolution remote sensing imagery.

---

[6] M. Frenkel, National Institute of Standards and Technology, *Global Information Systems in Science* http://pubs.acs.org/doi/abs/10.1021/je800877f

Remote sensing is a very complex and technologically advanced process and includes a complex chain of end-to-end data flow and image processing, in order to deliver useful data to the end user. It is quite an achievement that a sensor on a satellite, moving at 8.5 km/s at some 800 km above the Earth, can collect, store, and transmit data to the ground that can be processed into the images we have become familiar with today on Google Earth and other sources.

The fundamental process of extracting useful information from remote sensing data relies upon the fact that objects that are of similar physical makeup can be identified by analyzing the energy that is reflected or emitted from it. This “spectral signature” is unique to each type of matter, and this allows us to discriminate between clean and polluted water, mature or immature crops, healthy or diseased forests, etc.

At the heart of each remote sensing system is a detector, which is the heart of the system. Energy is emitted from the Sun and some small fraction enters the Earth’s atmosphere, interacts with the surface, is reflected or emitted from the surface back through the atmosphere, and falls upon the detector in a satellite after passing through the system’s optical telescope. Most passive remote sensing detectors are designed in accordance with the photoelectric effect, and there will be an emission of negatively charged particles (electrons) when a negatively charged plate of an appropriate light-sensitive material is subjected to a beam of photons.

These electrons can then be made to flow from the plate, collected, and measured as an electrical current or signal. After some noise reduction and other preprocessing, the signal is digitized on the satellite and is transmitted to the ground, where an extensive ground processing system further processes the data to produce the data which are then available for use. Image processing and remote sensing specialists further process the data to extract useful information about our world. Most passive remote sensing data are in a raster, or two-dimensional numerical array format, with a single digital value for each area on the ground where the sensor was focused for some milliseconds. This instantaneous field of view (IFOV) is the pixel size of the data and represents the spatial resolution of the image.

Other important resolution parameters include the temporal resolution (or how often the sensor samples the same location on the ground), the spectral resolution (or which part of the electromagnetic spectrum is measured), and the radiometric resolution (or how finely the signal can be measured). These resolution parameters define the characteristics of the sensor and are used to determine which data are useful for a given application. Active radar and Lidar systems work in a very different manner and are more complex, but, in the end, are processed into similar raster imagery for analysis or integration into a GIS.

There are many exciting new developments in the field of satellite remote sensing. These new developments include new higher-resolution systems operating down to 35 cm per pixel, hyperspectral sensors, high-resolution radar systems, new artificially intelligent processing techniques, as well as a range of new applications. The entire field of geomatics is evolving both in terms of its technology, applications, and data display capabilities. The key remaining issue is the extent to which geomatics is to continue to evolve as largely a governmental service offer to citizens as a public good much like meteorological satellite services or whether it will evolve as a commercial service to support an array of newly evolving commercial markets. Clearly some vital remote sensing services such as to support disaster warning and recovery services and those related to coping with climate change will remain as a public service supporting the public good, but other new services may evolve on a strictly commercial basis.

Conclusion

The chapters that follow in this section seek to provide a comprehensive and interdisciplinary overview of satellite remote sensing technology, services, and applications. These chapters also seek to address to some degree the relevant markets, economics, operations, regulation, and future trends. The specifics of spacecraft design and technology for remote sensing and other application satellites are not addressed in detail here. Many aspects of these technologies have already been addressed in the earlier section on telecommunication satellites. Thus a consolidated approach related to all applications is taken is addressed in Overview of the Spacecraft Bus in this Handbook.

This chapter thus addresses the common technical elements found in essentially all application satellites in terms of power systems, spacecraft bus technology and structural design, and onboard thruster systems. Specifically, Section 6 addresses solar, battery, and nuclear spacecraft power systems; thermal balancing and heat dissipation systems; orientation, pointing, and positioning systems; structural design elements; diagnostic systems; tracking, telemetry, and command systems; manufacturing and integration; and quality and reliability testing processes.

Launch Vehicles and Launch Sites addresses how the various application satellites are launched by different rocket systems from various launch sites around the world and the differences between solid, liquid, and hybrid launching systems.

Remote sensing satellites represent the smallest of the various space applications in terms of market size, but the application of this technology is of utmost importance. Today remote sensing satellite is used to undertake disaster recovery and carry out environmental and climate change science, while remote satellite imaging is actively applied to an ever-increasing number of economic fields that include agricultural, forestry, fishing, mining, and industrial applications and services. The commercial market for satellite-based remote sensing may be small in comparison to a field like space telecommunications, but the critical importance of the global applications makes this field enormously large.

Cross-References

▶ Astronaut Photography: Handheld Camera Imagery from Low Earth Orbit
▶ Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing

- ▶ Electro-Optical and Hyperspectral Remote Sensing
- ▶ Fundamentals of Remote Sensing Imaging and Preliminary Analysis
- ▶ Geographic Information Systems and Geomatics
- ▶ Launch Vehicles and Launch Sites
- ▶ Lidar Remote Sensing
- ▶ Operational Applications of Radar Images
- ▶ Overview of the Spacecraft Bus
- ▶ Processing and Applications of Remotely Sensed Data
- ▶ Remote Sensing Data Applications

References

S. Madry, J. Pelton, Satellites in the service to humanity, in *The Farthest Shore: A 21st Century Guide to Space*, ed. by J.N. Pelton, A. Bukley (Apogee Press, Burlington, 2010)

Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing

Michael J. Rycroft

> *"And God said, 'Let there be light,' and there was light."*
> *Genesis, chapter 1, verse 3, The Holy Bible, New International Version*

Contents

Abstract

Here, we consider a topic which is absolutely central to the successful operation of all satellites and spacecraft, namely, the basic principles and fundamental concepts of visible light in particular and of electromagnetic radiation in general. Both the wavelike nature of light (the speed of light being 300,000 km/s through free space) and its particle-like nature (as photons) are considered. We introduce its wave properties which explain the phenomena of reflection, refraction, diffraction, interference, polarization, and the Doppler effect. The photon properties explain blackbody radiation, continuous spectra, emission spectra, absorption

---

M.J. Rycroft (✉)
Cambridge Atmospheric, Environmental and Space Activities and Research (CAESAR)
Consultancy, Cambridge CB3 9HW UK
e-mail: michaelrycroft@btinternet.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_38

spectra, and the photoelectric effect. We mention how electromagnetic radiation is used actively for radio communications with Earth-orbiting satellites and passively for remote sensing investigations not only of the atmospheres of the Earth and other planets but also of distant stars and the structure of Universe.

Keywords

Electromagnetic radiation • Electromagnetic waves • Frequency • Gamma-rays • GPS • Infrared • Light • Milky Way • Photons • Radar • Radio waves • Spectra • Sun • Ultraviolet • Universe • Velocity of light • Wavelength • X-rays

Introduction

Our human eyes are tuned to receive radiation from the Sun, in the form of light. Light is one type of electromagnetic radiation which, from many different viewpoints, is crucial for space studies and space applications. Thus an understanding of both electromagnetic waves and the electromagnetic spectrum is essential for all readers of this volume. Electromagnetic radiation is used for both active and passive studies carried out from space – in the former, waves are transmitted and received (e.g., in a radar altimeter) and, in the latter, they are only received, for example, from a star or from the Sun, the star of our solar system. Electromagnetic radiation from the Sun, whose intensity peaks in the visible part of the spectrum, is the source of energy for almost all processes occurring on the Earth, and for all life on Earth, for the fauna and flora which abound. Light takes about 8 min to travel the $\sim$150 million kilometers from the Sun to the Earth; thus, when we observe the Sun we see it as it was 8 min ago.

Nowadays, electromagnetic waves give us much information on the Earth's atmosphere and its weather systems and on the atmospheres of our neighboring planets. Such passive remote sensing studies are carried out using the observations made by diverse instruments on geostationary or polar-orbiting satellites or aboard spacecraft. Such studies also give us valuable information on the Earth's sea and land surfaces, such as its vegetation cover and the development of urban sprawl.

Radio waves are used to command satellite operations from the ground. They are also used in order to transmit scientific data from satellites, as modulated telemetry signals, to ground stations or to transmit very high frequency radio signals or TV broadcasts from space to different parts of the globe. Radio waves are essential for the operation of positioning and navigation services using satellites, such as provided by the GPS (Global Positioning System) signals now commonly received by the ubiquitous "sat nav" equipment in our cars. A thorough appreciation of the behavior and properties of electromagnetic waves is therefore crucial when discussing the performance and reliability of applications satellites.

The Fundamentals

Visible light is a wave motion, for which electric and magnetic fields oscillate extremely rapidly. Two cycles, or two periods, of a sinusoidally varying wave motion are illustrated in Fig. 1. In this diagram, the horizontal distance between two successive crests of the wave is one wavelength of the wave, which is denoted by the Greek letter λ. The vertical distance between the maximum and the minimum shown is twice the amplitude of the wave. The intensity, or power, of the wave is proportional to the product of the wave electric field amplitude and the wave magnetic field amplitude. Light carries energy radially away from a source, with the light rays traveling in straight lines. The wave electric and magnetic fields are perpendicular to each other, and they are both perpendicular to the direction of propagation of the wave. That result is obtained by solving Maxwell's four equations of electromagnetism (Grant and Phillips 1990).

Light travels at an incredibly high speed. In free space (a vacuum), the velocity of light, c, equals 3×10^8 meters per second (m/s), or 300,000 km/s. For yellow light, its wavelength λ has the value of 600 nm, that is, 6×10^{-7} m, or 0.6 μm. The velocity of light, c, is equal to the product of (i.e., is obtained by multiplying) the frequency f of vibration, or oscillation, of the wave (in cycles/s, now termed Hertz, Hz) and the wavelength λ (in m). The equation $f = c/\lambda$ is the fundamental relation connecting these three quantities.

Inserting the numbers for yellow light, $f = 5 \times 10^{14}$ Hz, 500 million million vibrations per second. Now blue light has a wavelength ~400 nm and red light ~700 nm, which differ by nearly a factor of two – in between blue and red are all the colors of the rainbow. Their frequencies are 7.0×10^{14} Hz (blue) and $\sim 4.3 \times 10^{14}$ Hz (red). We can only see radiation in the visible (Vis) part of the spectrum, with wavelengths between 400 and 700 nm. At shorter wavelengths, there is ultraviolet (UV) radiation, with infrared (IR) radiation at longer wavelengths, as shown in the third line down from the top of Fig. 2 that shows the entire electromagnetic spectrum.

The electromagnetic spectrum stretches from the lowest radio frequencies, 8 Hz (to the left of the bottom of Fig. 2), where the wavelength is equal to the circumference of the Earth (~40,000 km), through radio frequencies, ~100 MHz (or 10^8 Hz) where frequency modulated (FM) commercial radio stations operate, through the microwave (radar) part of the spectrum, at 10 GHz (10^{10} Hz, often termed X-band),

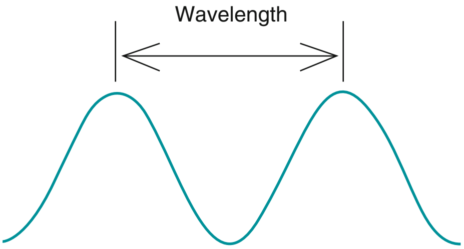


Fig. 1 Diagram showing one wavelength of a wave motion (From www.nrao.edu/index.php/learn/radioastronomy/radiowaves, courtesy of NRAO)

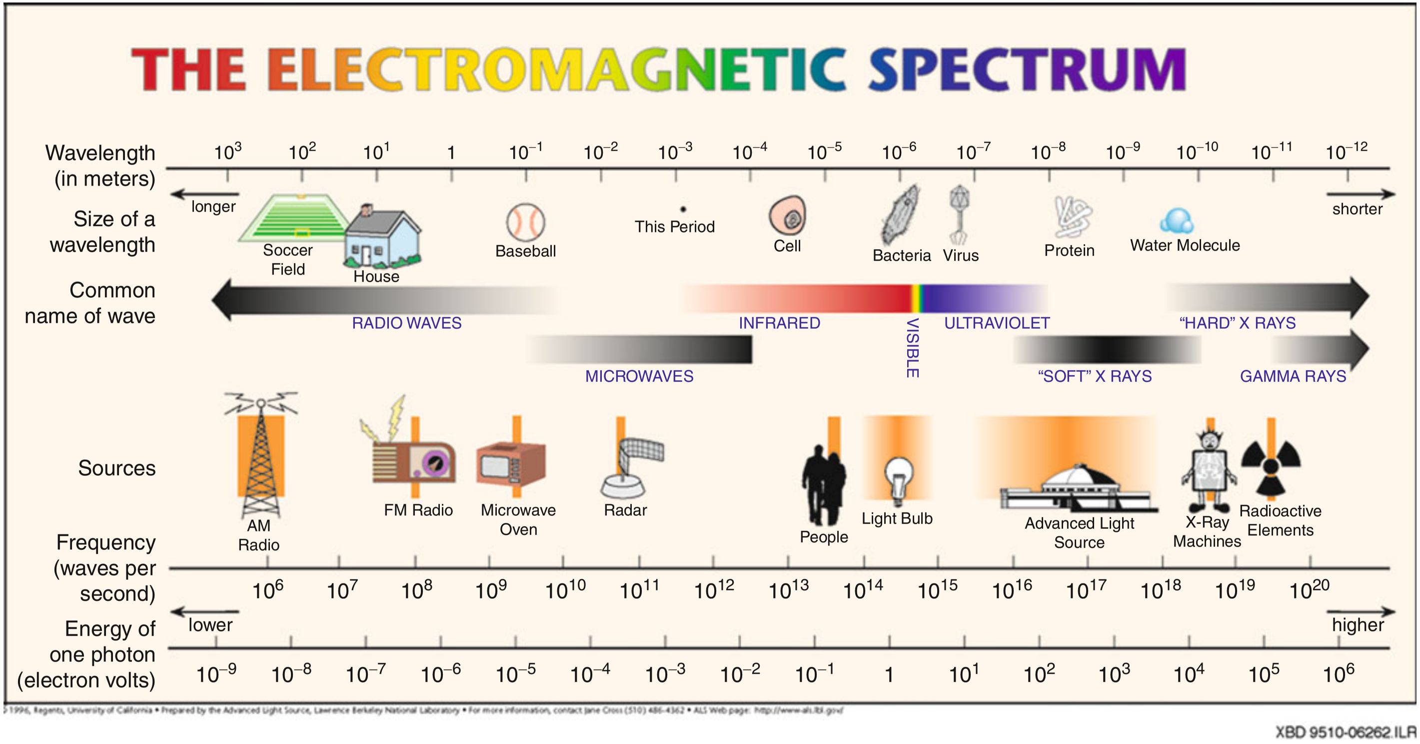


Fig. 2 Diagram showing the full electromagnetic spectrum, in all its glory; the wavelengths (*top*) range from 1 km on the *left* to 10^{-12} m on the *right*, and the corresponding frequencies, given at the *bottom*, go from radio waves of 300 kHz on the *left* to 3×10^{20} Hz gamma-rays on the *right*. The second row from the *top* illustrates different typical objects whose size is that of the corresponding wavelength. Different sources of electromagnetic radiation are shown above the frequency scale, which is itself above the scale specifying the energy (in electron Volts, eV) of one photon of that radiation (From www.lbl.gov/MicroWorlds/ALSTool/EMSpec?EMSpec2.html, courtesy of The Advanced Light Source, Lawrence Berkeley National Laboratory)

having wavelengths ~3 cm, and up to the infrared. Beyond the visible, the electromagnetic spectrum stretches up in frequency another million times, through X-rays and gamma-rays, radiation whose wavelengths are as small as a millionth of the wavelength of visible light.

It might be helpful to some readers to provide a musical – acoustic – analog of the visible part of the electromagnetic spectrum. The note of middle C on the piano has a frequency of 256 Hz, and the C an octave above has a frequency twice that, namely, 512 Hz. The full range of a piano is from the lowest bass, three and a bit octaves below middle C, at ~25 Hz to the top treble (three octaves above 512 Hz) at ~4 kHz, that is, it covers a range of 160 times (1.6×10^2 times) in frequency. So the full electromagnetic spectrum from 10 to 10^{21} Hz, that is, over 20 orders of magnitude in frequency, is equivalent to having five more bass "pianos" and three more treble "pianos," as well as the actual piano! The electromagnetic spectrum indeed spans an incredibly wide frequency range.

At higher frequencies it is valuable to think of light as a particle as well as a wave. A beam of light is then represented as a stream of particles, each of which is called a photon. A photon is both massless and without an electric charge; it travels at the velocity of light. The energy of an individual photon, a quantum of energy, which is shown on the bottom line of Fig. 2, is a well-defined quantity. It is given by the product of a fundamental constant known as Planck's constant (6.6×10^{-34} Joules. seconds, and abbreviated as Js) and the frequency of the electromagnetic wave of interest (in Hz). Therefore the energy of one photon is directly proportional to its frequency; for yellow light the energy of one photon is $\sim 3 \times 10^{-19}$ J, which may also be expressed as ~2 electron Volts (eV). A photon of blue light has an energy that is almost twice that of a photon of red light. X-rays have very much greater photon energies, ~1–100 keV (thousands of eV), with gamma-ray photon energies exceeding 1 MeV, that is, >1 MeV (millions of eV).

Reflection, Refraction, Diffraction, Interference, and Polarization: Important Properties of Electromagnetic Waves

When a beam of light strikes a mirror, it is reflected. The angle that the beam makes to the perpendicular to the mirror surface (which is sometimes called the normal to the surface), called the angle of incidence (i), is equal to the angle which the reflected beam makes to the perpendicular, known as the angle of reflection (r_l). This relation shows a critical property of the phenomenon of reflection. For a rough surface, the reflected waves have different directions so that we often consider the waves to be scattered by the surface.

When electromagnetic waves travel through any material medium, they may be partially absorbed. For waves traveling through a transparent medium, such as glass, rather than through a vacuum, the velocity of light becomes less than c. In fact, we can write an equation for the velocity of propagation through a medium, v, as $v = c/\mu$, where μ is called the refractive index of the medium; its value is always greater than 1. For glass, μ is ~1.5, and its value is larger for blue light than for red light.

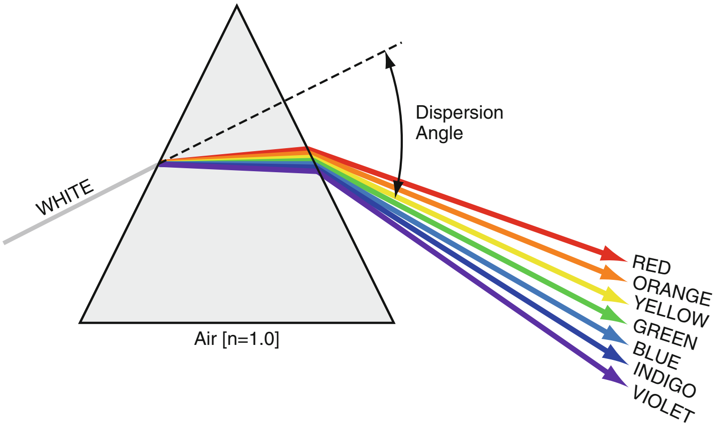


Fig. 3 Diagram illustrating how a glass prism separates a beam of white light into rays having all the colors of the rainbow. This is explained by the phenomenon of refraction (From www.thescienceclassroom.wikispaces.com, or www.heasarc.nasa.gov, courtesy of NASA)

In the year 1665, Isaac Newton carried out experiments on passing a beam of sunlight, white light, through a glass prism. As Fig. 3 indicates, the light beam does not travel along the dashed line, but it is bent, that is to say it is refracted, so that it emerges from the prism at an angle to the dashed line; this is called the dispersion angle. The ray is bent toward the normal to the prism surface, making an angle to the normal of r_r. Going from air, whose refractive index is unity, to glass of refractive index μ, the relation sin i = μ.sin r applies. The blue light is dispersed more than the red light. The beam of white light is split into all the colors of the rainbow by the glass as Fig. 3 shows. Thus the phenomenon of refraction allows us to investigate the spectrum radiated by a light source of interest, such as a star.

Figure 4 shows the spectrum of light emitted by the Sun, from blue at 400 nm at the bottom to red at 700 nm at the top. Each of the 50 horizontal lines shows the spectrum for a width of only ~6 nm. The Sun emits a broad – continuum – spectrum. On this continuum spectrum, dark – absorption – bands appear at generally very well-defined wavelengths. How these are formed will be mentioned later in this chapter.

Knowing about refraction makes it possible for us to design a lens which brings a beam of parallel light to a focus, for example, in the eyepiece of a telescope. With combinations of lenses, prisms, and mirrors, we can design telescopes of several different types (e.g., Newtonian, Cassegrain, or Coudé) to view distant astronomical objects. These range in complexity and performance from the first telescope of Galileo Galilei made in 1609 to view the Sun, when he discovered sunspots that are dark regions on the solar surface, to the Hubble Space Telescope or the Chandra X-ray telescope. Radio telescopes use metal parabolic reflectors to bring the radio beam from a satellite or a distant radio galaxy to a focus, where a sensitive receiver is placed.

The ionospheric plasma is the naturally occurring mixture of positive ions and electrons formed by the action of sunlight in the atmosphere at heights above ~80 km. The refractive index of this plasma is determined in part by the

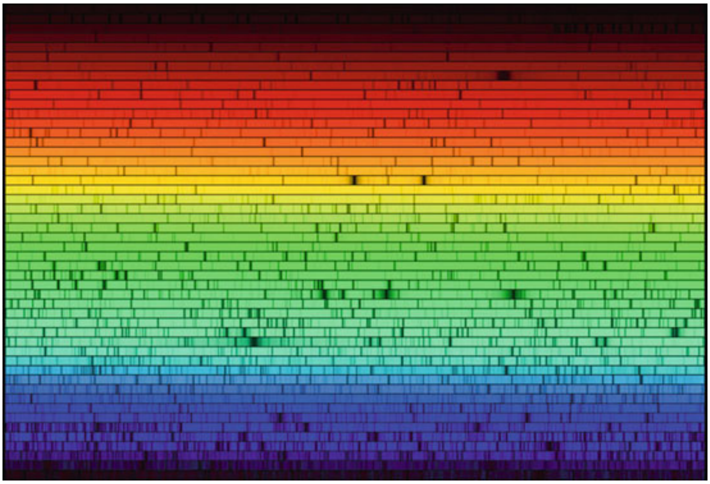


Fig. 4 The Sun's spectrum recorded by an instrument termed a spectrometer (From www.noao.edu, courtesy of NOAO/AURA/NSF)

concentration of electrons. In order to travel through (i.e., not be reflected by) the ionosphere, the command and/or telemetry radio signals must have frequencies exceeding the largest value of the electron plasma frequency along the propagation path from the ground to the satellite, or vice versa. That means that their frequency must exceed ~30 MHz. If the radio frequency is much larger than that, say, ~1 GHz, the refractive index is only slightly larger than unity. In fact, there are two values of the refractive index due to the presence of the Earth's magnetic field, which allows two types of wave to propagate. These are termed ordinary and extraordinary waves.

The performance of every telescope is limited by the phenomenon of diffraction. When light goes through a circular aperture or a slit, it does not travel straight through but is diffracted by some small angle. This phenomenon of diffraction limits the resolving power of a telescope; it determines the angular separation between two nearby objects in the sky that can be distinguished from one another. The 2.4 m diameter of the Hubble Space Telescope determines that its resolving power is equivalent to resolving two pinpricks of yellow light only 1 mm apart at a distance of 2 km, an incredible achievement.

The phenomenon of diffraction and of interference between light waves enables a diffraction grating to be created. This acts like a prism, bending light. Diffraction gratings are often used as spectrometers.

Another important property of an electromagnetic wave is its polarization. For a wave whose electric field always lies in the same direction, the wave is said to be linearly polarized. Alternatively a wave whose plane of polarization rotates as the wave propagates is called a circularly (or elliptically) polarized wave. The rotation can be either clockwise or counterclockwise; this property is what causes ordinary and extraordinary waves to exist. During propagation, the plane of polarization rotates – this is called the Faraday rotation of the plane of polarization. It enables the total electron content along the radio path between a satellite and a ground station to be calculated.

The Doppler Effect

When an observer is moving at a velocity v relative to a source of light, or if the source is moving relative to the observer, then the observer will notice a change in the wavelength of the light. Motion along the line of sight, away from the observer, causes an increase of the wavelength – this is termed a red shift. However, if the motion is toward the observer, a decrease of the wavelength is caused, termed a blue shift. This phenomenon is known as the Doppler effect. The reader may be more familiar with the acoustic analog. The siren of a police car approaching the observer increases in pitch, or frequency, whereas when it is moving away the frequency decreases below the transmitted frequency.

A useful equation is that the magnitude of $\delta\lambda/\lambda = \delta f/f = v/c$, for values of v which are very much less than (<<) c. Measuring the Doppler frequency shift of well-known spectral lines (discussed in the next section) leads directly to an estimate of the source velocity. The application of this result has demonstrated that the Universe is expanding, and that the velocity of more distant objects is larger than that for nearer objects. A space applications example of the Doppler effect is the changing frequency of a radio signal transmitted by a polar-orbiting satellite traveling at 7 km/s; it is increased by up to ~7/300,000, or ~0.002 %, as it approaches the ground station and is decreased by that amount as the satellite moves away.

Multiwavelength Studies and Black Body Radiation

Dust, clouds, and various molecular gases, including water vapor, the most important "greenhouse gas," and carbon dioxide, in the Earth's atmosphere absorb much of the Sun's infrared radiation. Molecular oxygen and ozone in the stratosphere absorb almost all the dangerous – to the DNA molecules in our bodies – ultraviolet radiation from the Sun. Therefore, the only way to carry out experimental gamma-ray, X-ray, ultraviolet, or infrared astronomical studies is to put an instrument aboard a rocket, satellite, or spacecraft. Space technology is essential for such fundamental scientific studies.

Where in the spectrum there is little absorption of radiation is termed a spectral window. The visible (Vis) part of the spectrum is one such region; that is clearly shown between 400 and 700 nm in the upper part of Fig. 5. There is a broad radio window at wavelengths between 0.1 and 10 m, which allows ground-based radio telescopes to operate, as illustrated in the lower part of Fig. 5. Longer wavelength radio waves cannot penetrate the ionosphere; they are reflected by the ionosphere.

Figure 6 plots the atmospheric opacity quantitatively as a function of frequency. An absorption of 1 dB (decibel), shown here as 1.E + 00, is rather negligible. However, 20 dB (two times 1.E + 01) shows absorption of the power of the radio

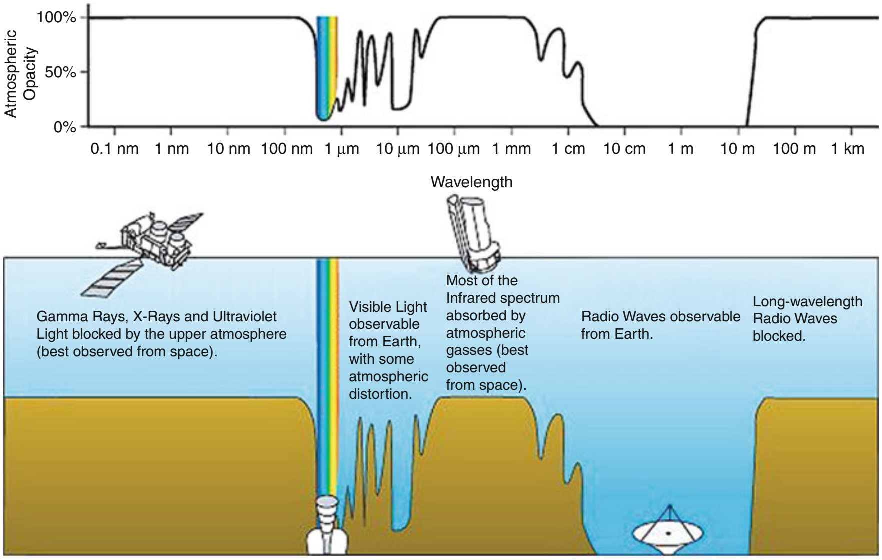


Fig. 5 The *upper* panel shows the percentage absorption (or opacity) due to the atmosphere, which is 100 % for gamma-rays, X-rays, and for ultraviolet, infrared radiation and very long wavelength radio waves. The *lower* panel demonstrates why space technology is required for fundamental scientific studies in these spectral ranges (From http://www.newworldencyclopedia.org/entry/Electromagnetic_spectrum)

signal by one order of magnitude (i.e., ten times); this would be noticeable. Absorption of 100 dB (i.e., by five orders of magnitude, a hundred thousand times, 1.E + 02) would be devastating for any communications system. Thus, this plot informs a communications engineer about those frequencies (such as around 60 GHz and near 118 GHz) which should definitely not be chosen for an effective communications system.

There are three basic types of spectra – the continuous spectrum, the emission (or bright-line) spectrum, and the absorption (or dark-line) spectrum. The continuous spectrum, an uninterrupted sequence of wavelengths, is a broadband spectrum; it is emitted by a hot gas at high pressure (e.g., the photosphere at the Sun's surface). The emission spectrum emitted by an atomic gas (e.g., hydrogen) at low pressure, that is, under rarefied conditions, is a set of discrete bright narrow lines. A particular spectral line is radiated when an electron in a certain excited state falls back to a lower level state or to its ground level state. (An electron is said to have been excited when a photon collided with it, in an atom or molecule, and raised it to a higher energy level.) Spectral lines occur at well-defined wavelengths; they are a characteristic of the particular chemical species radiating them. For molecular gases, the emissions are broader bands rather than narrow lines.

The absorption spectrum is observed when light from a bright source producing a continuous spectrum passes through a cooler gas which absorbs its characteristic line radiation. Radiation at these well-defined wavelengths is removed from the spectrum.

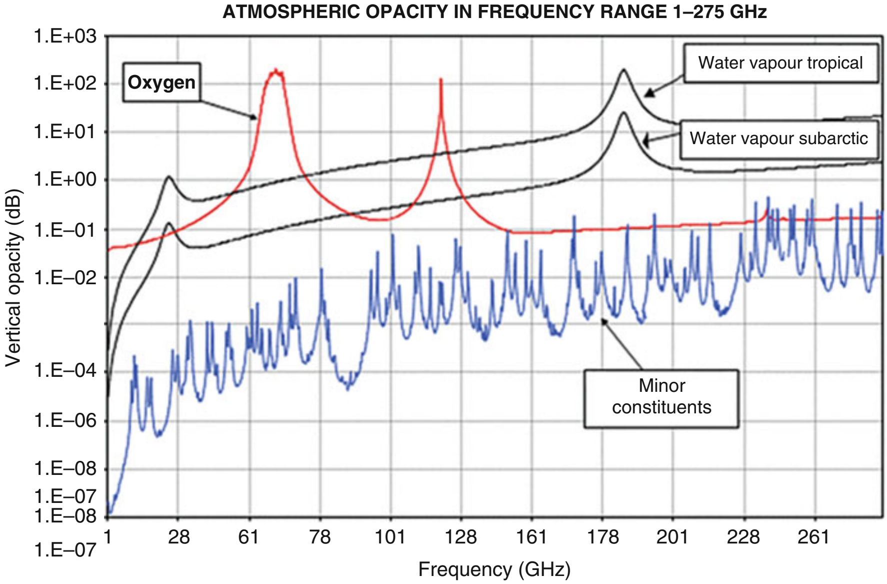


Fig. 6 The atmospheric opacity in the radio part of the spectrum from 1 to 275 GHz; here molecular oxygen and water vapor are the agents most responsible for the absorption (From ITU Report, RO7-SG07-C-0104!!MSW-E, page 11)

Figure 4 presented an example of an absorption spectrum for the Sun. Radiation is absorbed by the cooler gases in the chromosphere just above the solar surface (the photosphere). The thousand or so absorption lines evident in Fig. 4 are called Fraunhofer absorption lines. The most dominant pair of absorption lines is near the center of Fig. 4, in the yellow part of the spectrum, at wavelengths of 589.0 and 589.6 nm. These are due to absorption by atoms of sodium in the solar atmosphere.

The intensity – or brightness – spectrum of a continuum spectrum known as black body radiation, Planck's radiation law, is plotted as a function of wavelength in Fig. 7. A black body is an object that absorbs all the radiation which is incident upon it; it reradiates that energy as radiation which depends solely on the temperature of the object. Stefan-Boltzmann's law states that the total brightness of a black body (the area under the curve) is proportional to the fourth power of the absolute temperatures of that body. The wavelength of the spectral peak also depends on the temperature, expressed in degrees on the absolute (Kelvin) scale. On this scale, the freezing point of water is 273 K and its boiling point is 373 K. The origin of the Kelvin temperature scale is at −273 °C; a negative absolute temperature is impossible.

The range of wavelengths displayed on a logarithmic scale in Fig. 7 is enormous, 14 orders of magnitude. The range of intensities, also shown on a logarithmic scale (but not given quantitatively), is even larger, at least 30 orders of magnitude. The curve shown at 300 K is representative of the black body radiation emitted by the Earth and its atmosphere into space. Most of this radiation lies in the infrared part of the spectrum at wavelengths of ~1–10 μ. The curve at 6,000 K is close to the Sun's

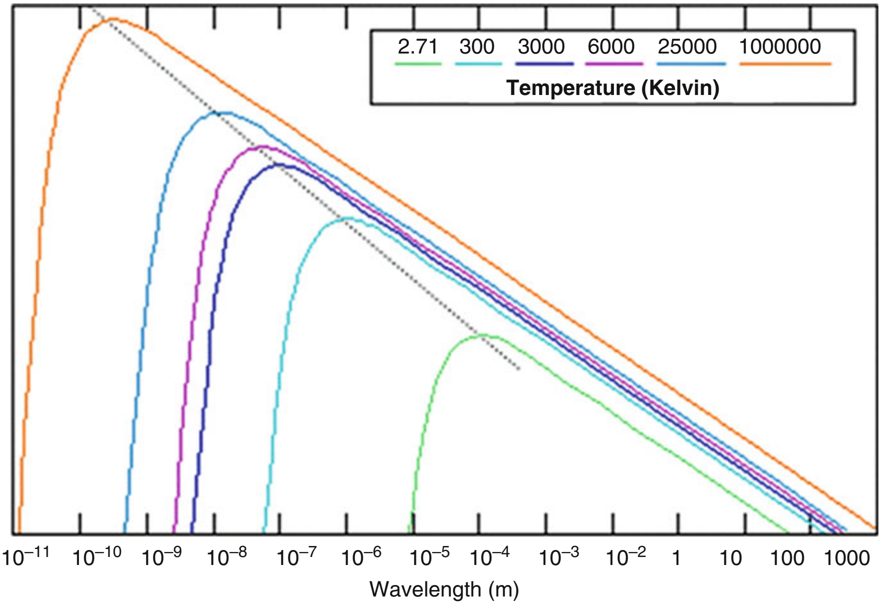


Fig. 7 The intensity (or brightness, on a vertical logarithmic scale of over 30 orders of magnitude) of the continuum spectrum (shown on the horizontal axis using a logarithmic scale from 10^{-11} to 1,000 m, a staggering wavelength range of 14 orders of magnitude) radiated by a black body at different temperatures. These range from microwaves approaching the far infrared, at 2.71° absolute (Kelvin, K), which is the temperature of the cosmic microwave background radiation remaining from the "big bang" origin of the Universe, to X-rays with wavelengths <1 nm at a temperature of a million degrees, which is the temperature of the Sun's outermost atmosphere, the corona (From http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=35774&fbodylongid=1696, or from the encyclopedia of science, www.daviddarling.info)

continuous spectrum; here most radiation is emitted at wavelengths <1 μ. The curve at a million K peaks in the X-ray part of the spectrum. The Sun's outer atmosphere, termed the corona, together with compact and very energetic stars, such as neutron stars, or material falling into a black hole, all emit X-rays. X-ray telescopes in space such as the Chandra Observatory investigate these areas. Figure 8 presents the spectral regions studied by the Chandra Observatory and by other space missions.

The curve below the Earth's spectrum in Fig. 7 is for black body radiation at 2.71 K. This radiation in the microwave part of the spectrum comes from the remnant of the "big bang" origin of the Universe 13.7 billion years ago. The dashed line in Fig. 7 shows Wien's displacement law; this law states that the wavelength of the most intense emission, the peak, is inversely proportional to the absolute temperature. The three fundamental laws of black body radiation introduced here can be proved only on the basis that the concept of radiation as streams of photons is valid.

Figure 9 presents a remarkable composite view of the color-coded intensities of electromagnetic radiation coming from all directions of our galaxy, the Milky Way, in different wavelength regions. The most energetic gamma-rays are shown at the

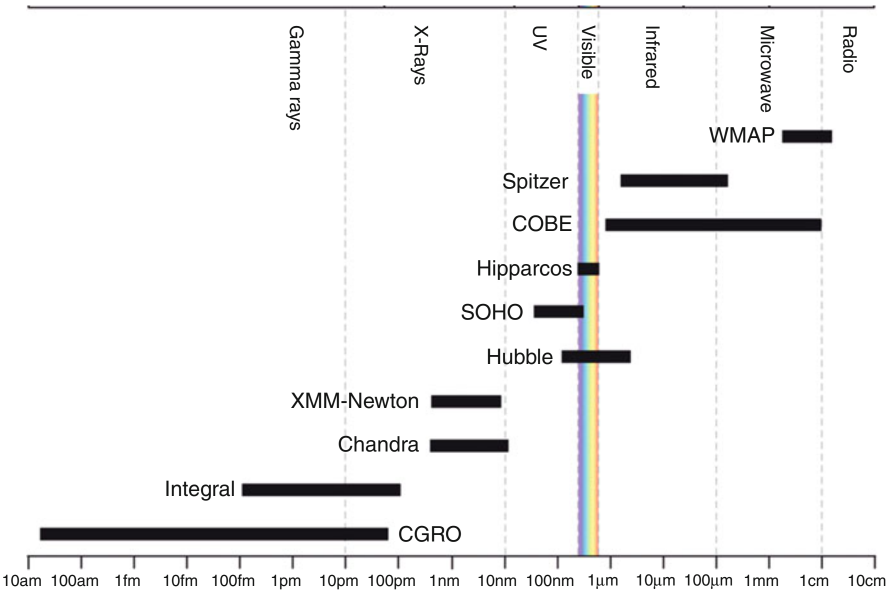


Fig. 8 Diagram indicating the wavelength ranges over 16 orders of magnitude (from 10 attometers, 10 am, or 10^{-18} m, to 0.1 m, or 10 cm) investigated by different space-borne instruments and missions. CGRO refers to the NASA Compton gamma-ray observatory, XMM to the ESA X-ray Multi-Mirror Newton observatory, and SOHO to ESA's solar and heliospheric observatory; COBE (Cosmic Background Explorer) and WMAP (Wilkinson Microwave Anisotropy Probe) both study the cosmic microwave radiation from the "big bang" origin of the Universe (From L.L. Christensen, R. Hurt, R. Fosbury, Hidden Universe, Wiley-VCH)

bottom, with radiation at X-ray wavelengths (e.g., from neutron stars) the next up, then visible radiation, and then infrared radiation from stars that are cooler than our Sun (in three different wavelength regions). Further up in Fig. 9 are displayed intensity maps of radio emissions at various frequencies, including that for atomic hydrogen at 1.42 GHz (at the wavelength of 0.21 m) which is emitted in the cold interstellar medium; it is strongest in star-forming regions. The lowest energies (lowest frequencies and longest wavelengths) are evident at the top. At all wavelengths, there is more radiation coming from the plane (disk) of the galaxy, where most of the $\sim 10^{11}$ stars are to be found. The radiation is especially strong toward the center of the galaxy, at the midpoint of these horizontal images.

Another Effect due to Photons, the Photoelectric Effect

Heinrich Hertz observed in 1887 that a charged metal surface exposed to ultraviolet light lost its electric charge. He had found that the illuminated metal emitted electrons; this effect is called the photoelectric effect. In 1921, Albert Einstein was awarded the Nobel Prize for physics for his 1905 theory that explained the

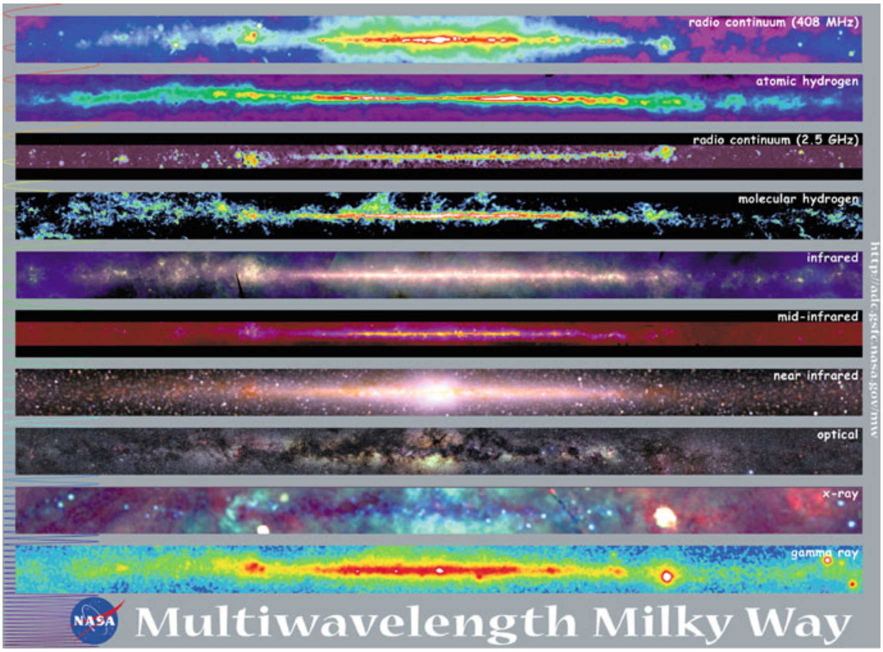


Fig. 9 A view of the entire Milky Way, our galaxy, shown in ten different wavelength regions; the most energetic processes are shown at the *bottom* and the least energetic at the *top*. Some observations are made from the ground and some from space. The central plane of the galaxy is clearly evident in all images, with the galactic center being especially evident in the *top* and *bottom* images (From http://son.nasa.gov/tass/content/emspec.htm, courtesy of NASA)

photoelectric effect. His theory requires that light is considered as a beam of photons, each of which has an energy equal to hf (in J), f being the frequency of the light and h Planck's constant. Electrons which are bound within metal atoms require at least this minimum energy, = hf, a few electron Volts (eV), to be ejected as photoelectrons from the surface of the metal (Fig. 10) and to cause a photoelectric current to flow. A lesser photon energy causes no photoelectric effect whatsoever. A greater intensity of light whose energy exceeds the threshold energy causes a greater photoelectric current. Different metals (different elements) have different threshold energies.

The photoelectric effect causes a satellite or spacecraft having a metal surface which is exposed to the UV radiation and X-rays contained in sunlight to gain a positive electric charge, through the loss of negatively charged electrons. There is a corresponding negative charge in the surrounding plasma in regions of shadow. If this charge becomes large, an electric discharge – a mini lightning discharge – can occur between the two electric charges of opposite sign; that can damage sensitive electronic equipment aboard the satellite. Other sources of danger to equipment operating in space aboard satellites and spacecraft are considered in the chapter "▶ Space Weather and Hazards to Application Satellites."

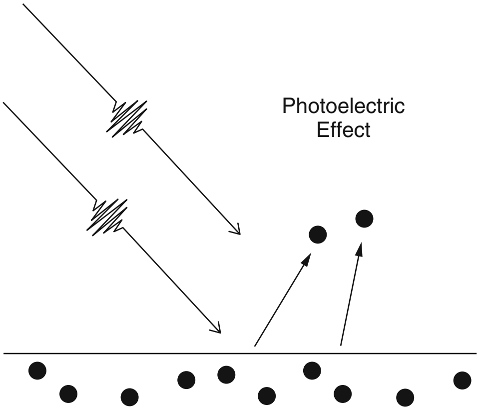


Fig. 10 Diagram showing two photons (wave packets) of ultraviolet light striking a metal surface and ejecting electrons (*dark dots*) from it, illustrating the photoelectric effect (From www.canadaconnects.ca/quantumphysics/1078/)

Conclusion

In this chapter, we have considered the principles and fundamental concepts of light and electromagnetic radiation as they pertain to the performance and reliability of applications satellites. The topic is a central and essential one. Visible light is the best known example of electromagnetic radiation. Its narrow spectrum from blue to red (with wavelengths varying by less than a factor of two) contrasts markedly with the enormously broad electromagnetic spectrum stretching from radio waves to gamma-rays (where the range of frequencies covered is by 20 orders of magnitude).

We have summarized both the wave and the particle (photon) properties of light and, more broadly, of electromagnetic radiation. We have discussed six different phenomena occurring when light interacts with matter, as well as the concept and basic properties of black body radiation. We have explained the Doppler effect. Almost all our knowledge of the Universe, and much of our understanding of the Earth's and other planetary atmospheres, derives from the passive reception of electromagnetic radiation emitted by atoms and molecules in stars, other celestial objects, and the atmospheres of planets. Both one way and two way radio communications between the ground and rockets, satellites, or space probes, and also the operation of GPS systems and radar altimeters, depend upon the active generation, transmission, and reception of electromagnetic waves.

References

I.S. Grant, W.R. Phillips, *Electromagnetism*, 2nd edn. (Wiley, Chichester, 1990), p. 542

Astronaut Photography: Handheld Camera Imagery from Low Earth Orbit

William L. Stefanov, Cynthia A. Evans, Susan K. Runco, M. Justin Wilkinson, Melissa D. Higgins, and Kimberly Willis

Contents

W.L. Stefanov (✉) • C.A. Evans • S.K. Runco
Astromaterials Research and Exploration Science Division, Exploration Integration and Science Directorate, NASA Johnson Space Center, Houston, TX, USA
e-mail: william.l.stefanov@nasa.gov; cindy.evans-1@nasa.gov; susan.k.runco@nasa.gov

M.J. Wilkinson
Texas State University/Jacobs Contract at Astromaterials Research and Exploration Science Division, Exploration Integration and Science Directorate, NASA Johnson Space Center, Houston, TX, USA
e-mail: justin.wilkinson-1@nasa.gov

M.D. Higgins
Jacobs at Astromaterials Research and Exploration Science Division, Exploration Integration and Science Directorate, NASA Johnson Space Center, Houston, TX, USA
e-mail: melissa.higgins@nasa.gov

K. Willis
Oceaneering at Astromaterials Research and Exploration Science Division, Exploration Integration and Science Directorate, NASA Johnson Space Center, Houston, TX, USA
e-mail: kim.willis-1@nasa.gov

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_39

Abstract

Photographic observations of the Earth by humans in low earth orbit, in contrast to unmanned orbital sensor systems, began during the 1960s as part of both the USA and former USSR manned space flight programs. The value of regularly repeated photographic observations of the Earth from orbit was demonstrated by later long-duration missions and led directly to the development of unmanned, multispectral orbital sensors such as the Multispectral Scanner and Thematic Mapper on board the Landsat series of satellites. Handheld imagery of the Earth has been continually acquired during both USA and USSR/Russian space station and former Space Shuttle programs and represents a rich dataset that complements both historical and current unmanned sensor data for terrestrial studies. This revised chapter provides an overview of astronaut/cosmonaut imagery and development of specific data collection programs, then moves on to discussion of technical aspects of both the historical film and current digital cameras used in orbit with information on how to access online datasets. Case studies are presented to highlight varied applications of handheld imagery for terrestrial research and natural hazard monitoring. Developments in time-lapse sequence photography, full georeferencing of astronaut photographs, and involvement with international disaster response efforts are discussed. The chapter concludes with discussion of future directions for digital handheld imagery of the Earth from manned orbital platforms such as the International Space Station (ISS).

Keywords

Astronaut • Cosmonaut • Camera • Space Shuttle • International Space Station (ISS) • Film • Human Space Flight • Mir • Skylab • Apollo • Geology • Geography • Oceanography • Atmospheric Science • Meteorology • Hydrology • Ecology • Urban • Hazards • National Aeronautics and Space Administration (NASA)

Introduction

Remote sensing – the detection of surface-material properties such as composition and texture without physical interaction with the material – is an important analytical approach for investigating and monitoring planetary surface processes. For many Earth scientists, remotely sensed data is strongly associated with unmanned satellites. Remotely sensed data typically is collected by automated sensors on satellites in high polar and sun-synchronous orbits at approximately 700–900 km altitude above the earth's surface. Following the advent of civilian, government-supported orbital platforms in 1972 (e.g., Landsat) remotely sensed data has become an important resource for Earth scientists.

There is another spatially and temporally extensive remotely sensed dataset available for terrestrial studies and applications: systematic photographic images of the Earth taken by astronauts from the NASA Mercury missions of the 1960s to the present International Space Station (ISS) crews. The astronaut photography dataset covers much of the earth's land and coastal surface, as well as including atmospheric phenomena such as hurricanes and aurora. Unlike the unmanned satellite-based sensors mentioned above, astronauts have used off-the-shelf equipment (film and digital cameras) to image the Earth, rather than mission-specific instruments. Such equipment limits astronaut photographs to the visible and near-infrared wavelengths in three bands (red, green, blue, and/or near-infrared with appropriate filters and films), similar to what is collected by aerial photograph surveys. The majority of astronaut photographs have been taken from altitudes of 300–400 km (185–250 miles) – the most notable exception being the Apollo missions to the Moon during 1969–1972. Currently, the ISS is the primary manned platform for astronaut photography, which is acquired exclusively with digital cameras.

Today, the process of acquiring high-quality and scientifically useful images of the Earth begins with astronaut training in the Earth sciences. This training is provided for ISS crews by the Crew Earth Observations (CEO) Facility operations team (Stefanov 2008), part of the Earth Science and Remote Sensing (ESRS) Unit at the NASA Johnson Space Center in Houston, Texas. Experts in a variety of disciplines provide ISS crew members with the science background appropriate to a variety of research image foci including glaciers, urban areas, natural hazards, aurora, coral reefs, and deltas (both coastal and inland). This training provides the astronaut with the scientific context of a given image request. Experience with hundreds of crew members over the years indicates that intellectually engaged crew members tend to take higher quality data.

Collection of digital imagery from manned space vehicles such as the ISS is request driven, with specific operational procedures and constraints. Each day ESRS personnel examine the predicted ISS orbital path for the coming 24 h to determine which science targets may be visible to the crew. This list is then filtered by predicted cloud cover and illumination conditions, daily crew schedule, and the ISS ground track relative to the target's latitude/longitude position. The resulting target list is then augmented with additional instructions and data – such as landmark features to aid in locating a target, desired camera focal lens, additional imagery, etc. – and uplinked to the ISS crew. Returned imagery of targets is routinely reviewed to determine whether imagery objectives have been met. Since 2012, NASA ISS remote sensing systems, including the CEO Facility, have participated in collection of data to support the International Charter, Space and Major Disasters (Stefanov and Evans 2015).

Historical Overview of Astronaut Photography

USA Space Program

Observing the Earth from space is one of NASA's longest-standing science experiments. Beginning in the mid-twentieth century, US astronauts have trained for and performed observations of the Earth from orbiting spacecraft, starting with the Mercury missions and continuing today on the International Space Station. The legacy of NASA's astronaut Earth Observations program is evident in the wealth of Earth observation data now available to the global citizenry. Viewing the Earth from space is now routine and fully integrated into the public's daily habits due to widespread Internet access, advances in both data visualization and serving technologies, and ready availability of sophisticated personal computing devices such as laptops, tablets, and smartphones.

Since the beginning of the US human spaceflight programs in the early 1960s, a broad community of scientists has been involved in defining Earth observations objectives and training crew members about Earth processes. Support for Earth observations increased rapidly during the 1960s to support development of new unmanned sensors and was reinforced by the intensive geological training for the Apollo program. Other authors have written extensively about these early programs, e.g., Amsbury (1989) and references therein; (Compton et al. 1989; Wilhelms 1993).

Handheld photography from space started in the Mercury program when payload volume expanded to include cameras and film for photographing the Earth. Structured Crew Earth Observations developed quickly during the Gemini program and resulted in dedicated experiments to develop systematic capabilities. The new views of the Earth from space excited both the public and scientists and were published in several technical reports (Anonymous 1968; Derr et al. 1972; Ewing 1965; Foster 1967; Foster and Smistad 1967; Gill 1967; Lowman 1965, 1969b) and popular venues, e.g., Lowman (1966a). The early Apollo test flights were used to evaluate specific imagery applications (stereo photography, documentation of changes, weather photography) and also functioned as a test bed for demonstrating the capabilities of different films (color infrared and filtered multispectral gray scale) for future multispectral sensors on satellites, such as the United States' Landsat program (Amsbury 1989; Kaltenbach 1970; NASA 1970a). During this time the astronauts worked closely with investigators to obtain the best images for these projects, demonstrating the success of interactive teams of scientists and astronauts.

By the mid-1960s, crews were actively engaged in geology training in key locations in the western USA. Even as the training for future lunar missions accelerated, NASA maintained support for Crew Earth Observations to support ongoing scientific investigations. Crew Earth Observations peaked during the Skylab missions, supporting the highly successful Earth Resources Experiment Program (EREP). EREP involved scores of scientists who selected sites, conducted field studies, trained the crews, and analyzed the orbital data (NASA 1974b, 1977). A fuller account of the early Earth observations training programs and additional references is contained in Amsbury (1989).

The Skylab EREP program was succeeded by the Space Shuttle Earth Observations Program (SSEOP), established in the early 1980s. The early activities of the office supporting this program are described in (Helfert and Wood 1989). Astronaut training in the Earth sciences was a prime function, with the goal of providing critical scientific background to the crews to enable the acquisition of scientifically meaningful astronaut photography. The office staff included scientists from several disciplines in the Earth sciences (meteorology, geology, oceanography, ecology, and geomorphology); they trained the crews, evaluated target requests, and planned and coordinated observation campaigns for each flight. Throughout the 1980s and 1990s, the SSEOP obtained a broad set of medium-resolution imagery using a variety of medium and large format cameras that complemented data from the unmanned Earth Observations satellites of the time. Some flights supported specific data collection campaigns such as the shuttle imaging radar missions (Cimino et al. 1986; Elachi et al. 1982; Evans et al. 1997) and the LIDAR In-space Technology Experiment, or LITE (Winker et al. 1996). Science targets and associated equipment (cameras, lenses, film complement) were selected based on the mission's orbit tracks, projected lighting, season, spacecraft altitude, and current events. Staff scientists provided last-minute updates to the crew a few days before flight, such as overviews of weather systems and noteworthy events (e.g., volcanic activity, biomass burning locations, tropical storms, and floods). During flight, the science team on the ground communicated with the crew daily about upcoming sites, global weather patterns, and additional targets based on current events.

The Shuttle-*Mir* program in the mid-1990s, followed by early missions to the ISS in 2000, allowed NASA's Earth scientists' insight into differences between shorter Space Shuttle flights and longer duration missions on a space station in terms of observational styles and appropriate scientific targets. In contrast to the SSEOP science strategy, based on the spaceflight parameters of short (less than 2 week duration) Space Shuttle missions and built around specifics of each flight (time of launch, season, mission objective, basic spacecraft attitude relative to Earth, orbital parameters including inclination and altitude), a modified scientific approach was adopted for astronauts and cosmonauts flying long-duration missions to the Russian *Mir* Space Station from 1996 to 1998. Building from both the shuttle experience and Russian Earth science approach (Glazovskiy and Dessinov 2000), the Shuttle-*Mir* Earth Observations strategy adjusted for the set orbital inclination, the changing views with seasonal shifts and changing solar illumination (*Mir*, like the current ISS, was not sun synchronous). The Shuttle-*Mir* program segued directly to the current CEO Facility program described above.

Over time, the ISS CEO Facility program has increasingly favored an interdisciplinary earth science approach, integrating other remotely sensed data and new emphasis on coastal processes, natural hazards, human footprints, and environmental change into both crew training and target selection. Astronauts from ISS partner organizations, such as the Canadian Space Agency (CSA), European Space Agency (ESA), and Japan Aerospace Exploration Agency (JAXA), are occasionally trained in Earth observations as part of the CEO Facility program as well. Recent NASA planning for potential missions to Mars, the Moon, or Near-Earth Objects, as well as increasing use of the ISS as a platform for Earth-observing sensors, has revived more intensive geological and remote sensing training for incoming astronaut-candidate classes based in part on the Apollo-era training program, with the expectation of additional benefits to Earth observations (Evans et al. 2011).

USSR/Russian Space Program

The history of the Soviet Earth Observations program during the early years (1960s) mirrored the US Earth Observations program, starting with Yuri Gagarin's flight in 1961. In the beginning much of the cosmonaut photography of the Earth was initiated by cosmonaut interest, rather than a systematic program-level approach. The early cosmonauts, using binoculars and handheld cameras, identified the potential for global, real-time observations of the planet (including dynamic events); general characterization of the Earth's surface from orbit, particularly study of geographically inaccessible regions; and the potential for comparative studies across different regions.

Glazovskiy and Dessinov (2000) detail the evolution of the cosmonaut-based Earth Observations program in the Soviet space program through the 1960s, 1970s, and 1980s. The Earth Observations program structure was established by 1974 to support the Salyut 3–5 missions. The program was designed in collaboration with geographers at the Academy of Sciences, the Chief of Cosmonauts, the Soviet Air Force, and the RSC-Energia engineers and film experts. The cosmonauts were trained in the Earth sciences and conducted experiments with powerful tracking binoculars equipped with cameras. Technologies and procedures were developed, including improvised motion tracking using attitude control mechanisms on the spacecraft. The capability for larger payloads enabled the use of large format cameras (e.g., KATE-140 with an 18 × 18 cm film format and longer lenses) complemented by conventional 35 mm and medium format cameras. Soviet scientists collaborated with scientists and engineers in the German Democratic Republic to develop the program and equip the spacecraft with camera systems.

The USSR Earth observations scientific program ramped up between 1977 and 1985 to support the long-duration missions on the Salyut 6 and 7 Space Stations. The program collaborated with dozens of institutions and included extensive cosmonaut training in state-of-the-art facilities at the Gagarin Training Center outside of Moscow. Earth Observation training flights in Tupolev-134 aircraft employed a strategy that involved close work between cosmonauts and scientists. Cosmonauts on the first Salyut 6 mission were tasked with building and refining the Earth observations procedures and testing the equipment, including fixed high-resolution nadir-looking cameras (KATE-140 and MKF-6M cameras), powerful binoculars, and handheld cameras. The cosmonaut crews were actively involved in calculating manual controls for motion tracking by maintaining spacecraft attitude with thrusters as a means of damping out the relative ground motion. Through this method, they collected high-resolution photography of designated earth targets. During these early space station missions, cosmonauts started observing ephemeral features and events on the Earth's surface, conducted studies on changes in human visual acuity in space, and conducted simultaneous observations with aircraft surveys.

The *Mir* program followed in 1986, continuing through the last mission in 2000. The Earth Observations program on *Mir* followed the strategy developed for the Salyut program, although financial support for scientists and equipment dropped during the collapse of the USSR and the following difficult economic times lasting through the mid-1990s. The Earth-observing assets on the *Mir* included the KATE and MKF-6M fixed cameras, conventional handheld cameras, medium format cameras, and a variety of films including a unique color infrared film. The Priroda module, a dedicated Earth resources facility, was launched to the *Mir* Station in 1996 with an international complement of remote sensing equipment. However, heavy power constraints precluded full operations of many of the Priroda instruments, and the capabilities were not realized.

Cosmonaut crews on the International Space Station are currently performing a variety of Earth observations tasks and experiments, including the "Uragan" program that began in 2001 (I. Sorokin and S.P. Korolev, pers. comm. 2010). This program is similar in many ways to the early NASA Apollo and Skylab EREP programs, in that it uses digital still cameras to optimize design and test of multispectral sensor configurations for monitoring and forecast of land surface processes and hazards from space.

Specifications of Astronaut Photography

Handheld Earth imagery has been acquired by a variety of film and digital cameras over the past 55 years, from both short- and long-duration missions of spacecraft in low earth orbit. This chapter focuses on astronaut photography specifications for NASA missions, spacecraft, and equipment due to the availability of extensive and accessible documentation. A discussion of USSR/Russian handheld photographic equipment up to 2000 is presented in Glazovskiy and Dessinov (2000).

In contrast to purpose-built and mission-specific multispectral and hyperspectral imaging systems for airborne and orbital platforms, the majority of cameras, binoculars, etc. used for handheld astronaut Earth photography have been commercial off-the-shelf equipment that has seen little or no modification prior to launch and on-orbit use. Cameras or imagers meant for use during extravehicular activities (e.g., Apollo and ISS astronaut suit cameras) were modified both for space hardening – reduced outgassing and operation in the vacuum and hard radiation environments – and for the operational challenges posed by space suit design and ergonomics, e.g., operating the camera with pressurized suit gloves in low or zero gravity. A full discussion of lunar orbital and surface astronaut photography, training, and equipment is outside the scope of this chapter, but the interested reader is directed to Jones et al. (2010) and Woods (2009) for more information.

Cameras, Film, and Digital Media

Film cameras have been used in all of the NASA historical and current manned spaceflight programs, including on the retired Space Shuttle and International Space Station, with a variety of film types used for Earth observations (Tables 1, 2, 3, 4, 5, and 6). The Astronaut Office, Orbiter Photo/TV working group, and the SSEOP, in conjunction with the JSC photo lab, Kodak, and camera hardware vendors conducted a variety of tests on various film and hardware combinations, with the end goal of producing the highest resolution images with the greatest dynamic range and the best color reproduction, and minimized image artifacts (e.g., film graininess or image vignetting). The tests were driven by factors that included market and environmental considerations (film production and streamlined film development processes), image resolution results (e.g., using faster ISO 100 films and faster lenses, enabling shorter exposure times), on-orbit operational flexibility (e.g., using auto exposures rather than calculating appropriate f-stops) and, ultimately, new digital technologies.

Original film negatives of handheld Earth observations imagery are archived in a cold storage facility at the NASA Lyndon B. Johnson Space Center (JSC) in Houston, TX. Much of the historical film archive has been digitally scanned into standard formats (such as JPEG and TIFF) and is available online for review and download (see discussion below on online access to data). Many frames of Earth observation data, particularly for the early manned missions, have not been scanned in a systematic fashion however. Scanning of individual film frames of interest can be requested through the JSC Information Resources Directorate.

Both the former Shuttle and ISS programs – and international partners – transitioned to exclusive use of digital 35 mm single lens reflex (SLR) and video cameras in 2004 (Table 7). Coupled with improvements in operational uplink/downlink of data between ground stations and orbiting spacecraft, the use of digital cameras has both increased the volume of imagery and decreased the lag time between acquisition and public availability of data. A variety of interchangeable fixed focal length and zoom lens have also been (and continue to be) used with both the film and digital cameras, producing a wide range of ground resolutions in the handheld imagery dataset (Table 8).

More recently, improvements in camera technology, crew photographic technique, and improved lenses have enabled collection of images with center pixel resolutions approaching 2 meters/pixel. Most handheld astronaut photography is also oblique – has a viewing angle relative to the earth surface less than 90° – to some degree which introduces varying ground resolution per pixel across a given image. For these reasons, it is difficult to assign a standard per pixel ground resolution to the astronaut photography dataset as is common with other orbital sensor datasets.

In addition to increased pixel resolution, dramatic improvements in handheld digital single lens reflex (DSLR) camera capabilities now enable a different class of crew photography: time-lapse imagery. During Expedition 28 in 2011, NASA astronauts (M. Fossum, R. Garan) began experimenting with the automated

Table 1 Specifications of handheld Earth Observations equipment and activities during the NASA Mercury program

Mission (dates flown)	Camera/lenses/ other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Mercury 3 (5 May 1961)	Maurer model 220G 70 mm camera (fixed camera system)	Super Anscochrome	None	None	None	General information photography	More than 150 photographs taken of sky, clouds, and ocean	Lowman (1964, 1965, 1966a, b), Short and Lowman (1973), Swenson et al. (1966), Underwood (1968)
Mercury 4 (21 July 1961)	Mauer model 220G 70 mm camera (fixed camera system)	Super Anscochrome	None	None	None	General information photography	Photography taken over the Atlantic Ocean, North and Central Africa	Grimwood (1963), Lowman (1964, 1965, 1966a, b), Short and Lowman (1973), Swenson et al. (1966), Underwood (1968)
Mercury 6 (20 February 1962)	Ansco Autoset 35 mm camera/ 55 mm lens	Eastman color negative type 5250	None	None	None	General information photography of Earth, meteorological photography planned	48 photographs taken of clouds, oceans, northwest Africa	Cortright (1968), Grimwood (1963), Kuehnel (1972), Lowman (1964, 1965, 1966a, b), Short and Lowman (1973), Swenson et al. (1966), Underwood (1968)

(*continued*)

Table 1 (continued)

Mission (dates flown)	Camera/lenses/ other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Mercury 7 (24 May 1962)	Robot Recorder 35 mm camera	Eastman color negative type 5250	None	None	None	Photograph selected areas for terrain	155 photographs taken of Earth and Earth's limb	Cortright (1968), Grimwood (1963), Kuehnel (1972), Lowman (1964, 1965, 1966a, b), Short and Lowman (1973), Swenson et al. (1966), Underwood (1968)
Mercury 8 (3 October1962)	Hasselblad 500C 70 mm camera (NASA modified)/Zeiss 80 mm Planar f/2.8 lens	Super Anscochrome	None	None	None	Synoptic terrain photography experiment	Photography obtained of poor quality due to general overexposure	Grimwood (1963), Kuehnel (1972), Lowman (1964, 1965, 1966a, b), Short and Lowman (1973), Swenson et al. (1966), Underwood (1968)
Mercury 9 (15–16 May 1963)	Same as Mercury 8	Ultraspeed Anscochrome (fpc-289)	None	None	None	Same as Mercury 8	29 photographs taken of Earth and cloud features, including observations on visibility and color of Earth features	Cortright (1968), Grimwood (1963), Kuehnel (1972), Lowman (1964, 1965, 1966a, b), Short and Lowman (1973), Swenson et al. (1966), Underwood (1968)

Table 2 Specifications of handheld Earth Observations equipment and activities during the NASA Gemini program

Mission (dates flown)	Camera/lenses/ other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Gemini 3 (23 March 1965)	Hasselblad 500C 70 mm camera (NASA modified)/Zeiss 80 mm Planar f/2.8 lens	70 mm, Ektachrome MS SO-217/ Haze filter	None	None	None	General terrain and meteorology information photography	219 Earth-looking photographs taken	Carter and Stone (1974), Giddings (1975), Grimwood et al. (1969), Kuehnel (1972), Lowman (1966a, 1968), Lowman and Tiedemann (1971), NASA (1966, 1967a, b), Short and Lowman (1973), Underwood (1968)
Gemini 4 (3–June 1965)	Hasselblad 500C 70 mm camera (NASA modified)/Zeiss Planar 80 mm f/2.8 lens	70 mm Ektachrome MS SO-217 and 3401	None	None	None	Synoptic terrain photography experiment S005, synoptic weather photography experiment S006	219 Earth-looking photographs taken	Badgley et al. (1969), Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood et al. (1969), Kuehnel (1972), Lowman (1966a, 1968, 1972), Lowman and Tiedemann (1971), Lowman et al. (1967), NASA (1966, 1967a, b, c), Short and Lowman (1973), Stevenson et al. (1968), Underwood (1968), Wobber (1968), Zeitler et al. (1971)
	Zeiss Contarex 35 mm camera/ 250 mm lens	Anscochrome D-200 film/ Haze filter						
	Maurer 16 mm motion picture camera	Ektachrome film						

(*continued*)

Table 2 (continued)

Mission (dates flown)	Camera/lenses/other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Gemini 5 (21–29 August 1965)	Hasselblad 500C 70 mm camera (NASA modified)/Zeiss Planar 80 mm f/2.8 lens	70 mm Ektachrome MS SO-217 and 3401	None	None	None	Same as Gemini 4, with visual acuity experiment S008, surface photography experiment D006	250 Earth-looking photographs taken	Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood et al. (1969), Kuehnel (1972), Lowman (1966a, 1968, 1972), Lowman and Tiedemann (1971), NASA (1966, 1967a, b, c), Pesce (1968), Short and Lowman (1973), Stevenson et al. (1968), Underwood (1968), Wobber (1968), Zeitler et al. (1971)
	Zeiss Contarex 35 mm camera/250 mm lens, 1200 mm Questar lens	Anscochrome D-50 and D-200/Haze filter						
	Maurer 16 mm motion picture camera with spotmeter	Ektachrome film						
Gemini 6-A (15–16 December 1965)	Hasselblad 500C 70 mm camera (NASA modified)/Zeiss Planar 80 mm f/2.8 lens	70 mm Ektachrome MS SO-217 and type 2475/Haze filter	None	None	None	Same as Gemini 4	192 Earth-looking photographs taken	Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood et al. (1969), Kuehnel (1972), Lowman (1966a, 1968, 1972), Lowman and Tiedemann (1971), NASA (1966, 1967a, b, 1968d), Short and Lowman (1973), Underwood (1968), Wobber (1968), Zeitler et al. (1971)
	Maurer 16 mm motion picture camera with spotmeter	Ektachrome film						

Gemini 7 (15–16 December 1965)	Hasselblad 500C (NASA modified)/Zeiss Planar 80 mm f/2.8 lens	70 mm Ektachrome MS SO-217 and Ektachrome Infrared 8443	None	None	None	Experiment S005, S006, S008. Landmark contrast measurement experiment M412	429 Earth-looking photographs taken	Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood et al. (1969), Kuehnel (1972), Lowman (1966a, 1968, 1972), Lowman and Tiedemann (1971), NASA (1966, 1967a, b, 1968d), Pesce (1968), Short and Lowman (1973), Underwood (1968), Stevenson et al. (1968), Wobber (1968), Zeitler et al. (1971)
	Zeiss Contarex 35 mm camera/ 250 mm lens	Panatomic-X 3400 and type 2475/Haze filter						
	Maurer 16 mm motion picture camera with spotmeter	Ektachrome film						
Gemini 8 (16 March 1966)	Hasselblad 500C (NASA modified)/Zeiss Planar 80 mm f/2.8 lens	70 mm Ektachrome MS SO-217/ Haze filter	None	None	None	General terrain and meteorology information photography	19 photographs taken; one of Earth's limb, six Earth-looking oblique photographs	Carter and Stone (1974), Cortright (1968), Grimwood et al. (1969), Kuehnel (1972), NASA (1967a, 1968d), Short and Lowman (1973), Underwood (1968)
	Maurer 16 mm motion picture camera with spotmeter	Ektachrome film for Maurer cameras						
	Maurer 16 mm motion picture camera							

(continued)

Table 2 (continued)

Mission (dates flown)	Camera/lenses/other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Gemini 9-A (6 June 1966)	Hasselblad 500C (NASA modified)/Zeiss Planar 80 mm f/2.8 lens	70 mm Ektachrome MS SO-217/ Haze filter	None	None	None	Same as Gemini 8	362 Earth-looking photographs taken	Cortright (1968), Giddings (1975), Grimwood et al. (1969), Kuehnel (1972), Lowman (1968), Lowman and Tiedemann (1971), NASA (1967a, 1968d), Short and Lowman (1973), Stevenson et al. (1968), Underwood (1968), Zeitler et al. (1971)
	Hasselblad Super Wide Angle-C (NASA modified)/Zeiss Biogon 38 mm, f/4.5 lens							
	Maurer 70 mm space camera/ Xenotar 80 mm f/2.8 lens							
	Maurer 16 mm motion picture camera with spotmeter	Ektachrome film						

Gemini 10 (18–21 July 1966)	Hasselblad Super Wide Angle/Zeiss Biogon 38 mm, f/4.5 lens	Same as Gemini 9-A	None	None	None	Experiments S005, S006, M412. Color patch photography experiment M410	371 Earth-looking photographs taken	Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood et al. (1969), Kuehnel (1972), Lowman (1968), Lowman and Tiedemann (1971), NASA (1967a, b, 1968d), Short and Lowman (1973), Stevenson et al. (1968), Underwood (1968), Zeitler et al. (1971)
	Maurer 70 mm space camera/ Xenotar 80 mm f/2.8 lens							
	Maurer 16 mm motion picture camera with spotmeter							
Gemini 11 (12–15 September 1966)	Hasselblad Super Wide Angle/Zeiss Biogon 38 mm, f/4.5 lens	70 mm Ektachrome MS SO-368 film/Haze filter	None	None	None	Experiments S005, S006	238 Earth-looking photographs taken	Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood et al. (1969), Kuehnel (1972), Lowman (1968), Lowman and Tiedemann (1971), NASA (1967a, 1968d), Short and Lowman (1973), Stevenson et al. (1968), Underwood (1968), Zeitler et al. (1971)
	Maurer 70 mm space camera/ Xenotar 80 mm f/2.8 lens							
	Maurer 16 mm motion picture camera with spotmeter	Ektachrome film						

(*continued*)

Table 2 (continued)

Mission (dates flown)	Camera/lenses/ other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Gemini 12 (11–15 November 1966)	Same as Gemini 11	Same as Gemini 11	None	None	None	Experiments S005, S006 and ocean features photography	415 Earth-looking photographs taken	Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood et al. (1969), Kuehnel (1972), Lowman (1968), Lowman and Tiedemann (1971), NASA (1967a, b, 1968d), Short and Lowman (1973), Underwood (1968), Stevenson et al. (1968), Wobber (1968), Zeitler et al. (1971)

Table 3 Specifications of handheld Earth Observations equipment and activities during the Apollo program

Mission (dates flown)	Camera/lenses/other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Apollo 4 (9 November 1967)	Maurer Model 220G 70 mm sequence camera/Ektar 76 mm f/2.8 lens (exposure setting 1/500 of a second)	70 mm Ektachrome MS So-368	None	None	None	Acquisition and return of highest altitude photography (~18,520 km) taken of Earth	712 photographs of Earth (2136 Earth-looking photographs taken with the 70 mm camera during the Apollo 4, 6, 7, 8, 9, 10, 11, and 12 missions)	Dornbach (1968), Underwood (1968)
Apollo 6 (4 April 1968)	Maurer Model 220G 70 mm sequence camera/Ektar 76 mm f/2.8 lens	Same as Apollo 4	None	None	None	To obtain sequenced vertical photographs of Earth in stereo	Vertical or near-vertical photographs of Earth between ~183 and 306 km altitude. 372 Earth-looking photographs taken	Amsbury (1969), Badgley et al. (1969), Dornbach (1968), Kaltenbac (1969a), Lowman (1972), NASA (1968a, b, 1970b), Short and Lowman (1973)
Apollo 7 (11–22 October 1968)	Hasselblad Model 500C (NASA modified)/Zeiss Planar 80 mm f/2.8 lens Maurer 16 mm sequence camera/ 75 mm lens	70 mm Ektachrome MS SO-368, SO-121, and B&W type 3400/ Wratten 2A, 25A (red) and 58 (green) filters used with SO-121 film	Leitz monocular telescope (10X magnification, objective diameter 40 mm)	TV equipment, hand held or mounted. (Black & White)	Data tape recorder. Speed 9.5, 38, 305 cm/ s; Tape 686 m	Synoptic terrain photography experiment S005, synoptic weather photography experiment S006	Photographs taken of Earth between ~159 and 367 km altitude. 502 Earth-looking photographs taken. Hardware training for general photography and experiments given for each Apollo mission	Amsbury (1969), Carter and Stone (1974), Kaltenbach (1969b), Kuehnel (1972), NASA (1968c, 1969a, 1970b, 1973a), Short and Lowman (1973)

(*continued*)

Table 3 (continued)

Mission (dates flown)	Camera/lenses/other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Apollo 9 (3–13 March 1969)	Hasselblad Model 500EL (NASA modified)/Zeiss Planar 80 mm f/2.8 lens	70 mm Ektachrome MS SO-368/ Haze filter	Same as Apollo 7	Same as Apollo 7 (first color TV flown on Apollo 10)	Same as Apollo 7	Multispectral photography experiment SO65. Terrain, weather, air glow, and Earth's limb sites of opportunity were planned	127 vertical or near-vertical photographs taken of Earth in each of the four bands of the SO65 experiment. 1157 Earth-looking photographs taken with the handheld Hasselblad 70 mm camera. (A number of excellent 70 mm handheld color photographs were taken of the Earth during Earth orbit and enroute to and from the Moon on the Apollo 8, 10, 11, 12, 13, 14, 15, 16, and 17 missions)	Amsbury (1969), Badgley et al. (1969), Bannert (1972), Carter and Stone (1974), Kaltenbach (1970), Kuehnel (1972), Lowman (1969a, 1972), NASA (1969b, c, 1970b, 1973a), Nicks (1970), Schowengerdt and Slater (1972), Short and Lowman (1973)
	SO65 experiment:							
	Hasselblad Model 500EL/Zeiss Planar 80 mm f/2.8 lens/ Photar 15 filter	70 mm Ektachrome IR, SO-180						
	Hasselblad Model 500EL/Zeiss Planar 80 mm f/2.8 lens/ Photar 58B filter	70 mm B&W Panatomic-X, type 3400						
	Hasselblad Model 500EL/Zeiss Planar 80 mm f/2.8 lens/ Photar 89B filter	70 mm B&W Infrared SO-246						
	Hasselblad Model 500EL/Zeiss Planar 80 mm f/2.8 lens/ Photar 25A filter	70 mm B&W Panatomic-X, type 3400						
	Maurer 16 mm data acquisition camera (DAC)/75 mm lens	Ektachrome SO-368 for Maurer DAC						

Table 4 Specifications of handheld Earth Observations equipment and activities during the Skylab program

Mission (dates flown)	Camera/ lenses/other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Skylab 2 (25 May–22 June 1973)	Multispectral photographic facility (MPF, s190): Multispectral photographic camera (S190A). Six Itek 70 mm boresighted cameras/high precision f/2.8, 21.2° FOV lenses	MPF: 70 mm film: Station 1- IR B & W (0.7–0.8 μm) Station 2- IR B & W (0.8–0.9 μm) Station 3- Aerochrome IR color type EK2243 (0.5–0.88 μm) Station 4- color SO-356 (0.4–0.7 μm) Station 5- Panatomic-X B & W SO-022 (0.6–0.7 μm) Station 6- Panatomic-X B & W SO-022 (0.5–0.6 μm)	Leitz Trinovid 10 × 40 binoculars (space modified)	Westinghouse color 25–150 mm zoom. Focus range 20″ to infinity	Same as Apollo 7 video tape recorder, recording time 30 min	Obtain data with the Earth resources experiment package (EREP) sensors. Obtain general information photography of Earth	35,000 frames cycled through S190A during Skylab 2, 3, and 4 (Not all of these were exposed or Earth-looking). 32,994 Earth-looking photographs taken with S190A and 5569 Earth-looking photographs taken with S190B during Skylab 2, 3, and 4 302 Earth-looking photographs taken with the handheld Hasselblad 70 mm camera, and 355 Earth-looking photographs taken with the handheld Nikon 35 mm cameras A number of Earth-looking TV passes were performed Note: The S190A, S190B, and S191 are instruments in the Earth resources experiment package (EREP) which obtained information of the Earth in numerous spectral regions during the three manned Skylab flights	Kaltenbach et al. (1974), Kenney (1974, 1975), Lockwood and Sauer (1975), Lowman et al. (1973), Mcniel and Devalcourt (1974a), NASA (1973a, b, 1974b), Stevenson (1974), Underwood and Holland (1973a)
	Actron Earth Terrain Camera (AETC, S190B)/5 in. format, f/4 lens, 18 in. focal length	AETC: color SO-242 (0.4–0.7 μm)						

(*continued*)

Table 4 (continued)

Mission (dates flown)	Camera/ lenses/other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
	16 mm DAC boresighted on Visual Tracking System Infrared Spectrometer (S191) field of view	B & W EK3414 (0.5–0.7 μm) Ektachrome SO368						
	Hasselblad Model 500EL (electric data camera system with reseau plate), 70 mm camera/ 100 mm lens	70 mm Ektachrome SO-368						
	Nikon 35 mm cameras (5)/ 55 mm and 300 mm lenses	Ektachrome SO-368, SO-168, 2485, and 2443 type						
	Maurer 16 mm DAC/5, 10, 18, 25, 75, and 100 mm lenses	Ektachrome SO-368 and SO-168						

Skylab 3 (28 July–25 September 1973)	S190A, S190B, Hasselblad, Nikon, and DAC's same as Skylab 2	S190A, Hasselblad, Nikon, and DAC's film & filters same as Skylab 2	Same as Skylab 2	Same as Skylab 2	Same as Skylab 2	Obtain data with the Earth resources Experiment Package (EREP) sensors Obtain general information photography of Earth	642 Earth-looking photographs taken with the handheld Hasselblad 70 mm camera, and 537 Earth-looking photographs taken with the handheld Nikon 35 mm cameras A number of Earth-looking TV passes were performed	Kaltenbach et al. (1974), Kenney (1974, 1975), Lockwood and Sauer (1975), Lowman et al. (1973), Mcniel and Devalcourt (1974b), NASA (1973a, b, 1974b), Stevenson (1974), Underwood and Holland (1973b)
	Earth Terrain Camera (S190B)	SO-242 (0.4–0.7 μm) EK3414 (0.5–0.7 μm) Infrared color EK3443 (0.5–0.88 μm)						
Skylab 4 (16 November 1973–8 February 1974)	S190A, S190B, Hasselblad, Nikon, and DAC's same as Skylab 2 & 3	S190A, Hasselblad, Nikon, and DAC's film & filters same as Skylab 2 & 3	Leitz Trinovid 10 × 40 binoculars (space modified)	Same as Skylab 2	Same as Skylab 2	Obtain data with the Earth resources experiment package (EREP) sensors. Obtain general information photography of Earth	1260 Earth-looking photographs taken with the handheld Hasselblad 70 mm camera, and 1255 Earth-looking photographs taken with the handheld Nikon 35 mm cameras A number of Earth-looking TV passes were performed	Kaltenbach et al. (1974), Kenney et al. (1974, 1975), Lockwood and Sauer (1975), Lowman et al. (1973), NASA (1973a, b, 1974b), Stevenson (1974), Underwood and Holland (1974)
	Earth Terrain Camera (S190B)	SO-242 (0.4–0.7 μm) EK3414 (0.5–0.7 μm) Infrared color (high resolution) EK3443 (0.5–0.88 μm)						

Table 5 Specifications of handheld Earth Observations equipment and activities during the Apollo-Soyuz project

Mission (dates flown)	Camera/lenses/ other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Apollo-Soyuz Test Project (15 to 24-Jul-1975)	Two 70 mm Hasselblad Model 500EL cameras: One reflex camera with 50 mm and 250 mm lenses	SO-242/ Wratten 2A; SO-368 (QX807)/film coated with equivalent of Wratten 2A	Leitz Trinovid 10 × 40 binoculars (space modified)	ASTP television system (camera, lens, monitor) Operated in Apollo and Soyuz. Zoom 6-1 and 3-1 Range 25–150 mm, 9–27 mm. F-stop 4.4–44, 3.5–35 Focus 51 cm -inf., 30.5 cm -inf. (modified Westinghouse)	Modified Sony Cassette-Corder Model TC-SS for voice recorder and playback Video tape recorder recording time 30 min. Data tape recorder same as Skylab	Earth Observations and photography experiment MA-136	1916 Earth-looking photographs taken; 751 good exposures obtained A number of Earth-looking TV passes were performed	NASA (1974a, 1976, 1977)
	One electric Data Camera System with reseau plate with 60 mm and 100 mm lenses	Infrared color, type 2443 equivalent/ Wratten 12						
	One 35 mm Nikon camera/35 mm lens (interior photography), 300 mm lens (exterior photography)	Ektachrome						
	16 mm data acquisition camera/10 mm, 25 mm, and 75 mm lenses	Ektachrome						
	Spotmeter							

Table 6 Specifications of handheld Earth Observations film cameras and films used during the Space Shuttle and ISS programs

Cameras	Lenses	Film types
Hasselblad 500 EL/M, 70 mm, NASA modified	Zeiss 40 mm Zeiss 50 mm Zeiss 80 mm Zeiss 100 mm Zeiss 110 mm Zeiss 250 mm Zeiss 350 mm	Fuji, natural color positive, Velvia 50, CS-135-36, ASA 32, standard base Kodak B & W, Plus-X Aerographic Color negative, Vericolor III, 70 mm unperforated, process C-41 Natural color positive, Ektachrome Professional 5017, ASA 64, standard base Natural color positive, Ektachrome 5036, 200 Professional, ASA 200, standard base Natural color positive, Ektachrome Professional SO-117, ASA 400, standard base Natural color positive, Lumiere 100/5046, ASA 100, standard base Natural color positive, Lumiere 100x/5048, ASA 100x, standard base Natural color positive, Ektachrome MS, ASA64, thin base, fine grain Color positive, Ektachrome X Professional, ASA 64, standard base Color positive, Ektachrome SO-368, fine grained, with yellow dye layer equivalent to Wratten 2A Color positive, Ektachrome, high speed, ASA 400 Color positive, Ektachrome 64, 220 Roll format Color positive, Ektachrome 64 Color positive, Aerochrome II color reversal, ISO-A 32, process EA-5, standard base Color positive, Elite 100S, E6 reversal Color infrared, Aerochrome 2443, ASA 160, standard base Russian color infrared, 2 dye layer, estimated ASA 64
Linhof Aero Technika, 100 × 120 mm, NASA modified	Linhof 90 mm Linhof 250 mm	Kodak Natural color positive, Ektachrome QX 868, ASA 64, 5017 emulsion, thin base Natural color positive, Lumiere 100/5046, ASA 100, standard base Natural color positive, Lumiere 100x/5048, ASA 100x, standard base Color positive, Elite 100S, E6 reversal Color positive, Ektachrome X Professional, ASA 64, standard base Color positive, Ektachrome 64, 220 roll format Color positive, Aerochrome II color reversal, ISO-A 32, standard base Color positive, Aerochrome II Duplicating

(*continued*)

Table 6 (continued)

Cameras	Lenses	Film types
		Film, 70 mm, process EA-5 Color infrared, Aerochrome 2443, EA-5 process through June 1999, E-6 process afterwards, thin base
Rolleiflex, 70 mm, NASA modified	Zeiss 50 mm Zeiss 80 mm Zeiss 100 mm Zeiss 250 mm	Kodak, natural color positive, Ektachrome Professional 5017, ASA 64, standard base
Nikon F3 35 mm, NASA modified Nikon F4 35 mm, NASA modified	All lenses are interchangeable and autofocus: 16 mm 20 mm 28 mm 35 mm 35–70 mm zoom 55 mm 60 mm 70 mm 85 mm 180 mm 300 mm 400 mm 2X doubler	Fuji Natural color negative, NHG, ASA 400, standard base Color negative, 35 mm, ASA 800 Natural color positive, Velvia 50, CS-135-36, ASA 32, standard base Kodak B &W positive, Technical Pan Film 2415 Estar AH Base, ASA 100 Natural color negative, Ektar 100-3101, ASA 125, standard base Natural color negative, Ektapress 5030, ASA 1600, standard base Natural color negative, Vericolor III 5026, ASA 160, standard base Color negative, Kodacolor VRG/100, ASA 100, standard base Color negative, Ektar 25 Professional Film, ASA 25 Color negative, Pro 400, 35 mm, ASA 400 Color negative, Pro PMZ 1000, 35 mm or 120 mm, process C-41 Natural color positive, Ektachrome Professional 5017, ASA 64, standard base Natural color positive, Ektachrome Professional 5074, ASA 400, standard base Natural color positive, Ektar 25 Professional, ASA 25, standard base Color positive, Vericolor 400 Prof (VPH), ASA 400, standard base Color positive, Ektachrome X Professional, ASA 64, standard base Color positive, Ektachrome 64 T Professional Film, ASA 64 Color positive, Elite 100S, E6 reversal Color positive, EXR 500 Portra Color negative, 160NC, 35 mm, ASA 160 Color negative, 400NC, 35 mm, ASA 400 Color positive, 400VC, 35 mm, ASA 400

Table 7 Digital SLR cameras used in the Space Shuttle and ISS programs. Nikon D4 cameras are currently in use on board the ISS, with transition to Nikon D5 cameras expected over the next year

Manufacturer/model	Original Image size (mm)	Original Image size (pixels)
Sony HDW-700 high-definition television camcorder	–	1920 × 1035 interlaced
Kodak DCS460, RGBG array	27.6 × 18.5	3060 × 2036
Kodak DCS660, RGBG array	27.6 × 18.5	3060 × 2036
Kodak DCS760, RGBG array	27.6 × 18.5	3060 × 2036
Nikon D1, RGBG imager color filter	23.6 × 15.5	2000 × 1312
Nikon D2Xs, RGBG imager color filter	23.7 × 15.7	4288 × 2848
Nikon D3	36.0 × 23.9	4256 × 2832
Nikon D3X	35.9 × 24.0	6048 × 4032
Nikon D3S	36.0 × 23.9	4256 × 2832
Nikon D4	36.0 × 23.9	4928 × 3280
Nikon D800E	35.9 × 24.0	7360 × 4912

Table 8 Calculated ground resolutions for representative film and digital still cameras and lenses used on the International Space Station. Note: Altitudes are given as km above sea level (asl), and calculated resolutions do not include effective ground motion blurring caused by the high orbital velocity of the ISS (~27,500 km/h) relative to the Earth's rotation

Camera	Lens (mm)	Ground resolution in m/pixel at image center	
		Minimum altitude = 368 km asl	Maximum altitude = 386 km asl
Hasselblad 70 mm	110	35.4	37.1
	250	15.6	16.3
	350	11.1	11.6
Nikon 35 mm	300	13.0	13.6
	400	9.7	10.2
Kodak DSC	300	11.0	11.6
	400	8.3	8.7
	800	4.2	4.4

functions of the onboard DSLR cameras. The astronauts utilized a Bogen arm in the Cupola of the ISS to stabilize the camera and set the camera to take an image every 3 s for several minutes. The motion of the ISS in orbit captured by the sequence of still images was then assembled into dramatic time-lapse sequence videos, providing spectacular new views of the planet never seen before by the general public.

The downlinked still images were processed by the ESRS Unit, stitched together using video production software, and published on the Gateway to Astronaut

Photography of Earth (2016) website for the public. As educational supplements to these videos, the ESRS Unit has also created:

- Annotated time-lapse videos, highlighting city and place names
- Time-lapse video alongside a Google Earth tour, which plays simultaneously, enabling the user to see both geographical and geological feature names within the video
- Narrated time-lapse videos within which features in the video are described for the viewer

Since 2011, subsequent crew members on ISS have continued the acquisition of time-lapse photography and further developed the technique. Astronaut D. Pettit captured time-lapse imagery of star trails that had not been seen from the ISS before, and astronaut D. Burbank acquired impressive and rare time-lapse photography of comet Lovejoy (Fig. 1). The CEO Facility operations team has also tasked the crew with taking time-lapse photography for scientific and educational purposes, to include imaging a tropical cyclone moving toward land or taking a sequence of a strong aurora event (Fig. 2). The public and media response to this new class of imagery has been dramatic. The time-lapse sequence videos have been highlighted by numerous publications (e.g., Chicago Tribune, USA Today, etc.), websites (e.g., SpaceflightNow.com, Space.com, NASA.gov, YouTube, etc.), and TV broadcasts on most major networks, such as the Discovery Channel and the Public Broadcasting System.

Fig. 1 Astronaut photograph ISS030-E-15485 of comet Lovejoy as seen from the ISS. The image was collected as part of a time-lapse sequence on 22 December 2011

Fig. 2 Astronaut photograph ISS030-E-84660 of the aurora borealis as seen from the ISS. The image was collected as part of a time-lapse sequence on 4 February 2012. The ISS Japanese Experiment Module – Exposed Facility is visible at *image top*

Processing, Archiving, and Accessing Astronaut Photography

Digital camera imagery taken by astronauts is stored onboard the ISS and periodically transmitted to the ground using NASA or Russian downlink systems (both geosynchronous satellite network and line of site). Handheld digital camera imagery taken by NASA, JAXA, CSA, and ESA astronauts are initially processed by JSC archivists to determine the image subject category, e.g., Earth observation, with each image assigned a unique NASA identification number.

Imagery taken by cosmonauts is typically entered into the NASA system after receipt from Roscosmos. Once received on the ground and processed, the imagery is delivered to the ESRS Unit – or historically, to a precursor group – for cataloging and entry into the online Gateway to Astronaut Photography of Earth (2016) database. If the imagery was downlinked in a raw camera format, it is converted into a full-resolution JPEG format for public access through the database in order to reduce server and network loads; the raw camera files (if available) also can be requested through the Gateway to Astronaut Photography of Earth (2016) website.

The image cataloging process begins with determination of the image center point coordinates. In contrast to data collected by current automated orbital sensors, there is no geolocation data embedded in astronaut photographs – recall that off-the-shelf cameras are used. The time of image acquisition, if accurate, can be used with the known orbital position of the spacecraft (e.g., the ISS) to determine a nadir-viewing ground coordinate useful as an initial estimate of the image's center point coordinates. This estimate typically needs refinement, however, due to the unconstrained nature of handheld astronaut photography – all viewing angles, from essentially nadir to highly oblique, are available within the constraints of the spacecraft window, and the image center point may be some distance from the spacecraft nadir point. This lack of standard viewing angles requires manual determination of the image center point geographic coordinates by analysts using other georeferenced data – such as Landsat scenes or cartographic maps – with a typical location error of 0.1° in latitude and longitude.

Once the center point is determined, descriptive metadata is generated for the image that includes major visible features or landmarks, estimated cloud cover, and the calculated viewing angle/direction relative to the orbital position of the spacecraft. Together with the camera file metadata for each frame (digital camera data only), the image is added to the Gateway to Astronaut Photography of Earth (2016) database. While all received ISS and Space Shuttle Earth observations imagery is currently entered into the database, it is prioritized for purposes of cataloging.

While the CEO Facility maintains an impressive dataset now comprising over two million images, use of the data for scientific research, disaster response, and visualizations is minimal in comparison to other data collected from free-flying satellite platforms such as Landsat, Worldview, etc. The lack of full geolocation information makes it difficult to integrate astronaut photographs with other georeferenced data to facilitate quantitative analysis such as land cover/land use classification, change detection, or geologic mapping. The manual determination of image center points is both time- and labor-intensive, leading to delays in releasing geolocated and cataloged data to the public, and in particular the timely use of data for disaster response and humanitarian aid to stricken areas.

In order to address this inherent disadvantage of handheld astronaut photography, the GeoCam Space project was funded by NASA to develop an on-orbit hardware and ground-based software system for increasing the efficiency of geolocating astronaut photographs from the ISS. The hardware component consists of modified smartphone elements including cameras, central processing unit, wireless, and inertial measurement unit/accelerometers/magnetometers reconfigured into a compact unit that attaches to the base of the Nikon D4 camera and connects using the 10-pin connector or USB port. This provides a secondary, left or right facing camera perpendicular to the primary camera pointing direction. The secondary camera observes calibration targets with known internal X, Y, Z position affixed to the interior of the ISS to determine the camera pose corresponding to each image frame. This information is recorded by the GeoCam Space unit and indexed for correlation to the camera time recorded for each image frame.

Data – image, EXIF header, and camera pose information – is transmitted to the ground software system (called GeoRef) using the established Ku-band USOS downlink system. Once integrated on the ground, the camera pose information provides an initial geolocation estimate for the individual film frame. For nadir-viewing images this does not vary greatly from the ISS nadir position, but for oblique imagery this represents a significant advance in geolocation from the existing manual feature-matching approach. With the initial geolocation estimate, full georeferencing of an image is completed using a rapid tie-pointing interface, and

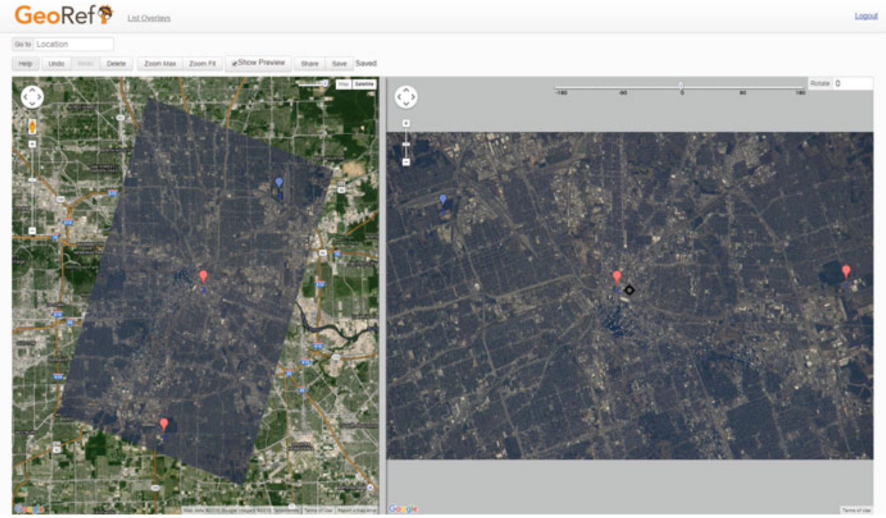


Fig. 3 Screen capture of the GeoRef tie-pointing interface, with astronaut photograph ISS038-E-30866 as an example georeferenced image

the resulting data is added to the publically accessible Gateway to Astronaut Photography of Earth (2016) online database in both Geotiff and Keyhole Markup Language (kml) formats (Fig. 3). Implementation of the GeoCam Space system has increased the efficient delivery of useful data to the public and is expected to encourage greater use of astronaut photography in applications and research requiring fully geolocated imagery.

The entire digital collection of astronaut photography (more than two million images) is accessible using both map- and metadata-based search tools from the Gateway to Astronaut Photography of Earth (2016), as well as the International Space Station instrument integration interface (2016) search tool (that also enables searching of other ISS remotely sensed datasets; Vanderbloemen et al. 2014). Handheld astronaut imagery can be downloaded at various resolutions free of charge from the Gateway to Astronaut Photography of Earth (2016) website. Keyhole Markup Language (kml) files for cataloged data with image center points also can be generated on-the-fly for direct input into geospatial browsers such as Google Earth. Digital images that have not been cataloged may also be queried using spacecraft nadir ground location coordinates, times of data acquisition, and lens focal length. Selected astronaut photographs and descriptive content are also available as part of the NASA content layer in the free geospatial browser Google Earth.

Platforms

Manned spacecrafts launched by the USA, Russia/former USSR, and other countries have uniformly employed asynchronous inclined elliptical orbits around the Earth.

This is due to the relative ease of establishing and maintaining this type of orbit (e.g., less fuel needed to attain orbit and perform attitude adjustments), minimization of loss-of-signal periods for radio communication, and accessibility to landing sites on both land and at sea (Green and Lopez 2009). While this type of orbit limits the degree of nadir-viewing land and sea surface that can be observed to the degree of inclination, it also provides for a wider range of illumination and viewing conditions than typically available for sensors on sun-synchronous, polar-orbiting platforms. NASA retired the remaining Space Shuttles from service in 2011, leaving the International Space Station as the only platform for regular handheld Earth observations for at least the coming decade. While China has announced plans to complete construction of a crewed space station in low earth orbit by 2022, any plans to include collection of handheld Earth imagery as a formal component of science activities are unknown.

International Space Station (ISS)

Originally conceived as the US Space Station Freedom, the ISS orbital inclination was initially chosen at 28° to facilitate launching from NASA Kennedy Space Center in Florida. With expansion of the program to include international partners the inclination was increased to 51.6° to accommodate the Russian launch site at Baikonur Cosmodrome (Eppler and Runco 2001). This inclination allows the ISS to overfly the temperate and tropical regions of the Earth – covering approximately 75 % of the Earth's land area and approximately 95 % of the Earth's population (Fig. 4). The ISS orbit varies in altitude from approximately 350 to 455 km asl; due to atmospheric drag, reboosting of the ISS to maximum altitude is required approximately every 90 days. Due to the westward precession of orbit tracks, the ISS has an approximate repeat time over the same location every 3 days, with similar lighting conditions being repeated every 3 months not correcting for seasonal lighting shifts. For purposes of Earth observation, the CEO Facility team generally limits daily target selections to regions with at least a 20° sun angle (elevation above local horizon) in order to have adequate illumination of ground targets. This constraint, combined with ISS orbit precession and seasonal precession of the Earth's orbit around the Sun, also produces intervals when only Northern or Southern Hemisphere ground targets meet the illumination criteria (Fig. 4). This constraint does not hold for atmospheric or night time ground targets.

The (ISS) Destiny Laboratory Window

The US Destiny Laboratory Module of the ISS has a window port built into its nadir-facing side (NASA 2015). The window consists of three panes of Corning 7940 fused silica which are approximately 56 cm in diameter, providing an approximately 51 cm clear aperture. The ISS program agreed to upgrade the glass in the Destiny window to a set of stringent optical performance requirements in 1996. The Destiny window has a wavefront error of $\lambda/15$ peak-to-valley over a 15.2 cm aperture relative to a reference wavelength of 0.6328 μm, which allows the use of up to a 30 cm telescope with no degradation of wavefront due to the glass. These properties give

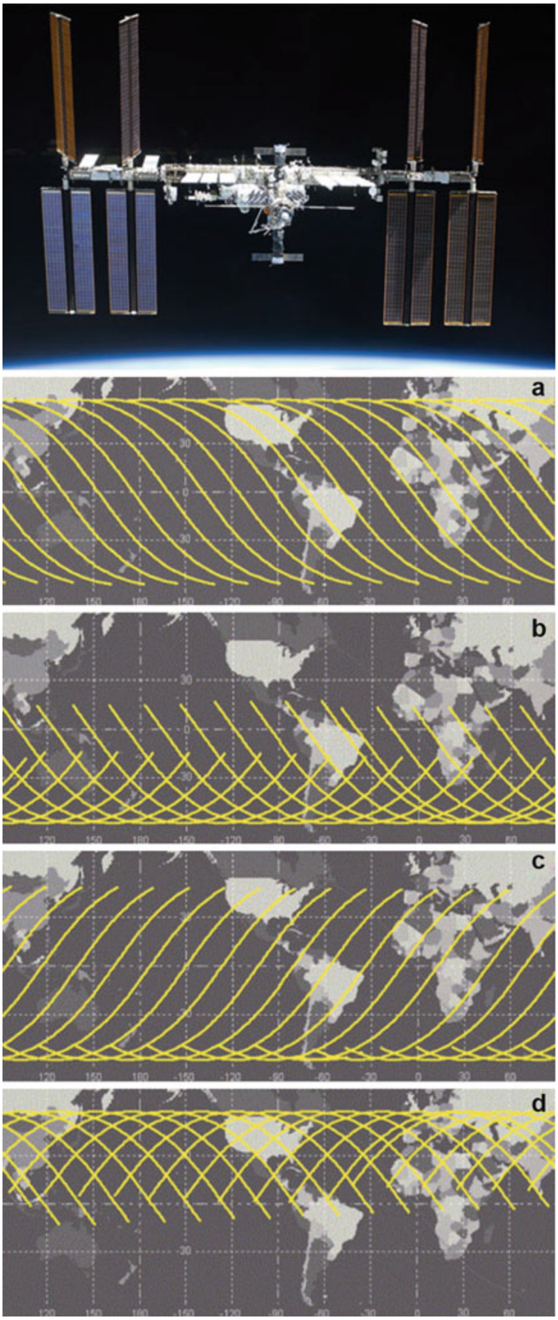


Fig. 4 International Space Station as viewed from Space Shuttle *Atlantis* during the STS-132 mission (image S132-E-13221, taken 23 May 2010). Westward orbit ground track precession for the ISS is depicted in frames a-d. (**a**) Successive descending orbit tracks, daylight illumination in both N and S hemispheres; (**b**) Daylight illumination in S hemisphere only; (**c**) Ascending orbit tracks, daylight illumination in both N and S hemispheres; (**d**) Daylight illumination in N hemisphere only

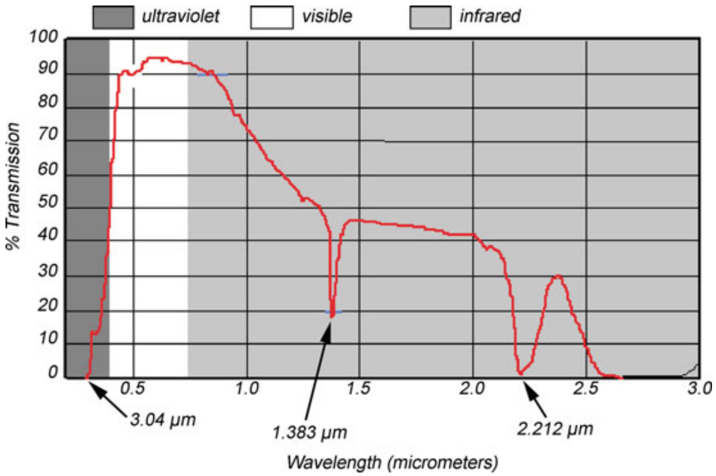


Fig. 5 Transmission curve for the center of the as-flown Destiny Laboratory window. Cutoff below 0.304 μm is caused by a reflective coating

the Destiny window the highest performance of any window ever flown on a crewed vehicle (Eppler and Runco 2001).

With the support of the ISS program, the flight article window was radiometrically calibrated prior to installation in the Destiny Laboratory in May of 2000. This calibration indicated that the window had better than 95 % transmittance in the visible region, with a steep drop-off in the ultraviolet (due to a reflective coating) and a gradual drop-off into the infrared wavelengths (Fig. 5). Window transmittance decreases to 50 % or less at approximately 1.3 μm.

In addition to the Destiny window, there are a number of other viewing ports on the ISS that are frequently used for Earth observations. Windows in the Russian service module (or Zvezda), albeit of lesser optical quality, provide both nadir and oblique viewing opportunities of the Earth. A moveable viewing module, the Cupola (Fig. 6), was transported to the ISS in 2010 during the STS-130 mission of Space Shuttle *Endeavour* and attached to the Tranquility module. The Cupola includes seven optical-quality fused silica and borosilicate glass windows that provide 360° viewing capability and was designed to support vehicle docking, remote manipulator arm operations, and Earth and space observations (European Space Agency 2011).

Window Observational Research Facility (WORF)

Complete utilization of the optical performance of the Destiny window would be impossible without a facility to allow stable positioning of research payloads in the window. Design and fabrication of the Window Observational Research Facility, or WORF, began in 1998. The WORF facility was transported to the ISS aboard the Space Shuttle *Discovery* in 2010 during the STS-131 mission.

Fig. 6 Astronaut photograph ISS022-E-66972 of the Sahara Desert as seen through the Cupola on board the ISS. Image acquired 17 February 2010

The WORF is based on an ISS-standard express rack (NASA 2015) to capitalize on the express philosophy in accommodating subrack payloads. The WORF is essentially an express rack with an approximate cubic meter-sized space in the middle (Fig. 7), centered on the Destiny window. This space – the payload volume – provides mounting surfaces for window payload hardware, including mounting on a stiff lower payload shelf that is designed to minimize transmission of ISS vehicle vibrations into the optical components of the payload. The interior of the WORF is sealed by means of an aisle-side hatch. The interior of the payload volume is painted flat black to minimize stray light and allow investigations of faint upper atmosphere phenomena such as aurora and noctilucent clouds.

The WORF can provide power, data, and cooling water for up to three payloads simultaneously by interfacing with existing ISS systems. At present, the WORF can provide an average downlink data rate on the order of 2 Mbps, although this may be improved with proposed communications infrastructure improvements to ISS. Investigators can operate their payloads autonomously at their institution, with up- and downlink data going through the Huntsville Operations Support Center at Marshall Space Flight Center in Huntsville, AL. The general design philosophy of WORF favors autonomous payloads, but crew members can operate payloads from the Destiny Laboratory aisle using an externally mounted laptop computer.

It is an axiom in the US manned space program that a trained crew member is one of the best analytical tools Earth observations can employ. The WORF is designed to accommodate crew stabilization devices and brackets to allow vibration-free operation of still cameras and video recorders. In addition, the aisle side hatch allows crew members to interface with a piece of equipment dubbed the "kayak shroud," which allows access to the interior of the WORF without glare from the Destiny Laboratory aisle interfering with Earth photography.

Fig. 7 The Window Observational Research Facility (WORF) installed over the US Destiny Laboratory module window on board the ISS. The Destiny window is visible through the WORF payload volume at *image center*. Image S131-E-8619 was taken 10 April 2010

Remote Sensing Using Astronaut Photography

Advantages and Limitations

The most significant distinction of astronaut photography as a research dataset is the most obvious one – these images are framed and acquired by a human being rather than an automated sensor system. The astronaut can make on-the-fly decisions about image targets (e.g., pointing), resolution of data collected (through selection of cameras lenses), and whether conditions are favorable for taking imagery (e.g., acceptable illumination and cloud cover) that are beyond the capabilities of automated sensor systems. A full orbit of the earth takes approximately 90 min, during which time the ISS crosses both illuminated and dark portions of the globe. The astronaut photograph dataset contains great variability in illumination conditions, look angle, spatial resolution (typically 4–40 meters/pixel), and repeat imagery of a given location on the earth's surface.

The majority of astronaut photographs of the Earth have been acquired using film and digital cameras sensitive to the visible blue, green, and red wavelengths of the electromagnetic spectrum (0.38–0.72 μm), with specific experiments conducted using filters and film sensitive to narrow bandpasses within this range and in the near-infrared wavelengths (~0.72–1.30 μm; see chapter ▶ "Introduction and History of Space Remote Sensing" by Madry, this volume, for a discussion of the historical development of remote sensing).

The use of off-the-shelf DSLR cameras is somewhat limiting for quantitative spectral analysis of digital astronaut photography due to the generally broad bandpasses of the visible blue, green, and red channels, particularly in images taken with moderate to low illumination (Fig. 8). Without application of wavelength-limiting filters, these broad bandpasses make application of traditional spectral analysis and classification techniques difficult due to potential contamination of spectral information from adjacent channels. High- to medium-resolution digital astronaut photographs, however, contain significant spatial information content as recorded by pixel-to-pixel brightness variations, and this information can be capitalized on using sophisticated classification techniques such as object-based image analysis (OBIA).

Object-based image analysis uses image segmentation to identify homogeneous image objects at several different scales, rather than the classical pixel-based (and single scale) classification approaches. Membership of pixels in a given image object is determined by rule-based analysis using fuzzy classification algorithms incorporating spectral character, shape, and neighborhood relationships across the class levels (Baatz et al. 2008; Blaschke et al. 2004). The application of OBIA to high-resolution digital astronaut photographs was explored for land cover classification of both urban and coastal ecosystems and found to perform comparably to classifications derived from orbital multispectral data (Stefanov and Vande Castle 2006).

Astronaut photography presents challenges with regard to preprocessing and quantitative analysis, but these challenges are not insurmountable with currently available image-processing software. Variable look angles, acquisition times, and resolutions inherent to the astronaut photography dataset make it a highly useful addition to more traditional sources of remotely sensed data, particularly for time-series analysis and change detection (Gebelein and Eppler 2006; Robinson et al. 2002; Stefanov et al. 2003). New technology such as the GeoCam Space system described above is expected to enhance this usefulness. Astronaut photography from the ISS supported investigations of atmospheric phenomena such as aurora and polar mesospheric (or noctilucent) clouds; sea ice transport and plankton blooms; and snow cover during the International Polar Year of 2007–2008 (Evans et al. 2006). More recently, astronaut imagery from the ISS has been used in investigations of atmospheric sprites (Jehl et al. 2013) and studies of the Earth surface at night (Kyba et al. 2015). The following examples illustrate applications of astronaut photography to geomorphic mapping, urban ecology, and volcano monitoring and disaster response. These examples are intended to convey a sense of the potential of astronaut photography for Earth remote sensing, and are by no means all-inclusive.

Mapping of Megafans (“Inland Deltas”)

Astronaut imagery has made an unexpected contribution to the study of world landscapes. It revealed the existence of numerous *megafans*, also known (inaccurately) as inland deltas, hundreds of km in radius, in many parts of the world, where only a few had been reported in local regional geological literature (Fig. 9). Because megafans are dramatic in their size – up to 650 km in radius, with a huge

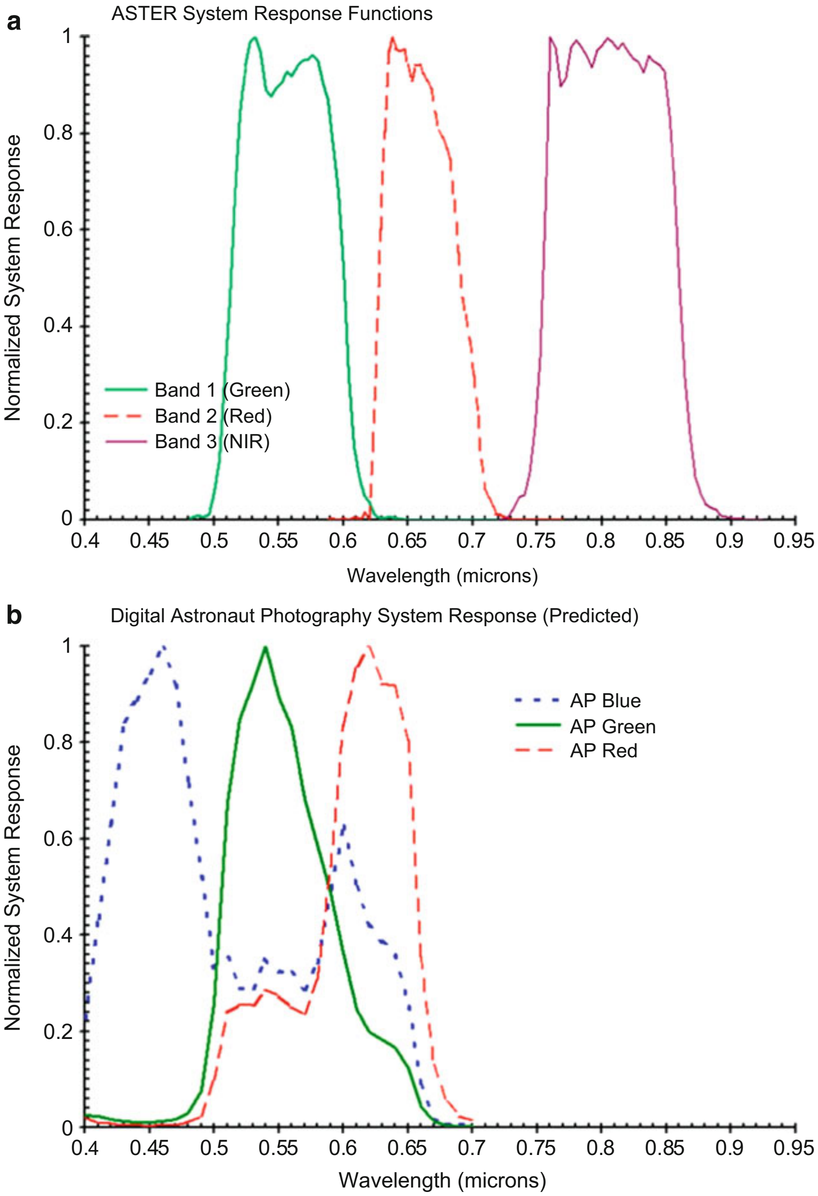


Fig. 8 Normalized system response curves for the visible *blue*, *green*, and *red* bands of the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor (**a**) and predicted response for Kodak DCS 760 camera flown on the ISS (**b**). Note the overlap below 0.6 system response between *blue*, *green*, and *red* curves for the Kodak camera

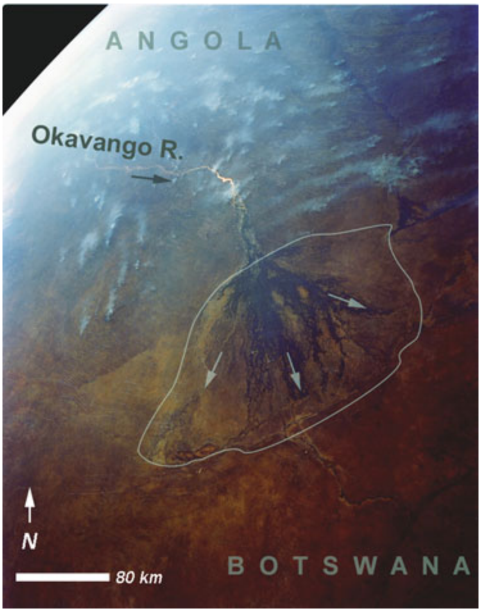


Fig. 9 Okavango megafan (outlined), northwest Botswana, is one of the most visible and photogenic large fans, with its vegetated fingers radiating from an apex, where the Okavango River (a highly reflective section appears in sunglint above the apex in this northwest-looking view) spreads laterally (*arrows*). The radius of this fan is 140 km. This is a well-known visual cue to astronauts circling over Southern Africa. NASA image STS043-151-32 was acquired on 8 August 1991

200,000 km^2 area for the largest – there had been several claims in the literature for the "largest fan" on the planet. The global astronaut view suggested that a systematic global survey was needed, especially since claims for the largest fan *excluded* the largest that had actually been described.

By showing unknown examples of these features in various parts of the world, handheld imagery provoked a wider systematic search, which proved productive scientifically. Combined with other imagery, the global view revealed the significance of large fans. First, large fans can be claimed to be a significant landform because it is widespread (Hartley et al. 2010). Prior regional studies had shown the existence of a few of these large features. Low local numbers perhaps lulled geologists and geomorphologists into thinking these features were merely the insignificant tail end of the alluvial fan population (*alluvial fans* are fan-shaped cones of sediment that accumulate at the foot of mountains, being very common in the American Southwest, with small radii, usually <25 km [Blair and McPherson 1994]). Cataloging modern megafans worldwide showed their presence on all continents, with >160 very large fans (radii >100 km: Wilkinson et al. 2010) now known between 55 °N and 55 °S. Megafans can be so large that they constitute major features on many continental surfaces, especially when nested – there are an estimated 1.2 million km^2 in South America alone. The significance of this landform is suggested by the fact that almost every flat zone (with areas >100,000 km^2) on the continents is occupied by megafans (Wilkinson 2010). An example is shown from the roughness map of the eastern Sahara Desert and plains of central South America (Fig. 10).

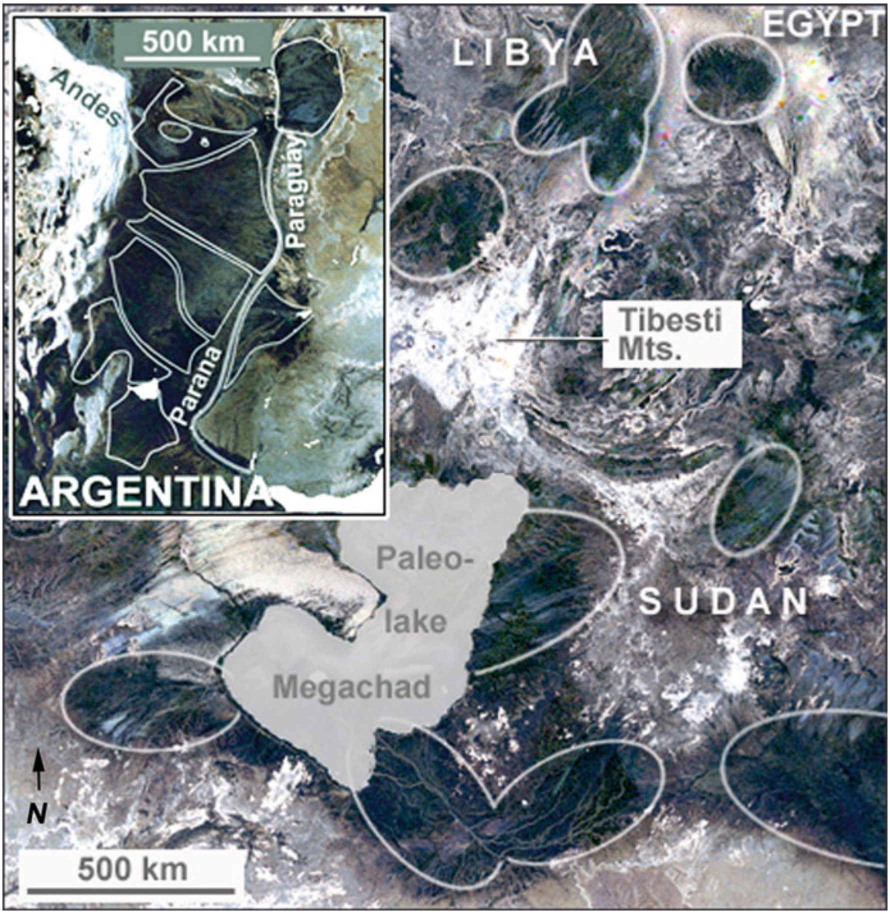


Fig. 10 Darkest tones in this roughness map of the Eastern Sahara Desert indicate areas of low slope, smooth topography covering very wide areas (note scale bars). All are surfaces formed by nested megafans (outlined). The inset figure is a similar map for the extensive plains of central Argentina and Paraguay provided for comparison. Rougher surfaces such as mountains and dune fields are light toned

Several aspects of this "find" in the astronaut imagery dataset are intriguing. Geologists and geomorphologists had assumed they knew the broad makeup of continental surfaces fairly well, especially in terms of major features, and especially in cloudfree more desertic zones where remotely sensed imagery had promoted familiarity with the landscapes. The variable perspective and ground resolutions of the astronaut imagery revealed an entire *set* of unrecognized landscapes and landscape relationships. The astronaut photography helped to clarify what now seems a simple point that rivers lay down vast quantities of material on continents, far from the oceans. Controls are an erosional zone, a river, and a neighboring basin to accommodate the river sediment. If the basin is wide the fans will be large.

Research based on the new understanding that megafan formation is a normal midscale component of river behavior on continental surfaces has been published in diverse fields. The megafan model is the basis for a theory of Amazonian landscape development (Wilkinson et al. 2010), a set of theories of fish speciation (Wilkinson et al. 2006), and the basis of a fluvial theory of evolution of the Martian surface where rover *Opportunity* has been performing its investigations (Wilkinson 2010).

Future study of astronaut imagery is likely to yield other scientifically important finds, because it has the potential to show us new features and processes on the Earth's surface. Also, the human brain "behind the lens" can accomplish the critical function of selection, thereby reducing the amount of imagery to be processed. A new view, even of the multiscale oblique kind, can lead to a significant change in what we see, e.g., recognition of the megafan landform class.

Urban Geography and Ecology

Large urban areas are easily recognizable from orbit, and cities have been a frequent subject of astronaut photographs since the Gemini program. The rich historical record of city photographs acquired during the Apollo, Skylab, and Shuttle-*Mir* programs was used by Robinson et al. (2000) to compare urban growth patterns and rates with population change for six North American metropolitan areas (Vancouver, Chicago, San Francisco, Dallas/Fort Worth, Las Vegas, and Mexico City). Photographs spanning a 28 year period were digitized, coregistered, and resampled to uniform pixel size to enable mapping of urban areal extent and measurement of change (Fig. 11). The results of the study provided insight into the variability of rates of urban land cover expansion relative to urban population growth in large North American metropolitan areas during 1969–1999. The first 2 years of this period predate publically available multispectral data from automated orbital sensors and highlight the temporal extent of the astronaut photography dataset.

The transition from film to digital cameras during the first decade of the twenty-first century facilitated the application of digital image processing and analysis techniques to astronaut photography. A comparison of true-color (RGB) digital camera imagery of the Paris, France metropolitan area, obtained from the ISS, with visible-near infrared data of similar spatial resolution from the Advanced Spaceborne Thermal Emission and Reflection (ASTER) sensor on board the Terra satellite (acquired 2 weeks earlier) was performed by Stefanov et al. (2003). The primary goal of this study was to identify the nature of yellow agricultural fields observed in the astronaut photograph that had no comparable signature in the ASTER data. Digital number values (DN, representing pixel brightness) recorded by both ASTER and the digital camera (Kodak 760 DCS) in the green and red bands were highly correlated for a number of vegetation and soil classes (Fig. 12). While the results of the study indicated that the data from both imagers could be reliably compared, the difference in coloration of the agricultural fields resulted from the flowering of rapeseed (*Brassica* sp.), a common commercial crop in the region. The digital astronaut photograph (image ISS004-E-10414, acquired on 24 April 2002) captured the phenological change of the rapeseed which had not occurred at the time

Fig. 11 Change in Mexico City urban extent from 1969 (*inner gold polygon*) to 1996 (*outer red polygon*) – an 112.3 % increase in built-up area. The Mexico City metropolitan area experienced less increase of built area relative to population growth over this time period compared to other North American cities examined by Robinson et al. (2000). Base image NM22-741-54B was acquired in December 1996

of the ASTER overpass 2 weeks earlier. The study highlights the power of combining temporally variable astronaut photography with the more regularly acquired automated sensor datasets to obtain denser time series for urban ecological and agricultural monitoring purposes.

Volcanic Eruptions and Hazard Monitoring

The inclined equatorial orbit of the ISS allows observation of active volcanoes located within (or near to) approximately 52° North or South latitude at variable times, providing the potential to capture data on eruptive activity outside the repeat frequency of polar-orbiting satellite sensors. Frequent communication between ISS crews, ground controllers, and even the general public through Internet social networking sites also provides a rapid-response capability for potentially hazardous eruptions. This capability was dramatically demonstrated on May 23, 2006, when ISS Expedition 13 Flight Engineer Jeff Williams observed volcanic activity at Cleveland Volcano, located within the Aleutian chain of islands extending westwards from Alaska (Fig. 13). Eruptions of Aleutian volcanoes can pose hazards to transcontinental airline flights because volcanic ash can disable jet engines. Cleveland Volcano was not heavily instrumented by the United States Geological Survey (USGS) at that time, and Williams was the first person to see (and report to the USGS) the eruptive activity from his vantage point in orbit.

The frequently serendipitous location of the ISS within observing distance of erupting volcanoes has also provided unique imagery of value to the volcanological community. The striking astronaut photograph of a pileus cloud forming during the large 2009 eruption of Sarychev Peak Volcano in the Kuril Islands chain is one of the

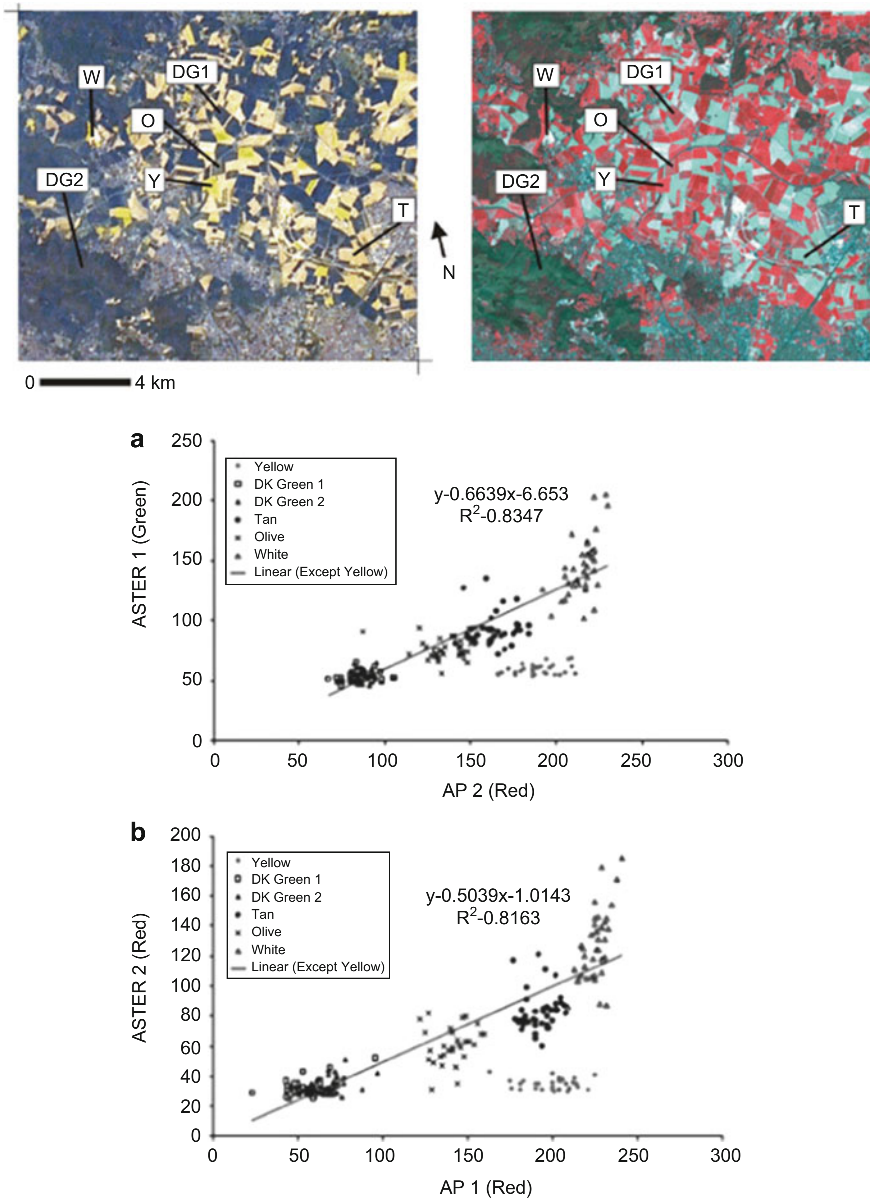


Fig. 12 Comparison of visible RGB digital astronaut photograph (AP; *upper left*) and visible to near-infrared (bands 4, 3, 2 as RGB) ASTER data (*lower left*) for agricultural fields in the northwestern Paris, France metropolitan area. Y – yellow: vegetation, high productivity; DG1 – dark green 1: vegetation, moderate to low productivity; DG2 – noncanopied vegetation; T – tan: bare soil; O – olive 1: sparsely vegetated soil; W – white: light colored soils and built materials. (**a**) green band correlation; (**b**) red band correlation

Fig. 13 Ash plume eruption from Cleveland Volcano on Chuginadak Island in the Aleutian Islands chain as seen from the International Space Station. The image (ISS013-E-24184) was acquired by the Expedition 13 crew on May 23, 2006

best records of this unusual feature (Fig. 14). The pileus cloud was formed by rapid lifting of a moist air mass by the eruption column, and sequence photography of the eruption allows for analysis of the cloud formation and plume dynamics (Lockwood and Hazlett 2010; Venzke et al. 2009).

The ISS, and handheld digital camera imagery, now play a role in data collection for disaster response. NASA remote sensing assets on the station began collecting data in response to activations of the International Charter, Space and Major Disasters (also known informally as the International Disaster Charter, or IDC) in May 2012, joining other NASA remote sensing assets able to respond to disaster events (Stefanov and Evans 2015). Handheld astronaut photography of disaster-struck regions has so far comprised the greatest contribution of data from the ISS, with data collected for 19 discrete events since May 2012 and delivered to the USGS for posting on the online Hazards Data Distribution System (2016; Fig. 15). The strengths of handheld digital camera imagery for this purpose include the human decision-making capacity to recognize whether or not useful data can be collected at the time of orbital overpass, as well as the ability to recognize a data collection opportunity that was not called out as a specific imaging target – two capabilities not yet shared by automated or ground-commanded sensor systems.

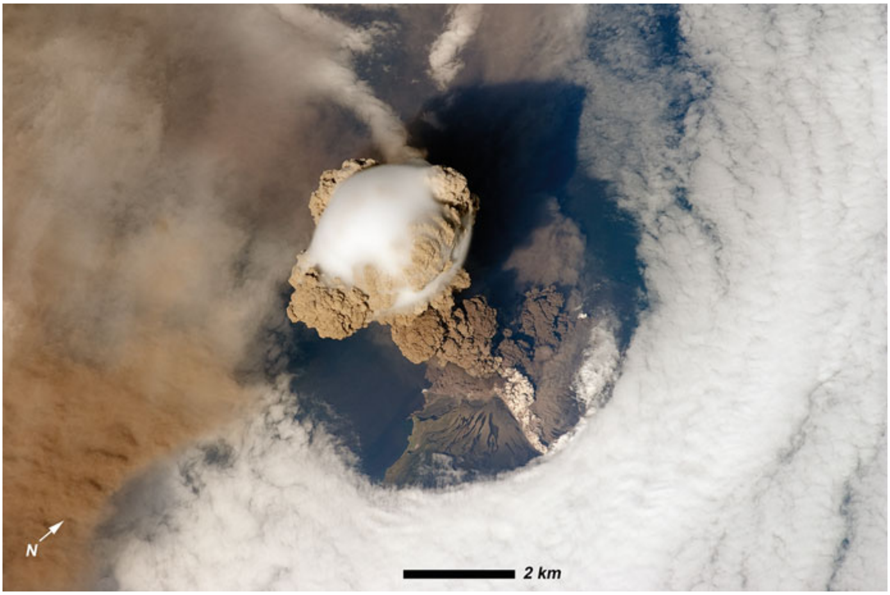
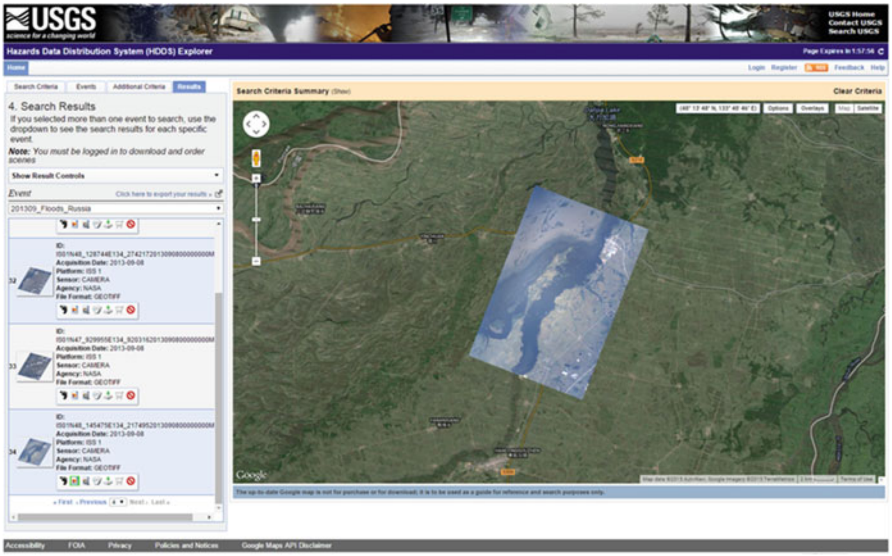


Fig. 14 Ash plume eruption, pyroclastic flows, and pileus cloud (*image center*) from Sarychev Peak Volcano, Matua Island in the Kuril Islands chain as seen from the International Space Station. The image (ISS020-E-9048) was acquired by the Expedition 20 crew on June 12, 2009



4. Search Results

If you selected more than one event to search, use the dropdown to see the search results for each specific event.

Note: *You must be logged in to download and order scenes*

Event

201309_Floods_Russia

32
ID:
Acquisition Date: 2013-09-08
Platform: ISS 1
Sensor: CAMERA
Agency: NASA
File Format: GEOTIFF

33
ID:
Acquisition Date: 2013-09-08
Platform: ISS 1
Sensor: CAMERA
Agency: NASA
File Format: GEOTIFF

34
ID:
Acquisition Date: 2013-09-08
Platform: ISS 1
Sensor: CAMERA
Agency: NASA
File Format: GEOTIFF

Fig. 15 Screen capture of the USGS Hazards Data Distribution System interface, illustrating inclusion of astronaut photography as an available dataset for selected events

Public and Education Outreach Applications

It can be posited that NASA fundamentally changed humanity's perspective of the planet Earth, and our own importance within the cosmos, with an astronaut photograph. That photograph was taken in 1972 during the Apollo 17 mission as the astronauts looked back at the small "blue marble" of Earth, surrounded by the vastness of seemingly empty and dark space, on their way to the Moon (Fig. 16). Some historians credit this image, and others like it, as helping to galvanize public opinion that led to important environmental legislation (such as the Clean Air and Water Acts) in the United States during the 1970s (Green and Jackson 2009). The public's fascination with looking at Earth from the vantage point of an astronaut on an orbiting platform continues today on the ISS (Fig. 16).

Public Outreach

Public access and distribution of handheld imagery of the Earth has evolved in step with the technological advances since the 1960s. Initially, the only way for the public to view astronaut imagery was to travel to the Earth Resources Observation and Science (EROS) Data Center in Sioux Falls, South Dakota; the Earth Data Analysis Center (EDAC) at the University of New Mexico in Albuquerque; or to the Media Service Center at NASA Johnson Space Center in Houston, Texas. With ready access to the Internet, today the global public can view and download astronaut imagery of the Earth through the Gateway to Astronaut Photography of Earth (2016), or GAPE, website (described above). Public dissemination of astronaut photography now includes social media outlets, including "tweeting" images directly from ISS via the social networking application Twitter, as well as third-party websites featuring astronaut photographs. Several astronauts have shared their unique views of the Earth by tweeting images along with commentary to the general public in near real-time, providing a direct connection with humans working and living in space. These same images can also be accessed through the GAPE website at various resolutions after working through the downlink and database entry procedure described above.

In recent years, the value of enlisting "citizen scientists" to participate in basic data analysis of large datasets has been proven by numerous online projects, such as Stardust@home and Galaxy Zoo (Wiggins and Crowston 2010). A public cataloging module, the "Image Detective" is now available on the GAPE. This module takes advantage of the public's interest in astronaut images and ready access to geospatial data (e.g., Google Earth) to provide an opportunity to participate in enhancing the metadata for the astronaut imagery of the Earth. The citizen scientist can review the dataset of existing Earth imagery, select images, and add metadata such as the location of the center point of the image, cloud percentage, and identification of major features. Planned enhancements to the Image Detective include integration with the GeoRef tool (described above) to enable full georeferencing of astronaut imagery.

Fig. 16 Apollo 17 "blue marble" view of the Earth showing the African and Antarctic continents. Image AS17-148-22727 was taken on 7 December 1972

Education Outreach

Originally eight NASA Teacher Resources Centers (now called Educator Resource Centers – the number has expanded to 11) distributed slides, prints, and 12-in. laser discs, of astronaut imagery on request to educators, museums, etc. Astronaut photographs of Earth have also appeared in textbooks, classroom lectures, and teacher workshops because they can easily be blended into existing classroom curricula without requiring sophisticated image processing or understanding of remote sensing – the visible wavelength, true-color imagery is intuitively interpreted by students. NASA education specialists from Oklahoma State University-Stillwater created a curriculum notebook based on astronaut handheld photography and satellite imagery to teach remote sensing as a tool for detecting Earth's features and changes. Themed slide sets of Earth phenomena such as weather/clouds and terrestrial impact craters from existing Shuttle imagery were made available to the public and to educators from the Lunar and Planetary

Institute in Houston, Texas. In the mid-1990s, the Houston Museum of Natural Science integrated digital astronaut images into an interactive computerized display in the museum where students could direct virtual flyovers of anywhere in the world. As with public outreach, the education outreach applications of astronaut imagery have also evolved over time.

An initial effort to provide online resources for educators was the Earth from Space photograph collection, now incorporated into the GAPE website. The content and format of the collection was informed by interactions with educators from all over the United States. Specifically, educators wanted searchable and themed imagery (e.g., cities, earth landscapes, hurricanes and weather) with extensive captioning, including an indication of the north direction on the image. Earth from Space was originally populated with Space Shuttle imagery and is today updated with ISS imagery; there are now over one million Earth images in the collection.

Direct interaction between educators, NASA scientists, and students are facilitated through the accessibility of astronaut photography. The ESRS Crew Earth Observation (CEO) Facility operations team and Teacher From Space Office, in cooperation with the NASA Aerospace Education Services Project-Montana/Nevada; Earth Observing System Education Project, University of Montana; and the Geospatial Research Group, Department of Geology, University of Montana developed an educational project called “Lewis and Clark’s Travels: The Astronaut View” in 2003. The ESRS Unit worked with the ISS Expedition 7 crew to take images along the 1804 path of Lewis and Clark to commemorate the 200th anniversary of their expedition. The ISS crew actively participated in this historical event and often drew parallels between the ISS mission and the Lewis and Clark Expedition. The images, highlighting locations along Lewis and Clark’s westward trek to the Pacific coast, were used by educators and historians and incorporated into museum displays, educational curriculum, and public websites across the United States from Philadelphia and Ohio through Montana to Oregon.

NASA continues to fund efforts to engage students in Science, Technology, Engineering and Math (STEM)-related authentic research that incorporates astronaut photography as a primary dataset. The Expedition Earth and Beyond (EEAB) project (2016) is a student involvement program that allows teachers and their students in grades 5–14 to be actively involved in the process of science. Student-driven, authentic research projects are used to study the Earth and in some cases to be compared to other planetary bodies. Student teams are mentored by scientists at the NASA JSC and in select cases may request new imagery be acquired from the ISS. The Sally Ride EarthKAM project (2016) also collects Earth imagery from the ISS using handheld digital cameras mounted in the Cupola or other Station windows to support educational activities for middle school students. Students request imagery of specific locations during several EarthKAM missions held throughout the year, and the cameras are programmed to collect imagery during specific time intervals when the locations will be visible.

Conclusion

Astronauts have taken images of the Earth using handheld cameras from the early days of space travel, and it is likely that they will continue to do so even with the development of ever-more capable automated sensor systems on free-flyer platforms and on the ISS. Indeed, the availability of regularly collected, standardized, and publically accessible remotely sensed data of Earth systems from orbital satellite platforms was enabled by early astronaut photography. Collection of handheld imagery of the Earth progressed from the essentially informal activities of USSR and USA astronauts in the 1960s through more formal programs and experiments during the 1970s and 1980s. Beginning in the late 1990s through today, collection of astronaut photography from long-duration orbital missions has enabled the use of this dataset for true time-series monitoring of selected targets as well as contributing to disaster response and management. The digital nature of current astronaut photography facilitates the application of standard image processing and analysis techniques, increases the volume of data acquired, and decreases the lag time between collection of data and availability to the public.

The human decision-making and response element inherent to astronaut photography continues to provide value and relevance within the current paradigm of Earth-observing automated sensors on satellite platforms. The ISS platform and the unconstrained pointing capability of handheld cameras allow for the collection of a dataset that is fundamentally different, yet fully complementary, to polar-orbiting sensor data due to the range of illumination conditions, viewing angles, and ground resolutions available. This enables the acquisition of several unique datasets, including high-resolution night time imagery of urban areas and detailed imagery of eruption plumes. The ready accessibility of digital astronaut imagery through online databases and collections fosters use by academic, government, and nonprofit organizations for a variety of purposes.

It is expected that handheld astronaut photography will continue to integrate with, and complement, further enhancements of remote sensing capability on the International Space Station. As off-the-shelf digital camera technology improves, so do the opportunities for modification and use in orbit to collect new, unique datasets. Addition of high-definition video cameras to the ISS presents the opportunity to capture and target dynamic events such as wildfire plumes, hurricanes, volcanic eruptions, etc. The emerging human spaceflight programs of other nations may choose to develop their own handheld Earth observation programs; if so, the historical and current programs of the USA and USSR/Russia will provide useful models for data collection and dissemination.

Cross-References

- ▶ Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing
- ▶ Fundamentals of Remote Sensing Imaging and Preliminary Analysis
- ▶ Introduction and History of Space Remote Sensing
- ▶ Processing and Applications of Remotely Sensed Data

References

D.L. Amsbury, Geological comparison of spacecraft and aircraft photographs of the Potrillo Mountains, New Mexico, and Franklin Mountains, Texas, in *Proceedings of the Sixth International Symposium on Remote Sensing of Environment*, vol. 2 (Environmental Research Institute of Michigan, Ann Arbor, 1969), pp. 493–515

D.L. Amsbury, United States manned observations of earth before the space shuttle. Geocarto Int. **1**, 7–14 (1989)

Anonymous, *Earth Photography from Gemini VI Through XII: National Aeronautics and Space Administration Special Publication 142* (National Aeronautics and Space Administration, Washington, DC, USA, 1968), 135 p

M. Baatz, C. Hoffmann, G. Willhauck, Progressing from object-based to object-oriented image analysis, in *Object-Based Image Analysis*, ed. by T. Blaschke, S. Lang, G.J. Hay (Springer, Berlin, 2008), pp. 29–32

P.C. Badgley, L. Miloy, L.F. Childs, *Oceans from Space* (Gulf Publishing Company, Houston, 1969)

D. Bannert, *Plate Drift in the Afar and Issas Territory (French Somalia) and Eastern Ethiopia as Seen on Space Photography*. NASA TN-D-6277 (1972), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19720020695_1972020695.pdf. Accessed 2 Feb 2016

T.C. Blair, J.G. McPherson, Alluvial fans and their natural distinction from rivers based on morphology, hydraulic processes, sedimentary processes, and facies assemblages. J. Sediment. Res. **A64**(3), 450–489 (1994)

T. Blaschke, C. Burnett, A. Pekkarinen, New contextual approaches using image segmentation for object-based classification, in *Remote Sensing Image Analysis: Including the Spatial Domain*, ed. by F. De Meer, S. de Jong (Kluwer, Dordrecht, 2004), pp. 211–236

L.D. Carter, R.O. Stone, Interpretation of orbital photographs. Photogramm. Eng. **40**, 193–197 (1974)

J. Cimino, C. Elachi, M. Settle, SIR-B-the second shuttle imaging radar experiment. IEEE Trans. Geosci. Remote Sens. **GE-24**(4), 445–452 (1986)

W.D. Compton, *Where No Man Has Gone Before: A History of Apollo Lunar Exploration Missions*. NASA History Series NASA SP-4214 (National Aeronautics and Space Administration, Washington, DC, 1989). 428 p

E.M. Cortright, *Exploring Space with a Camera*. NASA SP-168 (1968), http://history.nasa.gov/SP-168/sp168.htm. Accessed 2 Feb 2016

A.J. Derr, *Photography Equipment and Techniques: A Survey of NASA Developments*. NASA History Series NASA SP-5099 (National Aeronautics and Space Administration, Washington, DC, 1972). 25 p

J.E. Dornbach, *Analysis of Apollo AS-501 Mission Earth Photography*. NASA TM-X-58015 (1968), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19680026995_1968026995.pdf. Accessed 2 Feb 2016

C. Elachi, W.E. Brown, J.B. Cimino, T. Dixon, D.L. Evans, J.P. Ford, R.S. Saunders, C. Breed, H. Masursky, J.F. McCauley, G. Schaber, L. Dellwig, A. England, H. Macdonald, P. Martin-Kaye, F. Sabins, Shuttle imaging radar experiment. Science **218**(4576), 996–1003 (1982)

D. Eppler, S. Runco, Earth observations capabilities of the Window Observational Research Facility on board the International Space Station, in *Proceedings of the American Institute of Aeronautics and Astronautics Conference on International Space Station Utilization*, contribution AIAA-2001-91401 (American Institute of Aeronautics and Astronautics, Reston, VA, USA, 2001)

European Space Agency, *Cupola* (2011), http://www.esa.int/esaHS/ESA65K0VMOC_iss_0.html. Accessed 2 Feb 2016

D.L. Evans, J.J. Plaut, E.R. Stofan, Overview of the spaceborne imaging radar-C/X-band synthetic aperture radar (SIR-C/X-SAR) missions. Remote Sens. Environ. **59**(2), 135–140 (1997)

C.A. Evans, D.R. Pettit, S. Runco, G. Byrne, K. Willis, J. Heydorn, W.L. Stefanov, M.J. Wilkinson, M. Trenchard, in *International Polar Year Observations from the International Space Station*. Eos Transactions of the American Geophysical Union 87(52), Fall Meeting Supplement, Abstract IN41A-0878 (2006)

C.A. Evans, M.J. Wilkinson, W.L. Stefanov, K. Willis, Training astronauts to observe the Earth from the Space Shuttle and International Space Station, in *Analogs for Planetary Exploration*, ed. by W.B. Garry, J.E. Bleacher. Geological Society of America Special Paper 483 (Geological Society of America, Boulder, CO, USA, 2011), pp. 67–73

G.C. Ewing, *Oceanography from Space* (Woods Hole Oceanographic Institution, Woods Hole, 1965). Ref. No. 65-10, 469 p

Expedition Earth and Beyond Project (NASA Johnson Space Center, Houston, 2016), http://ares.jsc.nasa.gov/ares/eeab/index.cfm. Accessed 2 Feb 2016

W.B. Foster, *Earth Photographs form Gemini III, IV, V: National Aeronautics and Space Administration Special Publication 129* (National Aeronautics and Space Administration, Washington, DC, USA, 1967a), 266 p

W.B. Foster, O. Smistad, *Gemini Experiments Program Summary: National Aeronautics and Space Administration Special Publication 138* (National Aeronautics and Space Administration, Washington, DC, USA, 1967b), pp. 221–230

Gateway to Astronaut Photography of Earth (NASA Johnson Space Center, Houston, 2016), http://eol.jsc.nasa.gov. Accessed 2 Feb 2016

J. Gebelein, D. Eppler, How Earth remote sensing from the International Space Station complements current satellite-based sensors. Int. J. Remote Sens. **27**(13), 2613–2629 (2006)

L.E. Giddings, *Index Maps for Gemini Earth Photography*. NASA JSC-09581 (1975)

J.R. Gill, *Science Experiments Summary: National Aeronautics and Space Administration Special Publication 138* (National Aeronautics and Space Administration, Washington, DC, USA, 1967), pp. 291–305

N.F. Glazovskiy, L.V. Dessinov, Russian visual observations of Earth: historical perspective, in *Dynamic Earth Environments: Remote Sensing Observations from Shuttle-Mir Missions*, ed. by K.P. Lulla, L.V. Dessinov, C.A. Evans, P.W. Dickerson, J.A. Robinson (Wiley, New York, 2000), pp. 15–24

K. Green, M.W. Jackson, Timeline of key developments in platforms and sensors for Earth observations, in *Earth Observing Platforms & Sensors, Manual of Remote Sensing*, ed. by M.W. Jackson, vol. 1.1, 3rd edn. (American Society for Photogrammetry and Remote Sensing, Bethesda, 2009), pp. 1–48

K. Green, C. Lopez, Basics of remote sensing systems, in *Earth Observing Platforms & Sensors, Manual of Remote Sensing*, ed. by M.W. Jackson, vol. 1.1, 3rd edn. (American Society for Photogrammetry and Remote Sensing, Bethesda, 2009), pp. 49–106

J.M. Grimwood, *Project Mercury: A Chronology*. NASA SP-4001 (1963), http://history.nasa.gov/SP-4001/cover.htm. Accessed 2 Feb 2016

J.M. Grimwood, B.C. Hacker, P.J. Vorzimer, *Project Gemini Technology and Operations – A Chronology*. NASA SP-4002 (1969), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19690027123_1969027123.pdf. Accessed 2 Feb 2016

A.J. Hartley, G.S. Weissmann, G.J. Nichols, G.L. Warwick, Large distributive fluvial systems: characteristics, distribution, and controls on development. J. Sediment. Res. **80**, 167–183 (2010)

Hazards Data Distribution System (United States Geological Survey, Reston, 2016), http://hddsexplorer.usgs.gov/. Accessed 2 Feb 2016

M.R. Helfert, C.A. Wood, The NASA Space Shuttle Earth Observations Office. Geocarto Int. **1**, 15–23 (1989)

International Space Station Instrument Integration Interface (NASA Johnson Space Center, Houston, 2016), http://issearthserv.jsc.nasa.gov/i4.html. Accessed 2 Feb 2016

A. Jehl, T. Farges, E. Blanc, Color pictures of sprites from non-dedicated observations on board the International Space Station. J. Geophys. Res. Space Phys. **118**, 1–8 (2013)

E.M. Jones, K. Glover (eds.), *Apollo Lunar Surface Journal* (2010), http://history.nasa.gov/alsj/. Accessed 2 Feb 2016

J.L. Kaltenbach, *Science Report of the 70-Millimeter Earth Photography of the Apollo 6 Mission*. NASA Technical Note S-217 (National Aeronautics and Space Administration, Houston, TX, USA, 1969a)

J.L. Kaltenbach, *Science Screening Report of the Apollo 7 Mission 70-Millimeter Photography and NASA Earth Resources Aircraft Mission 981 Photography*. NASA TM-X-58029 (National Aeronautics and Space Administration, Houston, TX, USA, 1969b), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19690021200_1969021200.pdf. Accessed 2 Feb 2016

J.L. Kaltenbach, *Apollo 9 Multispectral Photographic Information*. National Aeronautics and Space Administration, Technical Memorandum X-1957 (1970), 34 p

J.L. Kaltenbach, W.B. Lenoir, M.C. McEwen, R.A. Weitenhagen, V.R. Wilmarth, *Skylab 4 Visual Observations Project Report*. NASA TM-X-58142 (1974), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19760009472_1976009472.pdf. Accessed 2 Feb 2016

G.P. Kenney, *Skylab Program, Earth Resources Experiment Package Sensor Performance Report Volume 7 (S190B): SL2, SL3 and SL4 Evaluations*. NASA MSC-05528 (1974), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19750008509_1975008509.pdf. Accessed 2 Feb 2016

G.P. Kenney, *Skylab Program, Earth Resources Experiment Package Sensor Performance Evaluation Volume 1 (S190A)*. NASA CR-144563 (1975)

H.A. Kuehnel, *Apollo Experience Report: Photographic Equipment and Operations During Manned Spaceflight Programs*. NASA TN D-6972 (1972)

C.C.M. Kyba, S. Garz, H. Kuechly, A. Sánchez de Miquel, J. Zamorano, J. Fischer, F. Hölker, High-resolution imagery of the Earth at night: new sources, opportunities and challenges. Remote Sens. **7**, 7–23 (2015)

J.P. Lockwood, R.W. Hazlett, *Volcanoes: Global Perspectives* (Wiley-Blackwell, Chichester, 2010)

H.E. Lockwood, G.E. Sauer, Processing corrections for Skylab photographic imagery. Photogramm. Eng. Remote Sens. **41**(4), 523–532 (1975)

P.D. Lowman Jr., *A Review of Photography of the Earth from Sounding Rockets and Satellites*. NASA TN D-1868 (1964)

P.D. Lowman Jr., *Space Photography – A Review; Photogrammetric Engineering*, vol. XXXI (American Society of Photogrammetry, Falls Church, VA, USA, 1965), pp. 76–86

P.D. Lowman Jr., *The Earth from Orbit: National Geographic Magazine*, November, 1966 (National Geographic Society, Washington, DC, USA, 1966a), pp. 644–671

P.D. Lowman Jr., Photography from space-geological applications. Ann. N. Y. Acad. Sci. **140**, 99–106 (The New York Academy of Sciences, New York, NY, USA, 1966b)

P.D. Lowman Jr., *Geologic Orbital Photography – Experience from the Gemini Program*. NASA X-644-68-228 (1968), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19680018143_1968018143.pdf. Accessed 2 Feb 2016

P.D. Lowman Jr., *Geologic Analysis of Apollo 9 Multispectral Terrain Photography*. NASA GSFC-X-644-69-423 (1969a)

P.D. Lowman Jr., Geologic orbital photography: experience from the Gemini program. Photogrammetria **24**(3–4), 77–106 (1969b)

P.D. Lowman Jr., *The Third Planet* (Weltflugbid, Zurich, 1972)

P.D. Lowman Jr., H.A. Tiedemann, *Terrain Photography from Gemini Spacecraft Final Report*. NASA X-644-71-15 (1971), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19710008933_1971008933.pdf. Accessed 2 Feb 2016

P.D. Lowman Jr., J.A. McDivitt, E.H. White, *Terrain Photography on the Gemini IV Mission – Preliminary Report*. NASA TN-D-3982 (1967), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19670017945_1967017945.pdf. Accessed 2 Feb 2016

P.D. Lowman Jr., H.V. Frey, W.E. Shenk, L. Dunkelman, *Manual for 70 mm Hand-Held Photography from Skylab*. NASA GSFC X-644-73-147 (1973)

J.L. Mcniel, C.C. Devalcourt, *Skylab-2 Handheld Photography Alphabetized Geographical Features List*. NASA CR-134245 (1974a), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740014486_1974014486.pdf. Accessed 2 Feb 2016

J.L. Mcniel, C.C. Devalcourt, *Skylab-3 Handheld Photography Alphabetized Geographical Features List*. NASA CR-140244 (1974b), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740024691_1974024691.pdf. Accessed 2 Feb 2016

National Aeronautics and Space Administration, *Gemini MidProgram Conference – Including Experiment Results*. NASA SP-121 (1966)

National Aeronautics and Space Administration, *Earth Photographs from Gemini III, IV, and V.* NASA SP-129 (1967a)

National Aeronautics and Space Administration, *Gemini Program Flight Summary Report: Gemini Missions I Through XII*. NASA MSC-GR-66-5 (Revision A) (1967b)

National Aeronautics and Space Administration, *Gemini Summary Conference*. NASA SP-138 (1967c), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19680005472_1968005472.pdf. Accessed 2 Feb 2016

National Aeronautics and Space Administration, *Apollo AS-502 Mission Data and Information List, 70mm Color Photography*. NASA Mapping Sciences Branch, MSC (1968a), http://apollo.sese.asu.edu/SUPPORT_DATA/ap06_index.pdf. Accessed 2 Feb 2016

National Aeronautics and Space Administration, *Apollo 6 Mission Report*. NASA MSC-PA-R-68-9 (1968b)

National Aeronautics and Space Administration, *Apollo 7 Mission Report*. NASA MSC-PA-R-68-15 (1968c)

National Aeronautics and Space Administration, *Earth Photographs from Gemini VI Through XII*. NASA SP-171 (1968d)

National Aeronautics and Space Administration, *Apollo 7 Mission Data and Information List, 70mm Color Photography*. NASA Mapping Sciences Branch, MSC (1969a). http://apollo.sese.asu.edu/SUPPORT_DATA/ap07_index.pdf. Accessed 2 Feb 2016

National Aeronautics and Space Administration, *Apollo 9 Mission Report*. NASA MSC-PA-R-69-2 (1969b)

National Aeronautics and Space Administration, *Apollo 9 Photographic Plotting and Indexing Report (Including Aircraft Underflights)*. NASA Mapping Sciences Laboratory, MSC (1969c)

National Aeronautics and Space Administration, *Earth Resources Program Synopsis of Activity* (Manned Spacecraft Center, Houston, 1970a)

National Aeronautics and Space Administration, *Ecological Surveys from Space*. NASA SP-230 (1970b), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19700017671_1970017671.pdf. Accessed 2 Feb 2016

National Aeronautics and Space Administration, *Handbook of Pilot Operational Equipment for Manned Space Flight*. NASA CD42-A/SL-997 (1973a), http://history.nasa.gov/alsj/JSC-07210PltOpsEquip.pdf. Accessed 2 Feb 2016

National Aeronautics and Space Administration, *Skylab Program-EREP Investigator's Information Handbook*. NASA MSC-07874 (1973b)

National Aeronautics and Space Administration, *Apollo-Soyuz Test Project Visual Observations Debriefing*. NASA JSC-09920 (1974a)

National Aeronautics and Space Administration, *Skylab Earth Resources Data Catalog*. NASA JSC-09016 (1974b), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19750012726_1975012726.pdf. Accessed 2 Feb 2016

National Aeronautics and Space Administration, *Apollo-Soyuz Test Project: Preliminary Science Report*. NASA TM-X-58173 (1976), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19760015986_1976015986.pdf. Accessed 2 Feb 2016

National Aeronautics and Space Administration, *Skylab Explores the Earth*. NASA SP-380 (1977), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19820004619_1982004619.pdf. Accessed 2 Feb 2016

National Aeronautics and Space Administration, *Reference Guide to the International Space Station*. NASA NP-2015-05-022-JSC (2015), https://www.nasa.gov/sites/default/files/atoms/files/np-2015-05-022-jsc-iss-guide-2015-update-111015-508c.pdf. Accessed 2 Feb 2016

O.W. Nicks, *This Island Earth*. NASA SP-250 (1970), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19710003091_1971003091.pdf. Accessed 2 Feb 2016

A. Pesce, *Gemini Space Photographs of Libya and Tibesti: A Geological and Geographical Analysis* (Petroleum Exploration Society of Libya, Tripoli, 1968)

J.A. Robinson, B. McRay, K.P. Lulla, Twenty-eight years of urban growth in North America quantified by analysis of photographs from Apollo, Skylab, and Shuttle-Mir, in *Dynamic Earth Environments: Remote Sensing Observations from Shuttle-Mir Missions*, ed. by K.P. Lulla, L.V. Dessinov (Wiley, New York, 2000), pp. 25–41

J.A. Robinson, D.A. Liddle, C.A. Evans, D.L. Amsbury, Astronaut-acquired orbital photographs as digital data for remote sensing: spatial resolution. Int. J. Remote Sens. **23**(20), 4403–4438 (2002)

Sally Ride EarthKAM Project (U.S. Space & Rocket Center, Huntsville, 2016), https://www.earthkam.org/home. Accessed 2 Feb 2016

R.A. Schowengerdt, P.N. Slater, *Final Postflight Calibration Report on Apollo 9 Multiband Photography Experiment S065, Technical Memorandum 3* (Optical Sciences Center, University of Arizona, Tucson, 1972)

N.M. Short, P.D. Lowman Jr., *Earth Observations from Space: Outlook for the Geological Sciences*. NASA X-650-73-316 (1973), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740004044_1974004044.pdf. Accessed 2 Feb 2016

W.L. Stefanov, Astronaut photography: hands-on remote sensing of the Earth. Phi Kappa Phi Forum **88**(1), 2–7 (2008)

W.L. Stefanov, C.A. Evans, Data collection for disaster response from the International Space Station. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **40**(7/W3), 851–855 (2015)

W.L. Stefanov, J. Vande Castle, *Ecological Landscape Classification Using Astronaut Photography*. Eos Transactions of the American Geophysical Union 87(52), Fall Meeting Supplement, Abstract B41A-0155 (2006)

W.L. Stefanov, J.A. Robinson, S.A. Spraggins, Vegetation measurements from digital astronaut photography. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **34**(7/W9), 185–189 (2003)

R.E. Stevenson, Observations from Skylab of mesoscale turbulence in ocean currents. Nature **250**, 638–640 (1974)

R.E. Stevenson, R.M. Nelson, *An Index of Ocean Features Photographed from Gemini Spacecraft*. Contribution 253 (Bureau of Commercial Fisheries Biological Laboratory, Galveston, 1968)

L.S. Swenson Jr., J.M. Grimwood, C.C. Alexander, *This New Ocean: A History of Project Mercury*. NASA SP-4201 (1966), http://history.nasa.gov/SP-4201/toc.htm. Accessed 2 Feb 2016

R.W. Underwood, Color photography from space, in *Manual of Color Aerial Photography*, ed. by J.T. Smith Jr., A. Anson, 1st edn. (American Society of Photogrammetry, Falls Church, 1968), pp. 365–379

R.W. Underwood, J.W. Holland, *Skylab 2: Photographic Index and Scene Identification*. NASA JL12-601 (1973a), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740007997_1974007997.pdf. Accessed 2 Feb 2016

R.W. Underwood, J.W. Holland, *Skylab 3: Photographic Index and Scene Identification*. NASA JL12-602 (1973b), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740007948_1974007948.pdf. Accessed 2 Feb 2016

R.W. Underwood, J.W. Holland, *Skylab 4: Photographic Index and Scene Identification*. NASA JL12-603 (1974), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19750019464_1975019464.pdf. Accessed 2 Feb 2016

L.A. Vanderbloemen, W.L. Stefanov, C.A. Evans, *A Researcher's Guide to: International Space Station Earth Observations*. Johnson Space Center Publication NP-2013-06-011-JSC (2014), http://www.nasa.gov/sites/default/files/files/Earth-Observation-Mini-Book-042814-508.pdf. Accessed 2 Feb 2016

E. Venzke, S.K. Sennert, R. Wunderman, Reports from the Smithsonian's Global Volcanism Network, June 2009. Bull. Volcanol. **71**(10), 1211–1212 (2009)

A. Wiggins, K. Crowston, Developing a conceptual model of virtual organisations for citizen science. Int. J. Organ. Des. Eng. **1**(1–2), 148–162 (2010)

D.E. Wilhelms, *To a Rocky Moon: A Geologist's History of Lunar Exploration* (University of Arizona Press, Tucson, 1993)

M.J. Wilkinson, Fluvial sediment accommodation and mesoscale architecture – some neglected perspectives, in *First International Conference on Mars Sedimentology and Stratigraphy*, Abstract 6065 (2010), http://www.lpi.usra.edu/meetings/marssed2010/pdf/6065.pdf. Accessed 2 Feb 2016

M.J. Wilkinson, L.G. Marshall, J.G. Lundberg, River behavior on megafans and potential influences on diversification and distribution of aquatic organisms. J. S. Am. Earth Sci. **21**, 151–172 (2006)

M.J. Wilkinson, L.G. Marshall, J.G. Lundberg, M.H. Kreslavsky, Megafan environments in northern South America and their impact on Amazon Neogene ecosystems, in *Amazonia, Landscape and Species Evolution*, ed. by C. Hoorn, F.P. Wesselingh (Blackwell, Chichester, 2010), pp. 162–184

D.M. Winker, R.H. Couch, M.P. McCormick, An overview of LITE: NASA's LIDAR In-space Technology Experiment. Proc. IEEE **84**(2), 164–180 (1996)

F.J. Wobber, Orbital photography: applied Earth survey tool. Photogr. Appl. Sci. Technol. **2**(7), 21–29 (1968), 56

D. Woods, *Apollo Flight Journal* (2009), http://history.nasa.gov/afj/. Accessed 2 Feb 2016

E.O. Zeitler, T.G. Rogers, *The Gemini Program Physical Sciences Experiments Summary*. NASA TM X-58075 (1971), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19720003207_1972003207.pdf. Accessed 2 Feb 2016

Electro-Optical and Hyperspectral Remote Sensing

Scott Madry and Joseph N. Pelton

Contents

Abstract

Remote sensing satellites have become increasingly sophisticated in terms of increased spatial, radiometric, and temporal resolution. Over the past few decades, sensing devices have become more sophisticated with not only higher spatial resolution but have also now become more capable at capturing data in much more precisely defined bandwidths or frequency ranges. This provides the ability to identify particular vegetation, forestry, wildlife and fish, and minerals – even camouflage – with greater precision.

This evolution of sensor capabilities has, however, led to new needs on the ground in terms of interpreting the data. The new interpretative needs – because much more data is captured – involve requirements for new and faster processing techniques on the ground. Or it has led to the need for "preprocessing of data" (i.e., discarding noncritical or nonmeaningful data) before being downloaded from the satellite. The point is that the more capable sensors that collect a larger

S. Madry (✉)
Global Space Institute, Chapel Hill, NC, USA
e-mail: Scottmadry@mindspring.com

J.N. Pelton
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_42

amount of data serves to alter the way the torrent of data downloaded from the sky is processed. This chapter explains the transition that is rapidly occurring in terms of the transition from multispectral imaging to the much more precise and data-intensive hyperspectral sensing – also called imaging spectroscopy.

Much more capable electro-optical arrays – usually using charge-coupled devices (CCDs) – allow the capturing of hyperspectral data much more efficiently. In the past with multispectral sensing data was collected in perhaps five or perhaps as many as ten broad frequency bands. Now data can be collected in much more precise and narrower frequency bands in the infrared, near-infrared, visible spectrum, and even ultraviolet bands.

This chapter discusses the transition from multispectral to hyperspectral sensing that is now in full swing. It notes that the first uses of hyperspectral sensing were for military and defense-related purposes, but now hyperspectral sensing – using the latest electro-optical arrays – is becoming central to civil Earth Observation programs. This transition has not only meant a change in the imaging process and the types of sensor devices included on remote sensing satellites, but it has also signaled the shift in data processing formats with data being processed as "data cubes." In this format spatial data is provided along the (X, Y axis), while the various frequency bands are displayed on the vertical or Z axis.

Keywords

Charge-coupled devices (CCD) • Data cubes • Data sets • Electro-optical sensors • Hyperspectral sensing • Imaging spectroscopy • Multispectral sensing • Post-processing • Preprocessing • Radiometric resolution • Signal-to-noise performance levels • Spatial resolution • Spectral band • Spectral resolution • Spectrometer

Introduction

Recent advances in remote sensing sensor devices and improved data processing techniques have led the way for the development of two key capabilities in the field. These newer capabilities are (1) much more capable electro-optical sensors (typically using charge-coupled device (CCD) array technology) that can collect data in precise and narrow frequency bands with great precision and (2) hyperspectral sensing or imaging spectroscopy that is aimed at collecting data in hundreds of different spectra that ranges from infrared, to near infrared, to the visible light band and up to ultraviolet light.

Initially hyperspectral sensing was used for military purposes such as to distinguish camouflage from real forestry or vegetation or the detection and identification of particular ore, minerals, or vegetation. Electro-optical sensing, which benefits from technology developed for modern digital cameras and other types of ground-based technology, has also developed rapidly in the past few years and is also being used for an expanding number of purposes as well (Khorram et al. 2012).

Imaging spectroscopy has been used in the laboratory by physicists and chemists many years for identifying the composition of materials. Spectroscopy can be used to detect individual absorption features due to specific chemical bonds in a solid, liquid, or gas and thus it is a highly versatile tool. The idea to employ hyperspectral techniques in satellite remote sensing began in the mid-1980s and was initially used by geologists for locating and identifying minerals and rock structures that might help identify petroleum as well as valuable ores.

Hyperspectral remote sensing combines imaging and spectroscopy into a single system which often includes large data sets and requires new processing methods. Hyperspectral data sets are generally composed of as many as 150 to over 300 spectral bands of relatively narrow bandwidths (5–10 nm), whereas, multispectral data sets are usually composed of about five to ten bands of relatively large bandwidths (70–400 nm). Thus, as noted above, hyperspectral data sets require much more intensive and demanding data processing. This is further complicated by the fact that detection of mineral or ores is dependent on many different factors in combination. These factors include the spectral resolution and coverage, the signal-to-noise performance (or net effective gain) of the spectrometer, the abundance of the material, and the strength of absorption features for that material in the wavelength region measured. When these factors are taken all together, the data processing requirements can be orders of magnitude larger than is the case with multispectral remote sensing.

Electro-optical sensors depend on capturing photons from the imaged area and convert them into an “electronic image” that “mirrors” the imaged area. This type of imaging is most frequently now achieved using a charge-coupled device (CCD) or a larger CCD array. A CCD is an integrated circuit etched onto a silicon surface so as to form light-sensitive elements represented by what are called pixels. Photons incident on this surface generate a charge that can be read by electronics and turned into a digital copy of the light patterns falling on the device. CCDs come in a wide variety of sizes and types and are used in many applications from remote sensing satellites to digital cameras. A very high-resolution CCD array – used for hyperspectral imaging – can produce a huge amount of data for processing. In some cases this can lead to a form of “preprocessing” that can serve to eliminate or reduce data that is considered not to be of particular importance so that the data downloaded to remote sensing data centers is scaled down to levels so that the incoming data does not overload the processing center’s capacity.

In the sections that followed, the technology and application for both hyperspectral sensing and electro-optical imaging will be presented in greater detail.

Introduction to Hyperspectral Imaging

In the early days of remote sensing satellites, the initial approach was to undertake what is called multispectral sensing. This was to have sensors that collected images across a wide range of spectrum that ranged from infrared through the visible light and up to the ultraviolet frequencies. This imaging typically was across five to ten broadbands that might be as wide as 400 nm. This produced a significant amount of data and important results. The infrared imaging could help map human and animal activity and identify heat pollution in the oceans, disease in vegetation and forests, and meteorological applications. The visible light imaging was useful for a wide range of applications from weather tracking to a wide range of industrial, agricultural, and environmental applications, while ultraviolet sensing complemented the images captured at lower-frequency ranges.

Hyperspectral imagery, however, represented a whole new range of capabilities that were much more targeted and precise. The much more narrow and targeted bandwidths associated with this type of imaging allows specific minerals or ores to be identified. With multispectral sensing one could distinguish trees from meadows and lakes and ponds. With hyperspectral sensing one can distinguish lead ore from copper ore or pitchblende from quartz, or in defense-related applications, one can identify specific sites of a strategic nature. The refinement in terms of much more narrowly targeted frequencies, greater spatial resolution, high-gain spectrometers, and other enhancements, however, means that hyperspectral sensing gathers much greater amounts of data.

This acquisition of much greater amounts of data ultimately requires a much more sophisticated means to process the collected information within a reasonable amount of time, or the value of the data tends to be depreciated. In order to process this data and display it efficiently, it is typically collected (and represented) as a data cube with spatial information collected in the X-Y horizontal “spatial” plane, and spectral information is represented in the Z direction. This works quite well in flat areas to “see” what the hyperspectral sensors are “seeing,” and it can also allow for “preprocessing” in terms of discarding data from areas that are not considered of particular interest or from areas that have previously been successfully imaged (Hyperspectral Remote 2016).

In recent years, hyperspectral sensors have become increasingly sophisticated in terms of being perfected to achieve higher- and higher-performance characteristics. Thus hyperspectral sensors have increased in the following regards:

- Increased spatial resolution
- Greater spectral coverage (within narrowed bandwidth slots across a wider spectrum)
- Higher net gain (signal to noise) by the sensing spectrometers
- Greater material absorption levels

All of these factors combine to make the “data cube” that is downloaded from remote sensing satellites designed to carry out hyperspectral imaging much more “dense.” This means that these data cubes will contain a much greater amount of data and thus allow much more practical use to be made of the data collected by hyperspectral imaging satellites.

Early History of Hyperspectral Sensing

In the early 1980s, the NASA Jet Propulsion Laboratory in California developed what was called the Airborne Visible-Infrared Imaging Spectrometer (AVIRIS) and flew it on a test aircraft. Before AVIRIS, technological limitations prevented spectrometers from being used on moving platforms and produce an accurate hyperspectral image that could be processed and displayed with reliable results. In time this allowed the US Air Force to adapt this technology to be used on a test remote sensing satellite. On August 1, 2000, shortly after its July 9, 2000 launch, the Fourier Transform Hyperspectral sensor aboard the US Air Force MightySat II.1 sent back its first "hypercube" data set from orbit. This began a new phase in remote sensing from space. This first image showing a portion of rural Colorado is shown in Fig. 1 (Satellite Sends 2001).

From this image, data processors on Earth could determine highway pavement, ponds, and lakes and various types of vegetation and forestry. In terms of defense-related applications, it was able to distinguish materials that might be used to camouflage military bases, airfields, or fields used to grow drugs. The success of this test of a hyperspectral sensor has led many other parallel applications in civil space programs.

The European Space Agency developed the Proba-1 remote sensing satellite with an onboard hyperspectral spectrometer known as CHRIS (see Fig. 2). This sensor, launched in 2001, was designed to monitor 62 bands but had a spatial resolution of only 17 m. This was still sufficient to produce highly effective images. Today data from this spacecraft supports many practical and scientific applications, such as land

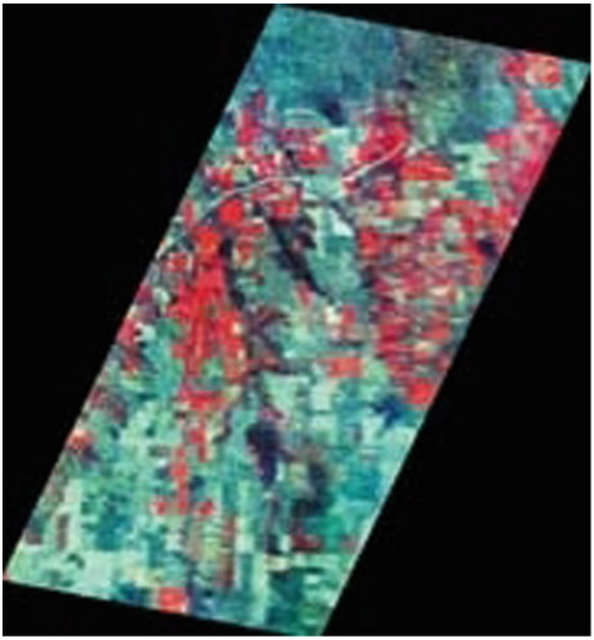


Fig. 1 Initial hyperspectral image from space captured in year 2000 (Image courtesy of US Air Force)

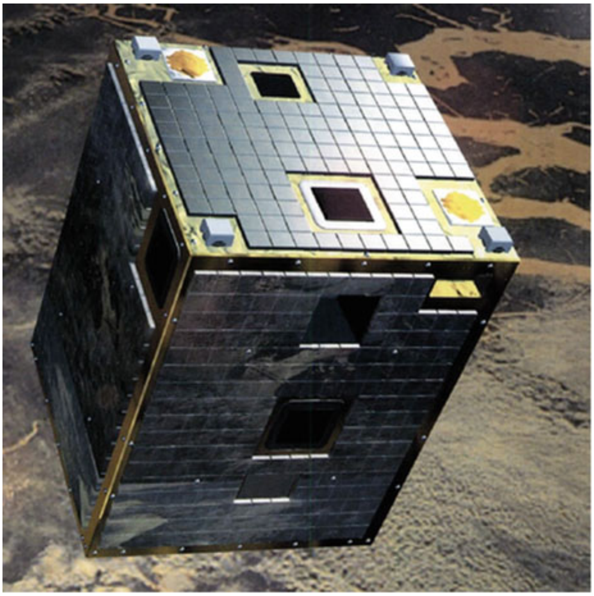


Fig. 2 Proba-1 low-earth orbit satellite by ESA with CHRIS spectrometer on board (Image courtesy of European Space Agency)

Fig. 3 Spectrometer image from Proba-1 satellite developed by ESA (Image courtesy of European Space Agency)

surface, coastal zone, and aerosol monitoring, and its imaging is actually being now used by 300 different scientific units in some 50 countries. One of the images from Proba-1 below shows with some clarity what is called the French Frigate Shoals. These shoals are an atoll about 800 km northwest of Hawaii (see Fig. 3).

NASA at the same time developed and launched the Hyperion hyperspectral sensor on its Earth Observing-1 satellite. With the success of these two hyperspectral sensors, a number of new projects to develop much higher capabilities in terms of bands and resolution were undertaken around the world.

Thus after decades of research and development, hyperspectral imaging is becoming increasingly mature. Today one can say that the transition from multispectral sensing to hyperspectral sensing is happening fairly rapidly. This relatively new technology is now greatly improving our ability to characterize the state of the Earth, monitor climate change and pollution, and locate both resources and problems around the world. In short it can probably be said with confidence that hyperspectral sensing has entered the mainstream of satellite remote sensing (Going Hyperspectral 2010).

Several new capabilities in the field are anticipated. One of these initiatives is the PRISMA satellite of Italian Space Agency (ASI). This is an Earth Observation System which will employ innovative electro-optical instrumentation for the hyperspectral imaging. The PRISMA mission combines a hyperspectral sensor with a panchromatic, medium-resolution camera. The electro-optical capability is important both in terms of sensing from space as well as processing on the ground.

The advantages of the combination of the hyperspectral electro-optical sensor and panchromatic camera are that in addition to the classical capability of observation based on the recognition of the geometrical characteristics of the scene (i.e., the panchromatic camera), there is the contrasting scene that is offered by hyperspectral sensors. The hyperpectral sensor can determine the chemical-physical composition of anything being viewed and do so regardless of whether it is solid, liquid, or gaseous.

The combined panchromatic and hyperspectral image offers the scientific community and users many applications in the field of environmental monitoring, resource management, crop classification, pollution control, and mineral prospecting. PRISMA is currently anticipated to launch by the end of 2013 (Prisma 2011).

Closely following the PRISMA smaller-scale demonstration project will be the full-scale German designed EnMAP hyperspectral satellite in 2014. This satellite will seek to map the entire surface of the world and do so in 200 narrowband frequency channels using the latest electro-optical techniques.

The primary goal of the German EnMAP satellite as designed by the German space agency DLR has been formally stated by EnMAP Project Scientist Prof. Hermann Kaufmann to be “to offer accurate, diagnostic information on the state and evolution of terrestrial ecosystems on a timely and frequent basis, and to allow for a detailed analysis of surface parameters with regard to the characterization of vegetation canopies, rock/soil targets, and coastal waters on a global scale.”

EnMAP has been designed to record biophysical, biochemical, and geochemical variables that will be displayed in hyperspectral data cubes such as shown below in Fig. 4. It is hoped that this unique ability to display the hyperspectral data in such a manner will serve to increase our understanding of the world’s chemical and biological makeup as never before possible.

Further in 2015, NASA plans to launch the so-called HyspIRI mission. This satellite will be able to obtain images in 210 separate spectral bands. The focus of HyspIRI will be to study volcanoes and volcanic eruption, water status, and nutrients

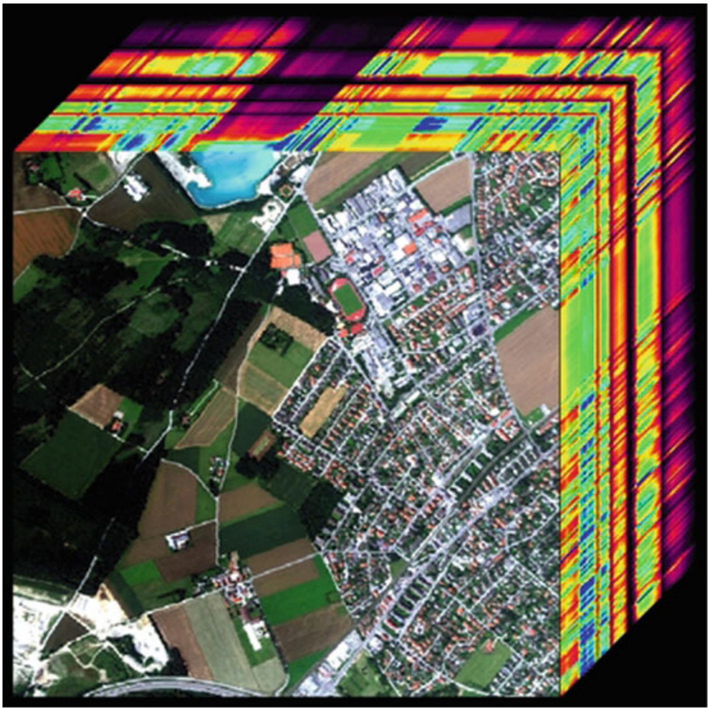


Fig. 4 A hyperspectral data cube that shows the spectra of the sensed spatial image (Image courtesy of DLR –The German Space Agency)

associated with various types of vegetation, deforestation, and new scientific information that might be able to help provide early warning of droughts. This satellite will supplement the findings from EnMap and other hyperspectral satellites.

Electro-Optical Sensing

The first hyperspectral imaging was restricted to the use of spectrometer technologies that are scientific instruments that have a number of constraints in terms of being able to collect data and effectively send it back to Earth for processing, as well as high costs associated with the equipment, the sensing and acquisition of data, and the final processing and display of the data. The development of electro-optical sensing and particularly charge-coupled devices (CCDs) and CCD arrays have served to lower the cost of hyperspectral sensing.

In short the advent of CCD devices brings many advantages to satellite-based hyperspectral sensing in terms of costs and ease of data display. The various satellites described above such as PRISMA, EnMap, and HyspIRI all use electro-optical CCD devices to gather hyperspectral images and then send them back to Earth as electronically coded information for processing.

In time remote sensing satellites designed to gather this type of information may also be designed with preprocessing capabilities to “discard” or “sort through” various images that have already been captured or where changes are not taking place.

Figure 5 above shows how a satellite equipped with a CCD array designed for hyperspectral sensing actually works to capture images into data sets that can then be efficiently sent back to Earth.

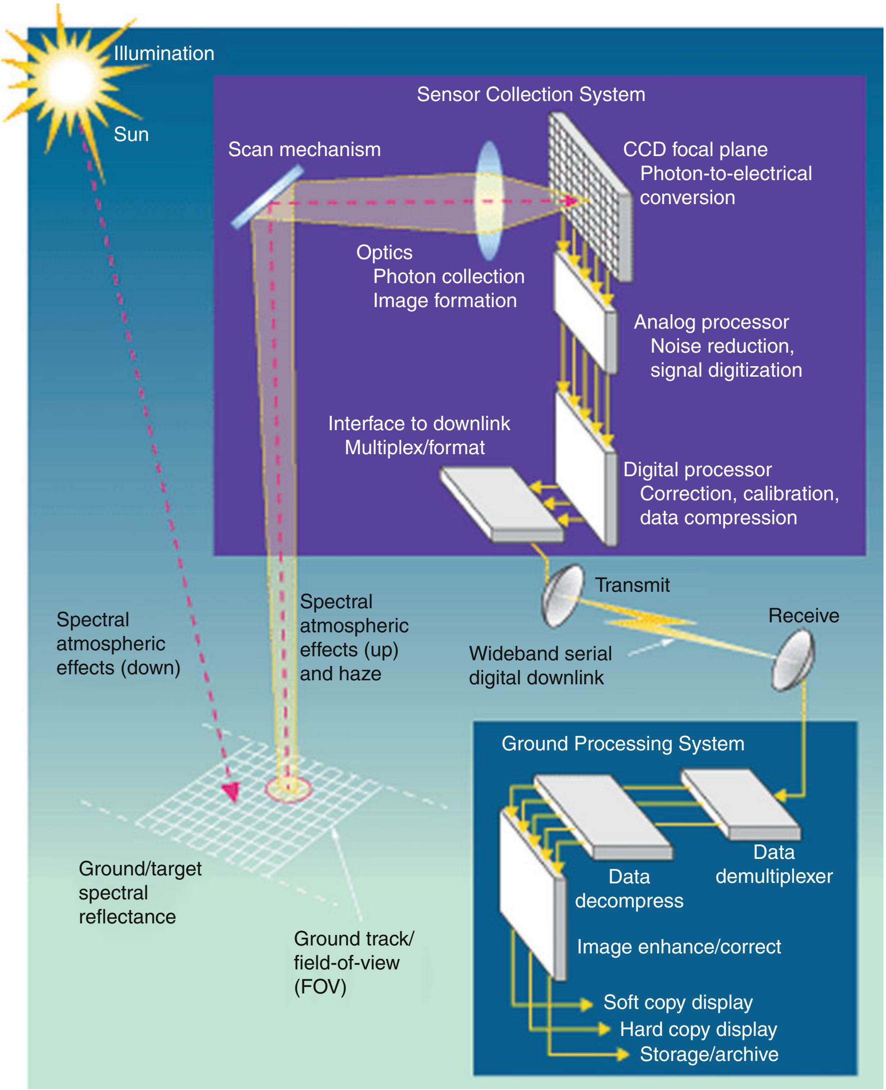


Fig. 5 Remote sensing using CCD for electro-optical imaging (D. L. Glackin)

Conclusion

The shift from multispectral sensing to hyperspectral sensing is still in the process of occurring. The value that comes from this new way of capturing and displaying data as hyperspectral data cubes is rather universally understood to represent a breakthrough in remote sensing technology. It brings with it the ability to analyze and understand much more precisely delineated images that this new type of remote sensing satellite and their sensors can tell us. This is especially true when hyperspectral imaging cubes are combined with panchromatic or radar satellites. There is still a great deal to be done in terms of perfecting the satellite sensor capabilities, the formatting of data cubes, the efficient transmission to ground stations, and the possible future of preprocessing, as well as more in-depth analysis of the data on the ground.

Today there are a growing number of applications that are evolving with regard to hyperspectral imaging. The fact that hyperspectral images tell us about gases (i.e., clouds, pollution, greenhouse gases, etc.), about liquids (oceans, lakes, rivers, pollution, droughts), and about solids (minerals, ores, vegetation, forests, etc.) strongly suggest that the scientific and practical applications can continue to expand for some time to come. The rapid evolution of charge-coupled device (CCD) technology and CCD arrays and their use in a broad range of industrial and consumer applications will doubtlessly help the evolution of this technology in terms of more effective remote sensing via satellites. It may especially lead to new economies.

Cross-References

▶ Developments in Hyperspectral Sensing
▶ Fundamentals of Remote Sensing Imaging and Preliminary Analysis

References

Going Hyperspectral, ESA Newsletter, About Proba-1, (2010), http://www.esa.int/esaMI/Proba_web_site/SEMFVBIK97G_0.html. Last accessed 2 Jan 2016

Hyperspectral Remote Sensing, (2016), http://www.csr.utexas.edu/projects/rs/hrs/hyper.html. Last accessed 2 Mar 2016

S. Khorram, F.H. Koch, C.F. van der Wiele, S.A. Nelson, *Remote Sensing* (Springer, New York, 2012)

Prisma – A.S.I.-Agenzia Spaziale Italiana, (2011), www.asi.it/en/activity/earth_observation/prisma. Last accessed 2 Jan 2016

Satellite Sends Hyperspectral Image from Space-Laser Focus World, (2001), http://www.laserfocusworld.com/articles/print/volume-37/issue-5/features/hyperspectral-imaging/satellite-sends-hyperspectral-images-from-space.html. Last accessed 2 Jan 2016

Operational Applications of Radar Images

Vern Singhroy

Contents

Abstract

This chapter provides some examples of the operational uses of satellite radar images. These include the uses of polarimetric radar images for crop classification and earthquake damage assessment, radar image fusion for mineral exploration, interferometric SAR techniques for landslide and volcanic monitoring, multidate radar image enhancement techniques for oil spill monitoring and flood mapping, and sea ice mapping from enhanced ScanSAR images. In the near future, new applications will be developed from current and future advanced SAR missions involving their high resolution, rapid revisits, and polarimetric capabilities.

Keywords

Radar applications • polarimetric and interferometric SAR • radar missions

---

V. Singhroy (✉)
Canada Centre for Remote Sensing, Natural Resources Canada, Ottawa, ON, Canada
e-mail: vern.singhroy@canada.ca

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_43

Introduction

This chapter describes the operational applications of radar images that have developed over the past 20 years. Table 1 provides a summary of current and future satellite missions and their technical specifications. The European and Canadian space agencies focus on launching high-resolution fully polarimetric C-band (5.7 cm) satellites. The German and Italian space agencies launch X-band (3 cm) satellites, and the Japanese space agency launches L-band radar satellites. All these multifrequency satellites provide images with different resolutions, viewing geometries, and polarization. In the next few years, there will be number of high-resolution radar constellation satellites that will be launched by the European and Canadian space agencies. These new-generation smaller radar constellation satellites will provide new applications for both civilian and military uses. For instance, the RADARSAT Constellation is designed as a scalable constellation of three small satellites. With a constellation, the time between successive imaging of the same part of the Earth (revisit time) is significantly reduced. The creation of a three-satellite constellation will increase the frequency of available information, as well as the reliability of the system, making it better suited to operational requirements. The RADARSAT Constellation will provide all-weather day and night data in support of maritime surveillance, disaster management, and ecosystem monitoring. In addition, these satellites will continue to provide images for research on information extraction and image fusion techniques.

In this chapter, operational examples in agriculture, geology, geohazards, ice, oil spills, and flood monitoring are discussed. Several emerging applications such as forestry, ship detection, and others are not addressed. Table 2 summarizes the numerous applications of radar images, some mature and others emerging. Although this table was prepared for RADARSAT applications, it represents most of the other radar satellite applications. This summary (Table 2, Van der Sanden and Ross 2004) also facilitates the identification of application fields and information extraction techniques that require further research and development.

Agriculture

The development of key applications for spaceborne SAR data in agriculture has been hindered by the limited information content of single-date, single-frequency, and single-polarization images such as available from first-generation SAR satellites. Some successes were achieved through the application of multitemporal SAR data sets, but images from optical satellites have remained the data source of choice. Optical sensors such as Landsat Thematic Mapper and SPOT are used extensively for crop inventory. However, because of cloud cover, gaps in optical image acquisition can cause inaccuracies in crop classification. In Canada, for instance, mid- to late-summer season optical images are essential for accurate crop classification that is used for crop acreage estimates. A combination of radar images acquired during the cloudy periods and optical images acquired during the cloud-free periods can

Table 1 Current and future radar satellite missions

Satellite	ERS-1	ERS-2	RADARSAT-1	JERS-1	Envisat	RADARSAT-2	AIOS	TerraSAR-X	COSMO-SkyMed	TanDEMX (TDX)	Sentinel	RADARSAT Constellation Mission (RCM)
Space agency	ESA	ESA	CSA	JAXA	ESA	CSA/MDA	JAXA	DLR/ Infoterra GmbH	ASI	DLR/ Astrium	ESA	CSA
Launch	1991	1995	1995	1992	2002	2005	2004	2006	2005	2010	2013	2015
Out of service since	2000			1998			2011					
Band	C	C	C	L	C	C	L	X	X	X	C	C
Wavelength (cm)	5.7	5.7	5.7	23.5	5.7	5.7	23.5	3	3	3	5.4	5.4
Polarization	VV	VV	HH	HH	HH/ VV	QUAD-Pol	All	All	HH/VV	All	Dual (VV VH, HH HV)	HH,VV,HV,VH, compact polarimetry
Incidence angle (°)	23	23	20–50	35	15–45	10–60	8–60	15–60	Variable	20–55	20–45	10–60
Resolution range (m)	26	26	10–100	18	30–150	3–100	7–100	1–16	1–100	1–16	5–40	1–100
Resolution azimuth (m)	28	28	9–100	18	30–150	3–100	7–100	1–16	1–100	1–16	5–40	3–100
Scene width (km)	100	100	45–100	75	56–400	50–500	40–350	5–100 (up to 350)	10–200 (up to 1,300)	5–100	20–400	5–500
Repeat cycle (days)	35	35 (3)	24	44	35	24	24–6	2–11	5–16	2–11	12	5–14
Orbital elevation (km)	785	785	798	568	800	798	660	514	619	514	693	592

Table 2 Application potential of RADARSAT-1 and RADARSAT-2[a] (van der Sanden and Ross 2004)

Application		Satellite	
		RADARSAT-1	RADARSAT-2
Agriculture			
	Crop type	–/+	++
	Crop condition	–/+	++
	Crop yield	–	–/+
Cartography			
	DEM interferometry	+	+
	DEM stereoscopy	++	++
	DEM polarimetry	N.A.	–/+
	Cartographic feature extraction	+	++
Disaster management			
	Floods	++	++
	Geological hazards	–/+	+
	Hurricanes	+	+
	Oil spills	+	+
	Search and rescue	–/+	+
Forestry			
	Forest type	–/+	–/+
	Clear-cuts	–/+	+
	Fire scars	–/+	+
	Biomass	–	–
Geology			
	Terrain mapping	–/+	+
	Structure	+	++
	Lithology	–/+	–/+
Hydrology			
	Soil moisture	–/+	+
	Snow	–/+	+
	Wetlands	–/+	+
Oceans			
	Winds	+	++
	Ships	+	++
	Waves	–/+	–/+
	Currents	–/+	+
	Coastal zones	–/+	+
Sea and land ice			
	Sea ice edge and ice concentration	+	++
	Sea ice type	+	+
	Sea ice topography and structure	–/+	++
	Icebergs	–/+	+
	Polar glaciology	–/+	+

Key: – minimal, –/+ limited, + moderate, and ++ strong

[a] Use of single-date images assumed

significantly improve crop classification accuracies (McNairn et al. 2002, 2009; Moran et al. 2004)

The polarization diversity offered by current radar satellites such as RADARSAT-2 can be expected to enhance the potential of spaceborne SAR for application to agriculture. The introduction of systems operating with longer wavelengths will also help in mapping high-biomass crops (e.g., corn) and related soil properties. Figure 1a shows an example of a RADARSAT-2 polarimetric classification of temperate crops near Ottawa, Canada (McNairn et al. 2009). A similar polarimetric classification was done for tropical crops in Guyana (Fig. 1b). Both these examples show the improvement made in recent years on using polarimetric SAR images for crop classification. In the near future, this classification will significantly improve crop inventory and land use mapping in cloudy tropical areas.

Soil moisture information also supports agriculture and hydrological management such as drought and flood prediction, crop irrigation scheduling, pest management, and others. The monitoring of soil moisture by means of spaceborne SAR has been studied extensively but has to become a viable operational application. The effect of soil roughness and the limited penetration depth, in both soil and potentially overlying vegetation, of C-band SAR systems is a major issue. Systems operating at lower frequencies, for example, L-band, have the capability to capture soil moisture information at larger depth in the soil profile.

Geological Applications

There are numerous studies dealing with geologic remote sensing using radar, optical, and thermal images. Many useful case studies are included in the books by Rencz (1999), Sabins (1997), Rivard (2011), and others. This chapter will focus only on new advanced SAR techniques for geological mapping, mineral exploration, and InSAR techniques for monitoring high-risk geohazards.

Recent results have shown that radar images have provided geologists with useful information for geomorphology, geological structure, and rock units (Lowman 1994; Singhroy et al. 1993; Singhroy 1992; Singhroy and Saint-Jean 1999; Saint-Jean et al. 1999). Geological mappings are facilitated by the use of high-resolution stereo SAR images which improve image interpretation techniques for terrain mapping and that some lineament orientations are enhanced by cross-polarized images (Saint-Jean et al. 1999). Polarimetric SAR images are providing specific information on the scattering mechanism of rock units and surficial materials and are particularly useful in mapping arid areas and planetary surfaces (Singhroy and Molch 2004b).

SAR Data Fusion for Mineral Exploration

The fusion of RADARSAT and other geophysical images in support of geological mapping and mineral exploration is outlined in this section. Singhroy et al. (1993), Singhroy (1996), Singhroy and Molch (2004b), Rivard et al. (1999), and others have

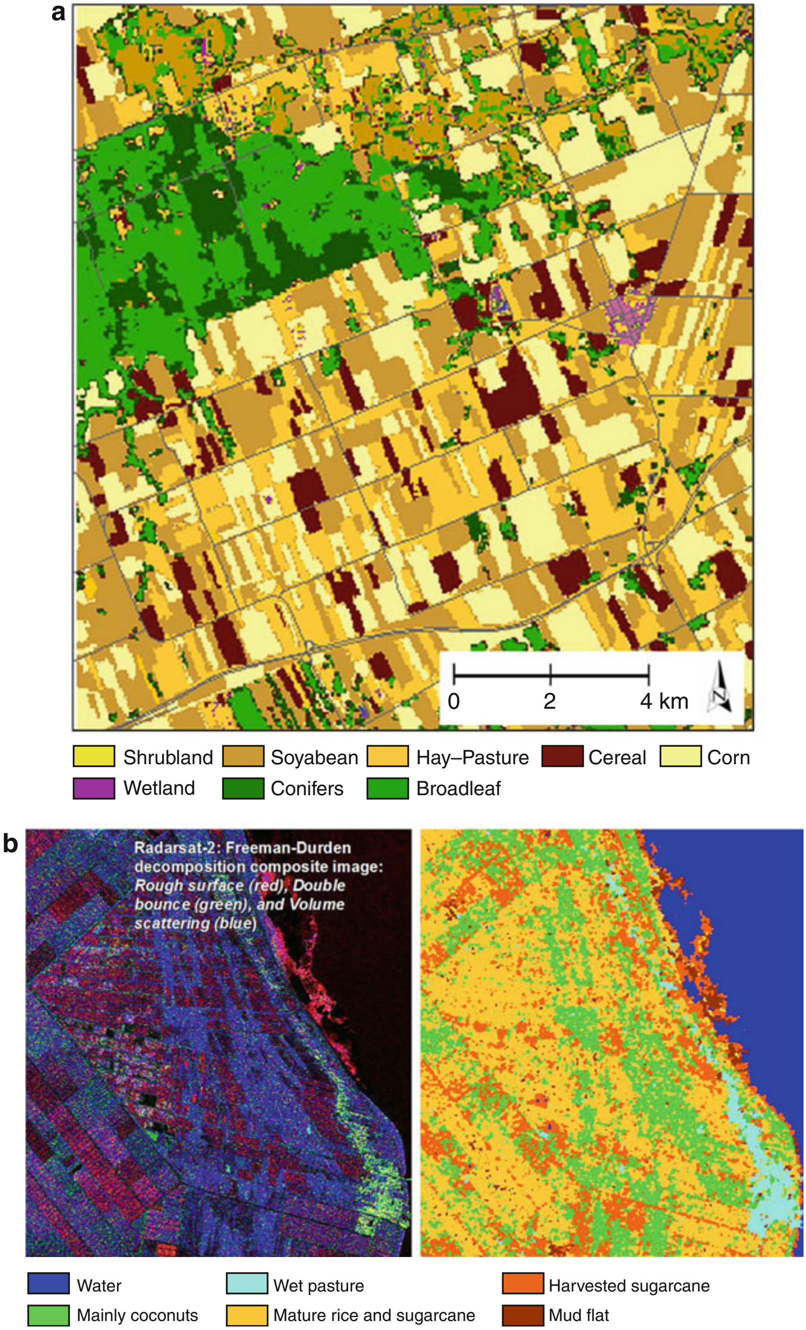


Fig. 1 RADARSAT-2 polarimetric classification of agricultural crops in (**a**) Ottawa, Canada, and (**b**) Berbice, Guyana

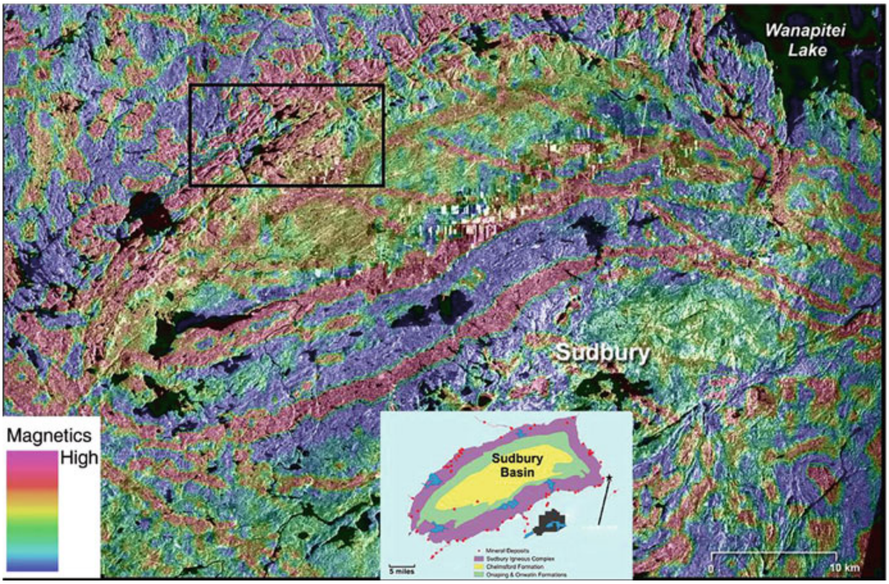


Fig. 2 RADARSAT image fusion map used for mineral exploration in Sudbury, Canada. This map was produced by fusing a RADARSAT standard mode image (25 m resolution) with vertical gradient magnetic image. Nickel, copper, and gold mines are found where high magnetic signatures intersect with faults shown as *linears*

developed interpretation techniques for RADARSAT-fused images to facilitate mineral exploration programs in Canada. This technique involves fusing RADARSAT images with vertical gradient magnetic, radiometric, optical, and topographic data, to produce thematic exploration image maps for exploration and geological mapping programs.

An example of RADARSAT-fused images is shown for the Sudbury Basin in Canada (Fig. 2). The Sudbury Basin is one of the world's richest mining areas and the oldest, largest, and best-exposed meteorite impact site. Current production of nickel, copper, and gold is about $2 billion a year. Figure 2 shows a standard RADARSAT-1 image (25 m resolution) integrated with magnetic data. The standard beam mode of RADARSAT, at 20–27° incidence angles, provides an excellent view of the topography and structural features of the Sudbury area. The geological structure and rock types, which control mineralization, are clearly shown on the fused SAR image. Areas in red outline the strong magnetic signature of the Sudbury Igneous Complex where most of the mines are found. NE-SW magma-filled dikes also have highly magnetic signatures. The other colors represent different rock types with different magnetic signatures.

Geological Hazards

Geological hazards such as earthquakes, volcanoes, and landslides are the main natural causes of damage to human settlements and infrastructures killing thousands every year. As population increases and habitation on hazardous land areas becomes more common, the risks posed by geohazards increase. The need to observe their behavior, understand them better, and mitigate their effects becomes ever more urgent (IGOS 2004).

Current state of the art in real-time monitoring of active geohazard areas developed for early warning is very expensive. Satellite radar interferometry (InSAR) is used to complement real-time monitoring such as GPS, seismometers, in situ field measurements (Singhroy 2009), and others. Provided coherence is maintained over longer periods, it is possible to measure surface displacement of a few centimeters per year. Using InSAR data with small perpendicular baselines, short time intervals between acquisitions, and correcting the effect of topography on the interferogram, reliable measurements of surface displacement are achieved.

Landslides: Interferometric synthetic aperture radar (InSAR) techniques are being used to measure small millimeter displacement on slow-moving landslides. An interferogram represents the phase differences between the backscatter signals in two or more SAR images obtained from similar positions in space. In the case of spaceborne SAR, the images are acquired from repeat-pass orbits. The phase differences between two repeat-pass images are the result of changes in topography; changes in the line-of-sight distance (range) to the radar, due to displacement of the surface; and change in the atmospheric conditions between scenes. On nonmoving target, the phase differences can be converted into a digital elevation model if very precise satellite orbit data are available. Landslide movements are measured in millimeters to centimeters per orbit cycle of the radar satellite. This orbit cycle can range from 44 days for ALOS, 24 days for RADARSAT-2, and a minimum of 2–5 days for TerraSAR-X and COSMO-SkyMed (Table 1). Figure 3 shows an example of InSAR technique for monitoring the Frank Slide in Canada. The Frank Slide is an active 30 × 106 m^3 rockslide-avalanche of Paleozoic limestone, which occurred in April 1903 from the east face of Turtle Mountain in southern Alberta, Canada. Seventy fatalities were recorded (Cruden and Hungr 1986). Detailed analysis on the use of SAR images for characterizing and monitoring the Frank Slide is reported by Singhroy and Molch (2004a) and Singhroy et al. (1998). Our results show that this landslide has moved about 150 mm from 2008 to 2011, using RADARSAT-2 InSAR analysis.

Earthquakes: Plate tectonics provides a framework for understanding earthquake activity. The earthquake-prone regions of Earth are well delineated and global seismicity information is readily available. However, there is a need to improve the identification and characterization of seismically active zones. Satellite radar interferometry is in the early stages in monitoring earthquake deformation aimed at producing more accurate hazard maps (IGOS 2004). Recently, Zhang et al. (2011) and others have used InSAR to monitor coseismic events and generated vertical displacement maps related to the 2008 Sichuan magnitude 7.5 and the 2010 Haiti magnitude 7.0 earthquakes. In addition, high-resolution optical data are widely used for earthquake damage assessment (Ajmar et al. 2011). Figure 4 shows the uses of RADARSAT-2 polarimetric images to map flooded areas in the urban coastal zones

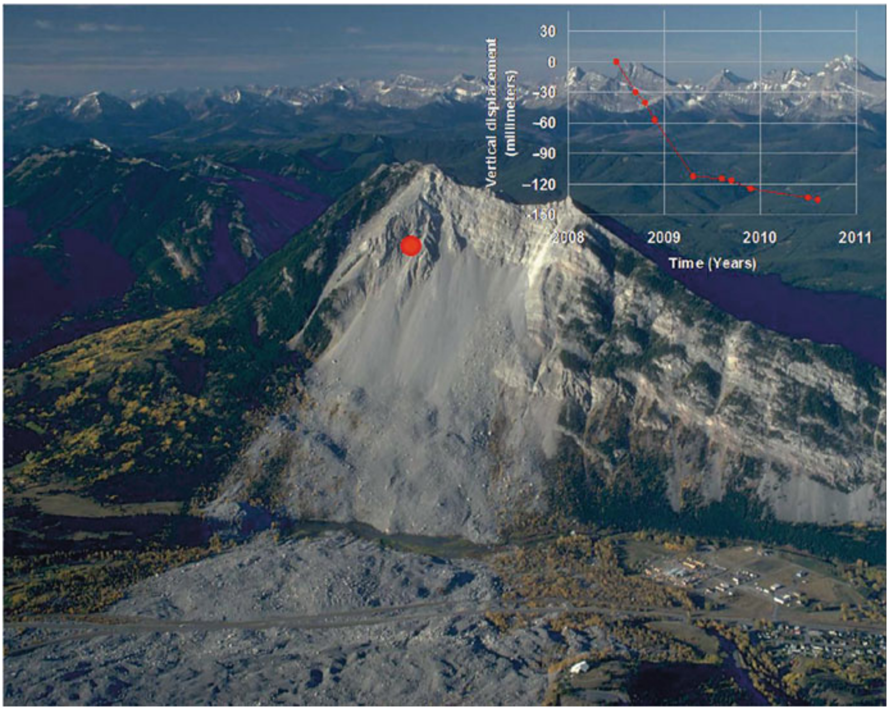


Fig. 3 Turtle Mountain, rock avalanche occurred in 1903, killing 70: RADARSAT-2 InSAR measurements show that the landslide is still active and has moved 150 cm since 2008

and adjacent rice fields. These areas were severely damaged from the 2011 magnitude 9 earthquake-triggered tsunami near Sendai, Japan.

Volcanoes: Volcanic hazards vary from one volcano to another and from one eruption to the next. The most damage is from pyroclastic flows and lahars. Civil authorities need clear information on detectable changes, such as ground cracking, associated earthquakes, and SO_2 emissions. Recently InSAR techniques have been used to monitor the deformation activity of active volcanoes for risk assessment to populated areas. Figure 5 shows the uses of D-InSAR techniques for post-eruption shrinkage rates at the Miyake-jima volcano, after the severe August 2000 eruption (Singhroy et al. 2004).

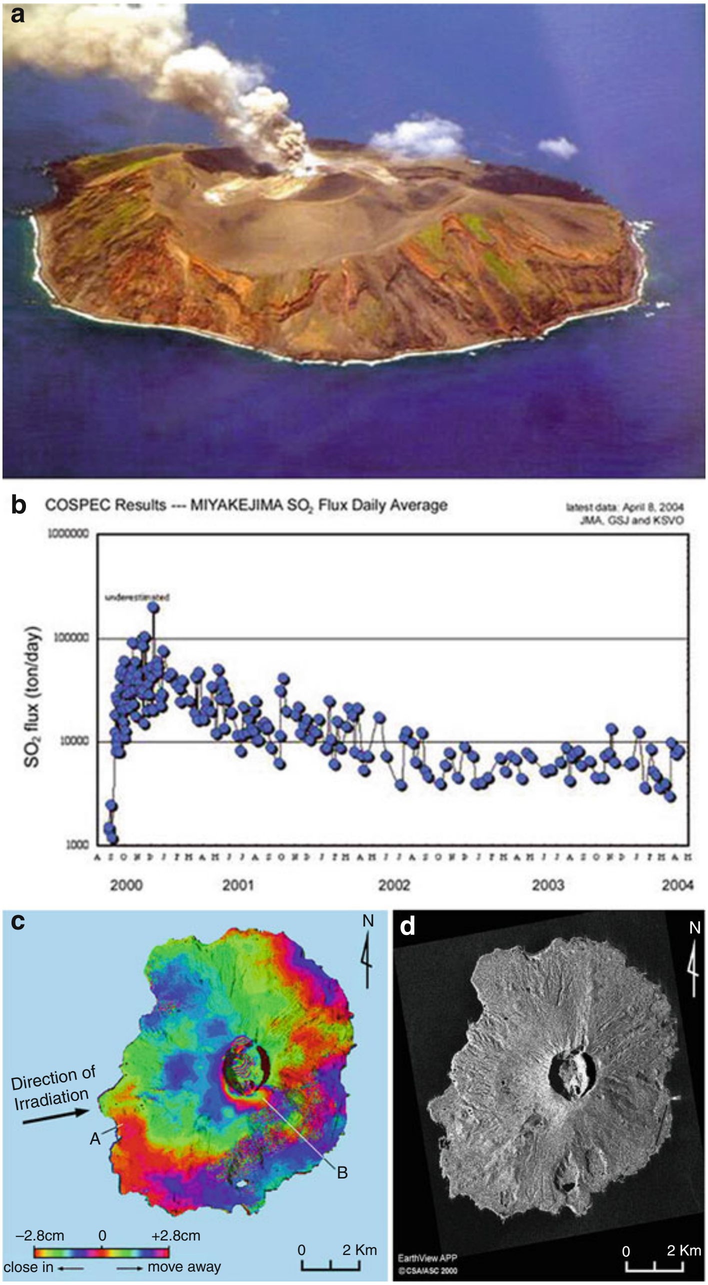
Flood Monitoring

Floods are among the most devastating natural hazards in the world, claiming more lives and causing more property damage than any other natural hazards. In 2011 alone flood devastations were experienced in Pakistan, China, Australia, Thailand, and Canada. Within the USA, an average of more than 225 people are killed, and more than US$3.5 billion in property is damaged by heavy rain and flooding each year

Fig. 4 Sendai, Japan, tsunami damage assessment image map showing flooded areas in urban coastal areas and adjacent rice fields. This map was produced from a polarimetric classification of RADARSAT-2 image taken soon after the 2011 magnitude 9 earthquake

(CEOS 2003). Satellite-derived flood maps produced in near real time are therefore invaluable to local and national agencies for disaster monitoring and relief efforts.

Satellite optical observations of floods are hampered by the presence of clouds, which normally prevent near real-time data acquisitions. It is well known that SAR is an excellent tool, which supports the monitoring of floods. The SAR data are not only restricted to flood mapping but can also be useful to the estimation of a number of hydrological parameters. For instance, radar images were used operationally to map flood extent and wet snow, to monitor wetlands, and to identify freshwater ice



RADARSAT D-InSAR 2000 11/10 $ 2000 12/04
INTERFEROGRAM

RADARSAT 42A F3f 2000 11/10
MASTER IMAGE

Fig. 5 Differential interferometric (D-InSAR) technique showing post-eruption shrinkage rates at the Miyake-jima volcano, Japan, after the severe August 2000 eruption (Singhroy et al. 2004)

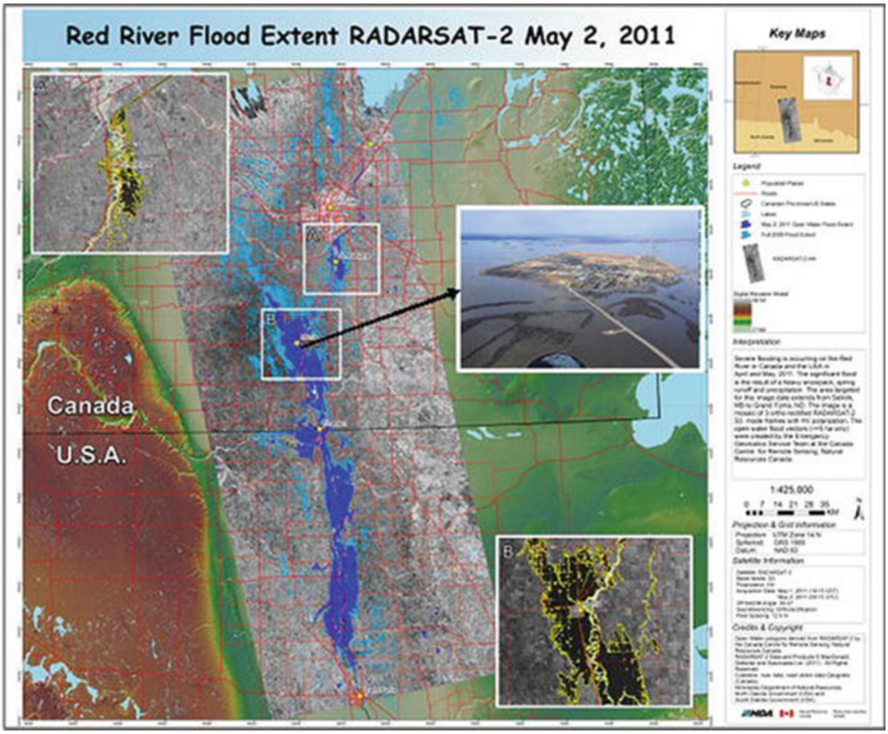


Fig. 6 RADARSAT-2 flood extent image map used in support of flood response programs during the 2011 Winnipeg spring floods in Canada. The radar image was combined with other GIS data such as land use, slope, aspect, and transportation and infrastructure networks

types (Pultz et al. 1997). Figure 6 shows an example of a RADARSAT-2 flood extent image map used in support of flood response programs during the 2011 Winnipeg spring floods in Canada. In this case, the RADARSAT-2 images were combined with other GIS data such as land use, slope, aspect, and transportation and infrastructure networks. The analysis was performed on the high-risk areas where decisions were made to evacuate populated areas affected by the flooding. The integration of GIS data layers with current satellite imagery requires time and effort, but also renders this type of product shown in Fig. 6 as valuable for disaster mitigation and prevention and response purposes.

Oil Spill Monitoring

Radar is ideally suited for the detection of sea surface slicks. Ocean wave properties, particularly capillary waves, are significantly dampened by the presence of surfactants with higher viscosities than water. Wave roughness is reduced, and radar backscatter through Bragg scattering is significantly lowered, causing a large contrast between the surrounding ocean and the area affected by oil.

Hydrocarbons, however, are the main target of interest for detection and monitoring purposes as their presence may be related to natural seep, indicating a subsurface hydrocarbon source or reservoir. These locations are of interest for petroleum exploration and for understanding the hydrocarbon basins. The second important application is for man-made oil spills. These can be deliberate acts of polluting waterways by ships discharging bilge oil or accidental leaks from ships or undersea pipeline infrastructure.

Significant benefits have been realized worldwide by transnational petroleum companies in petroleum producing basins (e.g., the Gulf of Mexico, the North Sea, and the NE coast of Brazil) using this technology. Figure 7a shows an example of RADARSAT-2 images to detect the 2010 BP oil spill in the Gulf of Mexico.

To prevent illegal dumping at sea, pollution monitoring and enforcement have relied on spaceborne SAR by several countries. The inset at Fig. 7b shows ships (white dots) discharging bilge oil near the Indian Ocean. Currently, legislation has been set up to allow coastal states to inspect all shipping within territorial waters and also to ensure that national legislation preventing any dumping applies equally to national- and foreign-owned shipping.

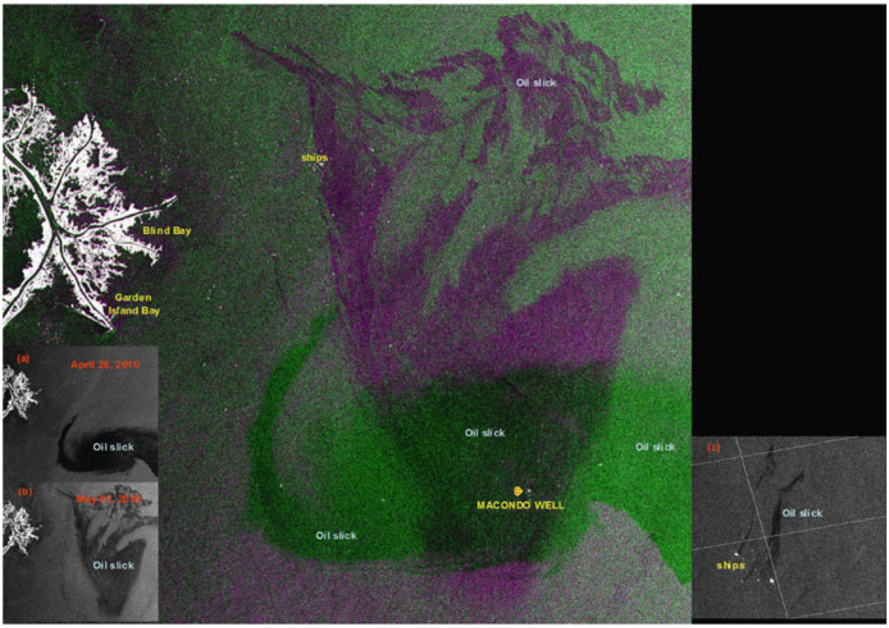


Fig. 7 (**a**) RADARSAT-2 image used to detect the 2010 BP oil spill in the Gulf of Mexico. The *insets* at (**a**) and (**b**) show oil slicks for April 28 and May 1, 2010. The colors represent the oil slicks during these two dates. The *white dots* show a large number of ships involved in the cleanup efforts. (**c**) Detection of illegal dumping of bilge oil from ships in the Indian Ocean

Research into combining information from multifrequency and multitemporal SAR offers an opportunity to advance the application of SAR for oil spill detection and classification.

Sea Ice Monitoring

In the polar regions, sea ice varies both spatially and temporally due to high variability in the environmental processes. National ice service agencies normally produce charts and maps showing current sea ice conditions. These maps must also show the location of the ice edge, concentration and distribution, stage of development, floe size, amount of pressure ridging or topography, location and orientation of ice openings, degree of ice compaction and divergence, and stage of decay during the summer melt season. Ideally, vessels at sea prefer to receive high-resolution satellite images that are less than 6 h old and have been interpreted to provide the above information necessary to avoid or exploit the ice (Skriver and Pedersen 1995; CEOS 2003). It is also well known that sea ice and icebergs pose a serious hazard to shipping and other maritime activities in these regions. Because of their high-resolution, all-weather, wide-swath, and ice detection capability, radar images are particularly useful for monitoring of sea ice conditions in the waters surrounding Canada and other polar regions.

The Canadian Ice Service (CIS) is using RADARSAT images to quantify various ice parameters. These include the mapping of ice extent (concentration) and ice type (stage of development), such as ice topography, the presence of open water or thin ice openings within the sea ice pack, stages of ice decay, and others (Johnson and Timco 2008). Maps of sea ice derived in real time from these data are used operationally to ensure safety of navigation by all vessels operating in the Canadian north. The data products of choice for the Canadian Ice Service, for example, are those that provide wide-area coverage (i.e., ScanSAR products). Figure 8 is an example of a September 2011 RADARSAT-2 ScanSAR image map of the sea ice condition near Ellesmere Island in the Canadian Arctic. These image maps are not only used for navigation safety, but to monitor the ice conditions that are being affected by climate change.

River Ice Monitoring

River ice governs the winter regime of most Canadian rivers and influences both the natural and human environment in various ways. The impact of ice covered on Canadian rivers typically peaks when breakup events result in the formation of ice jams that can cause structural and/or flood damage to nearby homes, businesses, and infrastructure.

Traditional methods of river ice breakup monitoring include visual airborne observation and riverside viewpoints. These methods offer excellent flexibility in

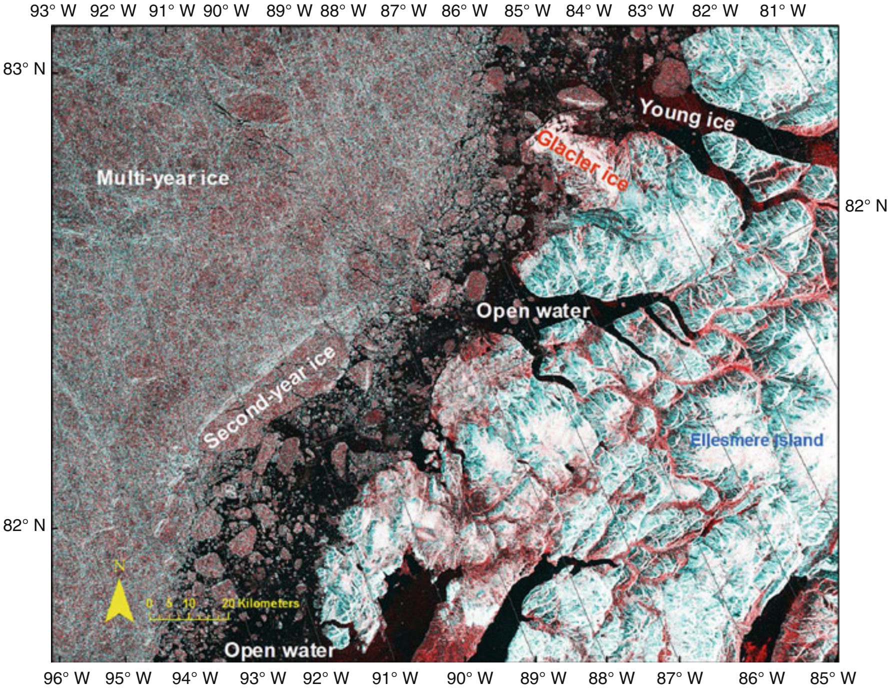


Fig. 8 RADARSAT-2 ScanSAR C-HH image map of the sea ice types and condition near Ellesmere Island in the Canadian Arctic. The image map is used for navigation safety and to monitor the ice conditions that are being affected by climate change

terms of observation timing and frequency but are expensive and provide limited spatial coverage. Satellite remote sensing systems make potentially outstanding tools for collecting current information on river ice because of its systematic, synoptic, and repetitive imaging capability. Synthetic aperture radar (SAR) satellites such as Canada's RADARSAT-2 are particularly well suited to the task because they can image independent of daylight and weather conditions. However, SAR satellites operate in certain orbits that limit their imaging capabilities in terms of timing and frequency. As such, SAR satellite and traditional observation must be considered as complementary rather than conflicting sources of information.

The utility of SAR satellites for the monitoring of ice cover breakup in rivers can be explained from the sensitivity of radar sensors to differences in the surface roughness of the ice surface. In addition, SAR sensors offer excellent sensitivity to the presence of open water. The map subsets shown in Fig. 9 discriminate between "water" and three classes each for ice cover conditions labeled as "sheet ice" and "rubble ice." Sheet ice covers are characterized by smooth textures, whereas rubble ice covers have rough textures. The features of most interest from the emergency management perspective, that is, ice jams, are a form of rubble ice with rough to very

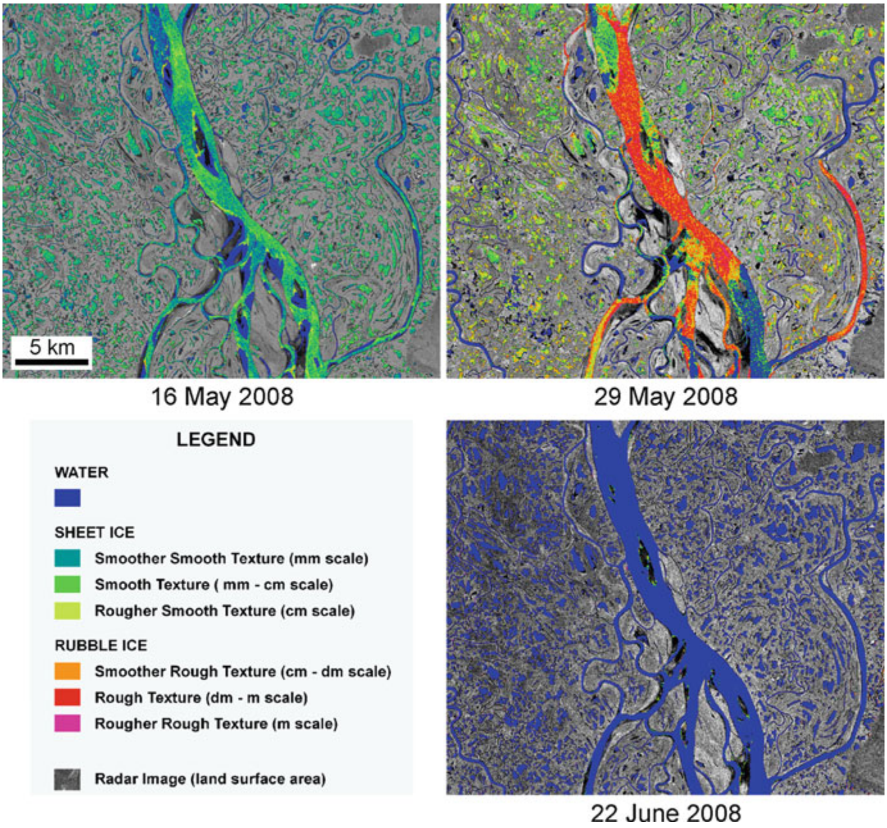


Fig. 9 RADARSAT-1-derived ice cover condition product subsets showing the progression of the 2008 spring breakup in the Mackenzie Delta just north of the town of Tsiigehtchic in Canada's Northwest Territories (van der Sanden et al. 2012)

rough textures. Generation and interpretation of a product time series are essential to determine with certainty whether an ice cover that exhibits the texture of an ice jam is actually jammed and represents a risk or is in fact moving downstream and therefore of little concern.

Conclusion

It is clear that radar images are being used routinely for a number of operational applications. Emerging applications are being developed from current and future advanced SAR missions involving their high resolution, rapid revisits, and polarimetric capabilities. There is a lack of understanding with regard to the relationship between the image information content and the polarization of the radar signal. There have been significant developments on the uses of both dual and fully polarimetric data in some applications, for example, in forestry. However, it is clear that future research should continue to explore the additional uses of fully polarimetric and compact polarimetric data that will be available from future constellations such as the CSA-RADARSAT and ESA Sentinel programs. In addition, significant developments have also been made on using InSAR techniques for geohazards and disaster mitigation. New 3D InSAR products and polarimetric InSAR techniques needed further evaluation.

References

A. Ajmar, P. Boccardo, F. Tonolo, Earthquake damage assessment based on remote sensing data. The Haiti case study. Italian J. Remote. Sens. **43**(2), 123–128 (2011)

CEOS, *The Use of Earth Observing Satellites for Hazard Support: Assessment and Scenarios NOAA* (Natural Resources, Canada, 2003), 147p

D.M. Cruden, O. Hungr, The debris of Frank slide and theories of rockslide-avalanche mobility. Can. J. Earth Sci. **23**, 425–432 (1986)

IGOS Geohazard Theme Report, *IGOS Geohazard Secretariat* (BRGM, Orleans, 2004), 50 p

M.E. Johnson, G.W. Timco, *Understanding and Identifying Old Ice in Summer*. Publication # CHC-TR-055. (Canadian Hydraulics Centre, Ottawa, 2008), 236 p

P. Lowman, Radar geology of the Canadian shield – a 10 year review. Can. J. Remote. Sens. **20**(3), 198–209 (1994)

H. McNairn, C. Duguay, B. Brisco, T. Pultz, The effect of soil and crop residue characteristics on polarimetric radar response. Remote Sens. Environ. **80**, 308–320 (2002)

H. McNairn, C. Champagne, J. Shang, D. Holmstrom, G. Reichert, Integration of optical and synthetic aperture radar (SAR) imagery for delivering operational annual crop inventories. J. Photogramm. Remote. Sens. **64**(5), 434–449 (2009)

S.M. Moran, C.D. Peters-Lidard, J.M. Watts, S. McElroy, Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. Can. J. Remote. Sens. **30**(5), 805–826 (2004)

T. Pultz, Y. Crevier, R. Brown, J. Boisvert, Monitoring of local environmental conditions with SIR-C/X-SAR. Remote Sens. Environ. **59**(4), 248–255 (1997)

A.N. Rencz, *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*, 3rd edn. (Wiley, New York, 1999), 700 p

L.A. Rivard, *Satellite Geology and Photo- Geomorphology* (Springer, Berlin, 2011), 270 p

B. Rivard, L. Corriveau, L.B. Harris, Structural reconnaissance of deep crustal orogen using RADARSAT and landsat satellite imagery and airborne geophysics. Can. J. Remote. Sens. **25** (3), 258–267 (1999)

F. Sabins, *Remote Sensing: Principles and Interpretations*, 3rd edn. (Freeman, New York, 1997), 432 p

R. Saint-Jean, V. Singhroy, M. Rheault, Understanding multi-polarized airborne C-SAR images for geological mapping in precambrian shield terrains. in *Proceedings Applied Geologic Remote Sensing, ERIM 13th International Conference*, vol. 1, Vancouver, 1999, pp. 411–418

V. Singhroy, Radar geology, techniques and results. Episodes **15**(1), 15–20 (1992)

V. Singhroy, Environmental and geological site characterization in vegetated areas: image enhancement guidelines, in *Remote Sensing and GIS: Applications and Standards*, ed. by V. Singhroy, D. Nebert, A. Johnson. ASTM Special Technical Publication No. 1279 (American Society for Testing and Materials, West Conshohocken, 1996), pp. 5–17

V. Singhroy, Chapter 15: Remote sensing techniques for geological mapping and exploration, in *Geoinfomatics for Natural Resource Management*, ed. by P.K. Joshi, P. Pani, S.N. Mohapatra, T.P. Singh (Nova Science, New York, 2009), pp. 333–349

V. Singhroy, K. Molch, Characterizing and monitoring rockslides from SAR techniques. Adv. Space Res. **33**, 290–295 (2004a)

V. Singhroy, K. Molch, Geological applications of RADARSAT-2. Can. J. Remote. Sens. **30**(6), 893–902 (2004b)

V. Singhroy, R. Saint-Jean, Effects of relief on the selection of RADARSAT-1 incidence angle for geological applications. Can. J. Remote. Sens. **25**(3), 211–217 (1999)

V. Singhroy, R. Slaney, P. Lowman, J. Harris, W. Moon, RADARSAT and radar geology in Canada. Can. J. Remote. Sens. **14**(4), 329–332 (1993)

V. Singhroy, K. Mattar, L. Gray, Landslide characterization in Canada using interferometric SAR and combined SAR and TM images. Adv. Space Res. **2**(3), 465–476 (1998)

V. Singhroy, H. Okhura, K. Molch, R. Couture, monitoring landslides and volcanic deformation from InSAR. in *Proceedings ISPRS Congress*, Istanbul, CD Paper, 2004, pp. 570–74

H. Skriver, L.T. Pedersen, Polarimetric signatures of sea ice in the Greenland Sea, in *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS '95*, Florence, 1995, pp. 1792–1794

J. Van der Sanden, S. Ross, *Applications Potential of RADARSAT 2- A Preview: CCRS*, Earth Science Sector Contribution Series 20043000 (Natural Resources, Canada, 2004), 86p

J.J. van der Sanden, T. Geldsetzer, N. Short, B. Brisco, *Advanced SAR Applications for Canada's Cryosphere (Freshwater Ice and Permafrost) – Final Technical Report for the Government Related Initiatives Program (GRIP)*, Ottawa, Natural Resources Canada, ESS Contribution Number 20120212, 2012, 80 p

G. Zhang, C. Qu, X. Shan, X. Song, G. Zhang, C. Wang, J. Hu, R. Wang, Slip distribution of the 2008 Wenchuan, Mag 7.9 earthquake by joint inversion from GPS and InSAR measurements; a resolution test study. Geophys. J. Int. **186**(1), 207–220 (2011)

LiDAR Remote Sensing

Juan Carlos Fernandez Diaz, William E. Carter, Ramesh L. Shrestha, and Craig L. Glennie

Contents

J.C.F. Diaz (✉)
NSF National Center for Airborne Laser Mapping (NCALM)/Department of Civil and Environmental Engineering, University of Houston, Houston, TX, USA

University of Houston, Houston, TX, USA
e-mail: jfernandezhon@yahoo.com; jfernan4@central.uh.edu

W.E. Carter • R.L. Shrestha • C.L. Glennie
NSF National Center for Airborne Laser Mapping (NCALM)/Department of Civil and Environmental Engineering, University of Houston, Houston, TX, USA
e-mail: carter4451@bellsouth.net; rlshrestha@uh.edu; clglenni@central.uh.edu

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_44

Abstract

Light detection and ranging (LiDAR), also known as laser detection and ranging (LaDAR) or optical radar, is an active remote sensing technique which uses electromagnetic energy in the optical range to detect an object (target), determine the distance between the target and the instrument (range), and deduce physical properties of the object based on the interaction of the radiation with the target through phenomena such as scattering, absorption, reflection, and fluorescence. LiDAR has many applications in the scientific, engineering, and military fields. LiDAR sensors have been deployed at fixed terrestrial stations, in mobile surface and subsurface vehicles, lighter-than-air crafts, fixed and rotary wing aircraft, satellites, interplanetary probes, and planetary landers and rovers. This chapter provides a high-level overview of the principles of operation of LiDAR technology and its main applications performed from space-based platforms such as satellite altimetry, atmospheric profiling, and on-orbit imaging and ranging.

Keywords

Active remote sensing • Atmospheric • Bathymetry • CALIOP • CALIPSO • DIAL • Differential absorption LiDAR • Doppler LiDAR • Fluorescence LiDAR • GLAS • ICESat • International Laser Ranging Service • Ladar • Laser altimeter • Laser detection and ranging • Laser remote sensing • LiDAR • Light detection and ranging • LLR • Lunar laser ranging • OBSS • Optical radar • Raman LiDAR • Satellite laser ranging (SLR) • Scattering LiDAR

Introduction

This chapter provides a description of light detection and ranging (LiDAR) as an active remote sensing technique. LiDAR has evolved over the past seven decades, and as a result, there are many different types of LiDAR systems in use today. Systems can be classified based on the application (atmospheric, mapping, bathymetry, navigation), based on the ranging technique (time of flight, triangulation, phase difference), based on the target detection principle (scattering, fluorescence, reflection), or even based on the platform that the system is deployed on (ground based, mobile terrestrial, airborne, spaceborne, marine, submarine). There are many reference works that cover LiDAR systems from alternative viewpoints. For example, *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere* (Weitkamp 2005) provides an indepth review of modern atmospheric LiDAR techniques, while *Topographic Laser Ranging and Scanning: Principles and Processing* (Shan and Toth 2009) provides a complete review of the main terrestrial mapping LiDAR techniques. In the context of a *Handbook of Satellite Applications*, this chapter provides a high-level overview of LiDAR systems with a focus in those based on spaceborne platforms and their main applications. The chapter starts with a brief historical timeline of the origins of the LiDAR technique; it is followed by a high-level technical overview of the principles of operation and the hardware that constitute a generic LiDAR system; and it concludes with descriptions of the main applications of LiDAR technology to and from spaceborne platforms.

Origins of LiDAR Technology

What we know today as LiDAR is the result of the convergence of efforts by different scientific communities to use visible light sources and detectors to resolve technical or scientific issues. LiDAR was pioneered by atmospheric scientists in the 1930s for the determination of atmospheric density profiles, refined as a way to obtain precise and accurate measurements of distances by geodesists and surveyors in the 1940s and 1950s, and taken to interplanetary distances by physicists studying relativistic effects in the 1960s.

Early proposals for the use of high-power searchlights to study atmospheric density and composition were developed by E. G. Synge in 1930 (Synge 1930) and M. A. Tuve et al. in 1935 (Tuve et al. 1935). Early successful measurements using bistatic systems consisting of a high-intensity searchlight and a telescopic photographic station separated by baselines of 2–18 km were conducted by J. Duclaux in 1936 (Hulburt 1937), E.O. Hulburt in 1937 (Hulburt 1937), and E.A. Johnson et al. in 1939 (Johnson et al. 1939). Using long-exposure photography, the setup by Duclaux was able to trace light scattering up to a height of 3.4 km, and the experiments by Hulburt reached heights of up to 28–30 km (Hulburt 1937). The limit of these photographic techniques was set by the saturation of the photographic film and the contrast between the beam intensity and night sky. An alternative to the saturation of the photographic film was the method proposed by Tuve et al. and implemented for the first time by Johnson et al. which consisted of modulating the intensity of the searchlight and using a photoelectric cell to detect the scattered radiation. The output of the photoelectric cell was amplified by an AC system tuned to the lamp modulating frequency. With this type of electric detection system, Johnson et al. were able to record light scattering to heights of 34 km (Johnson et al. 1939). These early atmospheric LiDAR experiments yielded scattering intensity information as a function of the height, but were not concerned with obtaining accurate range measurements. The need to obtain accurate range (distance) measurements using light beams came from the geodetic science community.

LiDAR as a tool to determine accurate range (distance) measurements for geodetic and surveying applications originated in the late 1930s as a technique named electronic distance measurement or EDM. The development of the first EDM instrument began in 1938 when the physicist and geodesist Erik Bergstrand, of the Swedish Geographical Survey Office, began to investigate the possibilities of using a Kerr cell as an electro-optical shutter to modulate a beam of light in an attempt to better measure the speed of light. Bergstrand's first operational instrument was reported to work in 1941 (Carter 1973). In August 1948, Bergstrand presented a paper at the meeting of the International Association of Geodesy (IAG) held in Oslo, Norway. In that paper, he explained that the process could be reversed and that by measuring the light's time of flight and using the known speed of light, it was possible to accurately compute the distance between the light source and a retroreflector. Soon after that IAG meeting, Bergstrand licensed the distance measuring concept to the Swedish AGA (Svenska Aktiebolaget Gasaccumulator) company to develop a commercial EDM instrument. AGA produced the first EDM instrument in the early 1950s and marketed it as the Geodimeter, short for geodetic distance meter. The instrument used a Kerr cell to modulate the light and a mercury vapor lamp as the light source. Refinement of the Geodimeter by AGA continued through the 1950s and 1960s (Fernandez-Diaz 2007).

During the 1940s and 1950s while Bergstrand was developing the EDM technique, atmospheric scientists continued to build upon the early scattering measurements by using pulsed searchlights. These pulsed light sources enabled the researchers to measure the range to the scattering particles using the time-of-flight principle rather than the original triangulation method. In the book *Meteorological Instruments*, published by W.E.K. Middleton and A.F. Spilhaus in 1953, the acronym LiDAR was coined for this type of time-of-flight technique (Wandinger 2005). Around the same time, a group at Princeton University led by professor R.H. Dickey, working on gravitation research, investigated a concept of using a high-density and high-altitude artificial satellite to measure slow changes in the universal gravitation constant (G) by tracking the satellite orbit using retroreflectors and pulsed searchlights (Bender et al. 1973). This concept incorporated elements of both the atmospheric and geodetic LiDAR research. However, the pulsed light sources and photodetectors available at that time made its implementation impractical. A breakthrough in technology was needed which increased the power and intensity of the light beams.

The breakthrough came in November 1957, when Gordon Gould, a graduate student at Columbia University, coined the acronym LASER, for light amplification by stimulated emission of radiation, and described the principal components of the laser (Taylor 2000). The conceptual invention of the laser was followed by the first successful implementation by Theodore Maiman and his colleagues at Hughes Aircraft Company, who built the first solid-state pulsed laser using a ruby rod in 1960. That same year, Ali Javan and his colleagues from Bell Laboratories succeeded in building the first gas (HeNe) laser (Javan et al. 1961). Another important advancement was the development of Q-switching for ruby lasers in 1961 by F.J. McClung and R.W. Hellwarth, which enabled the generation of short (nanoseconds) laser pulses that packed relatively large amounts of energy (McClung and Hellwarth 1962). The photons produced by a laser are from a very narrow wave band, have very similar phase and polarization, and travel nearly parallel to one another. These attributes make it relatively simple to create a highly collimated beam of light (its divergence is essentially limited by the aperture of the transmitter and the atmosphere) that yields strong returns from even very distant targets.

In May 1962, L.D. Smullin and G. Fiocco were successful in obtaining ruby laser returns from the bare lunar surface (Smullin and Fiocco 1962) and between June and July 1963 obtained atmospheric returns from heights between 60 and 140 km (Fiocco and Smullin 1963). These experiments ignited an exponential development in LiDAR technology in these fields of research. Within the following decade, atmospheric scientists had demonstrated all the basic atmospheric LiDAR techniques in use today (Wandinger 2005).

The physicists and geodesists working on relativity and gravitation obtained the first ruby laser returns from an artificial satellite (Beacon Explorer 22-B) equipped with corner cube reflectors (retroreflectors) on October 31, 1964 (Carter 1973; McGarry and Zagwodzki 2005). This became the origin of what is currently known as satellite laser ranging or SLR, which uses LiDAR to measure ranges from ground stations to satellite-borne retroreflectors with millimeter-level precision and from which it is possible to obtain highly accurate orbits for critical satellites such as GPS, GLONASS, Galileo, Jason, ERS, and others (The International Laser Ranging Service). However, even before the Beacon Explorer was launched, scientists realized that low-orbiting satellites imposed several challenges such as very short visibility times and Earth's gravitational perturbations that would limit the quality of the relativistic experiments. To overcome these limitations, they had proposed the idea of placing retroreflector arrays on the surface of the Moon, which could be used to bounce back a laser beam shot from the Earth. These lunar retroreflector arrays would allow yield better results than the ones obtained by Smullin and Fiocco in 1962 (Smullin and Fiocco 1962) and by Grasyuk et al. in 1964 (Bender et al. 1973), because they would result in "point" returns, with negligible time spread compared to returns from a patch of lunar topography.

On July 21, 1969, during the Apollo 11 mission, Neil Armstrong oriented and leveled the first lunar retroreflector array (LRRR) on the surface of the Moon. The first successful return signals from the LRRR were obtained on August 1, 1969, at Lick Observatory, and on August 20, 1969, at the McDonald Observatory (Alley et al. 1969). Additional retroreflectors arrays were deployed on the Moon by the Apollo 14 and 15 missions, and French-built retroreflectors arrays were deployed by the Soviet Lunokhod 1 and 2 rovers (Dickey et al. 1994). To this date, observatories are still bouncing laser pulses from these retroreflectors in a technique called lunar laser ranging (LLR). This has provided numerous contributions to a number of scientific fields such as gravitational physics, relativity, astronomy, lunar science, geodesy, and geodynamics (Dickey et al. 1994).

Down on Earth, during the 1960s, there was also an exponential development of the EDM technique. In 1967, AGA introduced its Geodimeter Model 8, which was its first to use a helium–neon laser, and doubled the range of the lamp units from

30 to 60 km. Meanwhile, other companies were working on laser-based EDMs with the ability to determine ranges using weak return signals from natural targets rather than from retroreflectors. Examples of these reflector-less EDMs are the instruments manufactured by Spectra Physics such as Mark II and Mark III (Geodolite). These, or similar instruments, were used in the mid-1960s as the first airborne LiDAR profilers and even bathymetric LiDAR systems (Fernandez-Diaz 2007). As lasers with higher pulse rates were developed and scanners of different designs were added to distribute measurements over swaths of terrain, these laser profiling systems evolved into the high-resolution airborne mapping LiDAR systems operational today(Carter et al. 2007).

The first spaceborne LiDAR system was flown onboard the ANNA-1B (Army, Navy, NASA, and Air Force) satellite in 1962, which was a joint project between the agencies to test various satellite tracking techniques including interferometry, Doppler, and strobe lights (Simons 1964). ANNA-1B was equipped with two high-intensity optical beacons that when commanded produced a sequence of five flashes separated by 5.6 s. The flashes were recorded against star fields using stellar cameras (e.g., Wild BC-4 and PC-1000) at ground stations of the Minitrack Optical Tracking System (Harris et al. 1966).

The first spaceborne LiDAR based on a laser transmitter was flown during the Apollo 15 mission in July–August 1971. The Apollo 15 laser altimeter, based on a Q-switched ruby laser, was part of the metric camera system but was also capable of operating independently (Robertson and Kaula 1972). Similar laser altimeter systems were flown on the Apollo 16 and 17 missions in 1972, and their data were used for, among other things, to determine the lunar shape and infer its structure (Kaula et al. 1974). Between 1972 and the 1990s, there was a hiatus in the deployment of spaceborne LiDAR systems, but since 1990, there has been a continuous progression both in terms of numbers and technological development of the deployed systems. Table 1 presents a summary of past, current, and future space-based LiDAR systems. Their principles of operation and applications are described in the following sections.

High-Level Technical Overview of LiDAR

In principle, LiDAR consists of sending out optical energy, observing the interactions between the photons and the target, and measuring the distance between the emitter and the target. At the highest level, a LiDAR system consists of three main subsystems: an optical transmitter, an optical receiver/detector, and ranging/timing/control electronics. The designs of these elements vary greatly among systems and depend upon the targeted application. To help illustrate these concepts, Figs. 1 and 2 show a 3D model and optical diagram of the atmospheric scattering LiDAR (CALIOP) onboard the NASA/CNES CALIPSO satellite.

The optical transmitter is composed of a light source, usually a laser system, and optical elements used to modify (focus, collimate, expand, split) the light beam. The optical detector is comprised of a telescopic-type instrument that collects the

Table 1 Spaceborne LiDAR systems

Launch date	Spacecraft	System/application
October 31, 1962	ANNA-1B	High-intensity optical beacons (Simons 1964; Harris et al. 1966)
July 26, 1971	Apollo 15, Endeavour	Apollo laser altimeter (Robertson and Kaula 1972; Kaula et al. 1974)
April 16, 1972	Apollo 16, Casper	
December 07, 1972	Apollo 17, America	
1982	PANTHER	LORA/laser altimeter (Werner et al. 1995, 1996)
September 25, 1992	STS-64 (Discovery)	Lidar In-space Technology Experiment (LITE) (Winker et al. 1996)
May 20, 1995	Spektr/MIR	BALKAN-1 (Werner et al. 1995)
January 11, 1996	STS-72 (Endeavour)	Shuttle Laser Altimeter 1 (SLA-01) (Garvin et al. 1998)
February 17, 1996	Near Earth Asteroid Rendezvous (NEAR)	NEAR Laser Range finder (NLR) (Colea et al. 1996)
April 23, 1996	Priroda/MIR	l'Atmosphere par LIdar Sur SAliout (ALISSA) (Chanin et al. 1999)
November 7, 1996	Mars Global Surveyor	Mars Orbiter Laser Altimeter (MOLA-2) (Smith et al. 2001)
December 4, 1996	Mars Pathfinder/Sojourner	Microrover Flight Experiment/rover navigation (JPL 1997)
January 1, 1997	ALMAZ-1B	BALKAN-2 (Matvienko et al. 1994)
August 7, 1997	STS-85 (Discovery)	Shuttle Laser Altimeter 2 (SLA-02) (Carabajal et al. 1999)
August 10, 2001	STS-105 (Discovery)	Space Vision Laser Camera System (LCS) (STS-105 Shuttle Press Kit 2001; Piedboeuf et al. 2004)
January 12, 2003	Ice, Cloud, and land Elevation Satellite (ICESat)	Geoscience Laser Altimeter System (GLAS) (Abshire et al. 2005)
August 3, 2004	Mercury Surface, Space Environment, Geochemistry and Ranging (MESSENGER)	Mercury Laser Altimeter (MLA) (Cavanaugh et al. 2007)
April 11, 2005	XSS-11	Spaceborne Scanning Lidar System (SSLS) (Nimelman et al. 2006; Dupuis et al. 2008)/rendezvous and proximity operations.
July 26, 2005	STS-114 (Discovery)[a]	Orbiter Boom Sensor System (OBSS)a (NASA 2005) Laser Dynamic Range Imager (LDRI) (Smithpeter et al. 2000) Laser Camera System (LCS) (Deslauriers et al. 2005)

(continued)

Table 1 (continued)

Launch date	Spacecraft	System/application
April 28, 2006	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO)	Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) (Winker et al. 2004)
August 4, 2007	Phoenix Mars Lander	Meteorological Station (MET) Atmospheric LiDAR (Whiteway et al. 2008)
June 18, 2009	Lunar Reconnaissance Orbiter (LRO)	Lunar Orbiter Laser Altimeter (LOLA) (Ramos-Izquierdo et al. 2009)
August 28, 2009	STS-128 (Discovery)	TriDAR[b] (English et al. 2005; NEPTEC)
~2013	ADM-Aeolus	Atmospheric Laser Doppler Instrument (ALADIN) (Ansmann et al. 2007)
~2014	ICESat-II	Multi-beam laser altimeter (Abdalati et al. 2010; Yua et al. 2010)
TBD	Earth Clouds, Aerosols, and Radiation Explorer (EarthCARE)	Atmospheric backscattering and depolarization LiDAR (ATLID) (Le Hors et al. 2008)
TBD	Deformation, Ecosystem Structure, and Dynamics of Ice (DESDynI) (Donnellan et al. 2008)	L-Band polarimetric InSAR multi-beam laser altimeter
Future NASA	ASCENDS, ACE, LIST, GRACE II, 3DWinds	These missions have been proposed by the NRC decadal survey (National Research Council 2007) and might include LiDAR instruments
Future ESA	WALES, ASCOPE (Durand et al. 2007), BepiColombo (Thomasa et al. 2007)	These are Earth and planetary observation missions under study by ESA that might include LiDAR

[a] The OBSS made its first flight on STS-114 Discovery and has flown on every shuttle mission since

[b] TriDAR had its first space demonstration on STS-128 and was flown again during STS-131 (April 2010)

backscattered photons, spatial and spectral filters that discriminate the specific wavelengths intended to be detected, and an electronic photodetector that can be a simple photomultiplier or photodiode in the case of the mapping LiDAR or as elaborate as an spectrometer in the case of fluorescence or Doppler LiDAR. If the transmitter and the detector systems share the same optical elements, i.e., same optical transmit and receive paths, the system is considered to be monostatic. If the optical transmit and receive paths do not share elements, the system is defined as bistatic. From Figs. 1 and 2, it can be seen that CALIOP is a bistatic system, with a transmitter consisting of two independent lasers located parallel to the receiving telescope. Finally, the ranging/timing electronics enable the LiDAR to determine the distance to the target. In addition, LiDAR systems very often have mechanical, optical, or electronic scanning mechanisms that allow steering the light beam.

The design of a LiDAR system starts with the definition of the purpose or application that the system will serve. The application will dictate which interaction

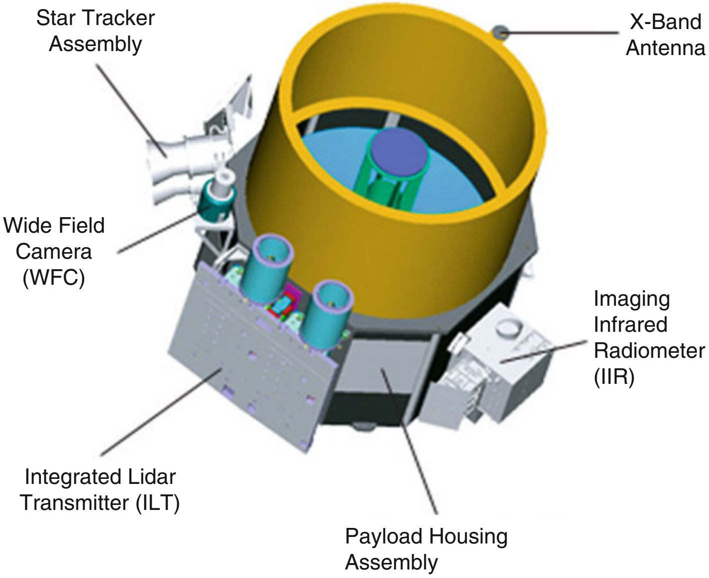
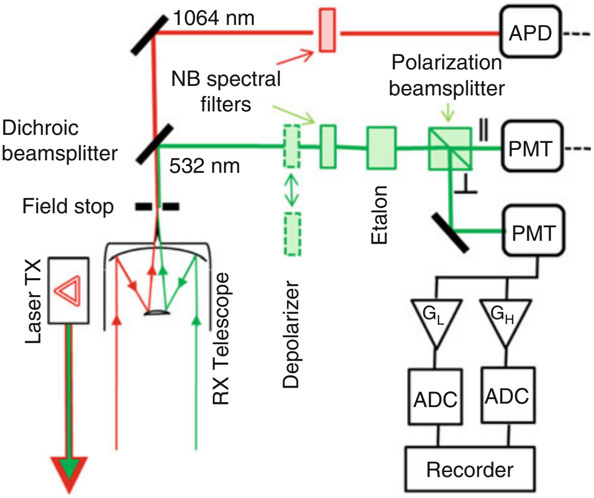


Fig. 1 3D optical model of the CALIOP LiDAR (Image courtesy of NASA)

Fig. 2 Optical diagram of the CALIOP LiDAR



between light and target needs to be detected (scattering, reflection, absorption, etc.) and the most suitable ranging method. The type of interaction between light and target dictates what particular wavelengths can be used and narrows down the light sources that can be selected. From this point, it remains to select the best available photodetector to sense that light–target interaction. To aid the design process, the LiDAR equation is used, which relates the expected received signal strength with sensor parameters such as transmitted optical energy and receiver telescope area, atmospheric parameters such as transmittance and scattering probability at the operating wavelength, and operating conditions such as expected range and target cross section. The following sections provide basic descriptions of the ranging methods, the light–target interaction phenomena, and the light sources and photodetectors that enable the operation of a system. These descriptions cover material that leads to different forms of the LiDAR equation.

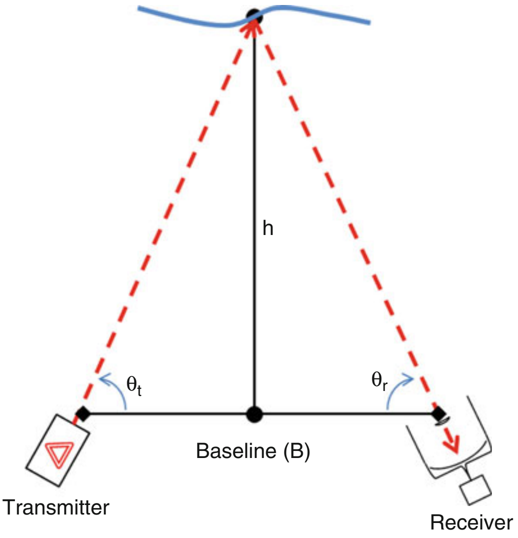
Ranging Methods

There are three main methods that can be used to measure the distance (range) between a LiDAR instrument and the target: optical triangulation, phase difference, and time of flight (TOF). It is also possible to employ hybrid approaches combining two of these methods. Each of these ranging methods has its own set of strengths and weaknesses and range of applicability (English et al. 2005).

Optical Triangulation

Optical triangulation was the ranging method used in the early atmospheric LiDAR experiments of the 1930s. As illustrated in Fig. 3, it is based on the geometry principle that knowing three elements of a triangle, it is possible to determine any other element of the triangle. In the case of the early atmospheric LiDARs, the first element that was known was the separation between the searchlights and the observing station. This leg of the triangle is known as the baseline. The other two

Fig. 3 Triangulation ranging principle



known elements of the triangle were the horizontal angles of the searchlight and photographic station.

Systems based on optical triangulation are ideal for short-range measurements (few meters) yielding micrometer-level precision at high data rates. However, its accuracy depends on the relation between range and baseline distance, and it degrades rapidly with increasing range ($\sim R^2$). It is also limited due to its sensitivity to noise from exterior illumination sources (English et al. 2005).

Phase Difference

Phase difference was the ranging method used in early geodetic EDMs such as the Geodimeter, and it is currently used in some ground-based and airborne mapping systems and on one short-range spaceborne imager. This method consists of modulating the intensity of a continuous wave (CW) laser using a superposition of sinusoidal waveforms with different spatial wavelengths. The range is determined by measuring the phase difference and the number of complete cycles between the emitted and return laser waveform. The main disadvantage with this method is that phase differences are not unique, as there is always an unknown number of complete modulating wave cycles that have occurred prior to the phase difference (phase ambiguity). Compared to the time-of-flight method, the phase difference methods provides higher measurements rates. If there is no a priori knowledge of the range (for geodetic systems), the maximum range of this method is half the spatial wavelength of the carrier frequency, and the range resolution is a function of the highest modulating frequency and the phase difference resolution (English et al. 2005).

Time of Flight (TOF)

The third ranging method uses discrete pulses of light rather than continuous emitting sources. The TOF principle is the simplest, and it consists of measuring the time between when the light pulse is emitted and the detection of a return signal. This two-way travel time (time of flight) is divided in half and multiplied by the speed of light in the respective medium, yielding the range between the instrument and the target. Early LiDARs that used light from lamp sources would create light pulses using optical chopper wheels or capacitive discharge devices (flash lamps). The development of Q-switching by McClung and Hellwarth in 1961 enabled the emission of very energetic laser pulses rather than the continuous wave beams. However, even when these pulses last for a relatively short time, generally in the order of a few nanoseconds, at the high speed that light travels, this translates into several centimeters in length (e.g., 1 ns = 30 cm). In order to obtain sub-centimeter accuracy, the recording and analysis of the entire emitted and return waveform must be performed, or a specialized electronic circuit called a constant fraction discriminator (CFD) can be used on the fly to precisely time a specific point on the waveform (generally the half point of the pulse amplitude at its leading edge). Systems that range to special design retroreflectors may use mode-locked lasers which produce very narrow pulses picoseconds in width.

TOF is the most common ranging method in modern LiDAR, because it provides unambiguous range measurements of distances limited only by the dispersion of the laser energy and the sensitivity of the detector. However, the TOF approach is limited in data collection rate by the laser repetition frequency (PRF) which is the number of laser pulses that can be emitted per second.

Hybrid Systems

Hybrid systems use two above ranging methods, combining the unique capabilities of each to overcome the limitations of a single method. For instance, a hybrid system that employs the triangulation and TOF methods can exploit the advantages of TOF for long ranges and the accuracy and speed of a triangulation system at short ranges (English et al. 2005).

Light–Target Interaction Phenomena

Recall that the “D” in LiDAR stands for detection, the detection of return optical energy backscattered from the target. Detection of a target is possible because there is an interaction between the emitted light energy and the target. There are several types of interactions, which usually depend on the relative size of the target and the wavelength of the radiation. The main interactions between light and matter employed by LiDAR technology are described next.

Scattering

Scattering is the physical phenomenon that occurs when electromagnetic radiation changes its original direction of travel due to interactions with matter in the form of atoms or molecules (Fig. 4). If there is only one particle, a single scattering process is produced. If the photon is scattered several times by different particles, the process is called multiple scattering. These matter and radiation interactions can occur with or without the apparent transfer of energy. In elastic scattering, the photons maintain their wavelength, thus conserving energy. Examples of elastic scattering include Rayleigh and Mie scattering. Inelastic scattering occurs when part of the photon energy is transferred into the scattering particle, thus changing its wavelength. Examples of inelastic scattering include Raman and Brillouin scattering. Based on the relative size of the scattering centers with respect to the wavelength of the radiation, scattering can be classified as Rayleigh scattering when the particles are small compared to wavelength, Mie scattering when the particle size and radiation wavelength are roughly of the same order of magnitude, and geometric scattering when the particles are much larger than the wavelength.

The backscatter component is the radiation that changes direction by approximately 180°, i.e., reverses direction (Fig. 4). Radars and LiDARs detect the backscatter component of the radiation that was emitted. In atmospheric LiDARs, Mie scattering is used to detect aerosols in the troposphere, while Rayleigh scattering is used to detect molecules in the stratosphere and mesosphere. Mapping LiDARs are based on geometric scattering as the targets are much larger than the optical wavelengths.

Reflection

Reflection is a particular type of geometric scattering following particular geometric relationships. There are two limiting theoretical models for reflective surfaces: a specular reflector is one from which incident radiation will be reflected in a single direction (like a mirror) following Snell's law, and a Lambertian reflector surface will spread the reflection over a wider pattern (Fig. 5). These are two limiting cases, and the actual reflection from most surfaces will be between these models. Mapping LiDAR detects reflected radiation from varied targets such as the solid rough surface of a planet (Lambertian behavior), diffuse targets like a forest canopy, or mirror-like surfaces such as a calm lake (specular behavior). An example of LiDARs that are based on specular reflections is those systems used for satellite or lunar laser ranging (SLR and LLR). To achieve extremely long ranges and millimeter-level accuracy,

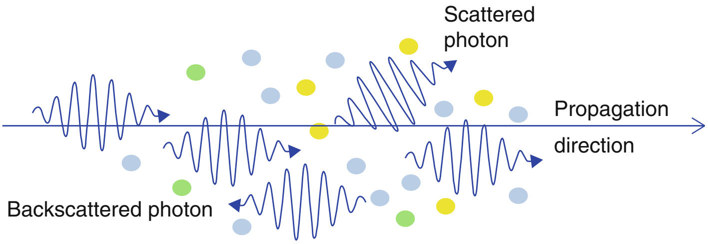


Fig. 4 Photon and matter interaction – scattering

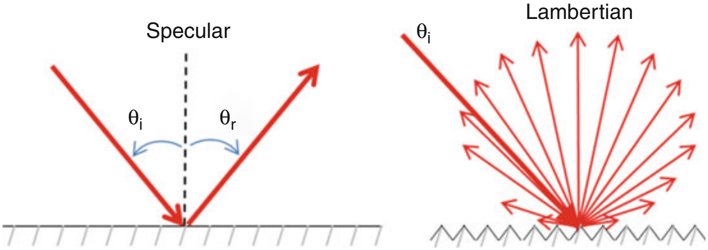


Fig. 5 Specular and Lambertian reflection patterns

corner cube reflectors (retroreflectors) are used to reflect the laser beam in almost exactly the opposite direction (within a few seconds of arc) in which it was emitted.

Absorption

Absorption is another possible result of the interaction of electromagnetic radiation and matter. For a photon to be absorbed, it has to be of a particular wavelength or energy, and because of the principle of conservation of energy, the absorption causes a change in the energy state of the atom or molecule by either an electronic, vibrational, or rotational transition. Differential absorption LiDAR (DIAL) systems compare the received backscattered signal for two or more different laser wavelengths to determine the differential molecular absorption coefficients. If the differential absorption cross sections for each wavelength are known, the concentration of the gas atoms or molecules can be directly deduced. Atmospheric constituents that can be detected by DIAL include ozone and water vapor. DIAL can also be used for industrial emission monitoring and forest fire detection.

Fluorescence

Fluorescence occurs when a molecule absorbs a photon and after a determined period of time emits another photon of the same or longer wavelength. It is considered resonance fluorescence when the emitted photons have the same wavelength of the absorbed photon and normal fluorescence when the emitted photons have longer wavelengths (lower energy). The process of normal fluorescence occurs in three stages: the excitation of the molecule by the incoming photon which happens on a timescale of femtoseconds (10–15 s), vibration relaxation which brings the molecule to a lower excited state and occurs on a timescale of picoseconds order (10–12 s), and emission of a longer wavelength photon and return of the molecule to the ground state which occurs in a relatively long time period of nanoseconds (10–9 s). Fluorescence LiDAR usually emits ultraviolet radiation and observes the reemission of photons in the visible range with a spectrometer detector which records the relative emission at different wavelengths. Applications of fluorescence LiDAR include vegetation studies and the detection of pollutants. For instance, minute amounts of oil in water can be detected because of the UV fluorescence properties of hydrocarbons.

Doppler

The Doppler effect consists of an apparent shift in frequency or wavelength of waves (sound or electromagnetic) as a result of the relative motion between the emitter and the observer. These relative motions can be due to movement of emitter, observer, or medium (in the case of sound waves) or even the simultaneous motion of all three of them. If the relative motion makes the emitter and observer become closer, the wavelength of the wave will appear to get shorter (blue shift), whereas if the distance becomes larger, the wavelength will appear to get longer (red shift). In addition to the well-known frequency shift, the Doppler effect also causes the broadening of spectral line features in a process that is temperature dependent. Turbulence and winds are manifestations of the collective motion of the atmospheric molecules and particles. Light scattered along the line of sight (LOS) of the propagating laser beam will experience Doppler shifts and linewidth broadening due to the relative motion of the atmospheric elements with respect to the LiDAR system and due to changes in atmospheric temperature. Thus, Doppler LiDAR is applied to determine air temperature, wind speeds, and directions. The Doppler shifts are proportional to the ratio of wind speed and the speed of light as

$\Delta\lambda = -\lambda_0 \frac{c}{v\times\cos(\theta)}$ where $\Delta\lambda$ is the wavelength shift, λ_0 is the reference or emitted wavelength, c is the speed of light, and $v \times \cos(\theta)$ is the wind speed component along the LOS. The spectral linewidth broadening is given by

$\sigma_\lambda = \frac{1}{\lambda_0}\sqrt{\frac{k_B T}{m}}$ where k_B is the Boltzmann constant, T is the particle temperature, and m is the particle mass.

There are two main ways for measuring the Doppler shift and linewidth broadening using LiDAR: direct detection and coherent (heterodyne) detection (Wandinger 2005). In direct detection Doppler LiDAR, the wavelength shift is determined by a spectrometer instrument which employs narrowband spectral filters and measures the backscattered radiation at each band. Coherent Doppler LiDAR is based on the emission of modulated pulses of single-mode single-frequency laser radiation. The detected backscattered signal is mixed with the signal of a local oscillator, and by detecting the beat frequency, the frequency shift is determined. To determine the sign of the shift, a frequency offset is introduced between the emitted pulse and the local oscillator.

Depolarization

Depolarization is not a LiDAR detection technique per se; however, because the polarization of the laser radiation emitted by a LiDAR is well known, it is possible to measure how much radiation is backscattered with the same polarization and at a perpendicular polarization. In atmospheric LiDARs, depolarization provides information about the nature of the scattering particles, as Mie scattering theory indicates that depolarization is caused by nonspherical scatterers. In mapping, LiDAR depolarization can be used to characterize surface roughness.

Light Sources

A light source is a basic part of a LiDAR system. During the early days of LiDAR experimentation, the light sources were mercury or sodium vapor lamps. Currently, the light source will most likely be a laser. Laser is an acronym for light amplification by stimulated emission of radiation. Traditional lasers consist of an optical resonator which contains an optical gain medium. This gain medium, or lasing material, is pumped with optical or electrical energy (semiconductor lasers) causing the electrons in the lasing material to be excited to a higher nonequilibrium level, and stimulated emission occurs when an interacting photon causes an electron to drop from the higher level to its ground state releasing an additional photon at the same wavelength as the interacting photon. If this stimulated emission builds up within the optical resonator to a point where the gain of the process overcomes the cavity losses at a given resonant mode, then lasing is achieved, and a relatively high-coherent beam of light will be emitted. Coherence refers to the laser beam's spatial and spectral characteristics; a perfectly coherent laser beam will travel in a single direction (spatial coherence), and the photons would be of a single wavelength, polarization, and phase (spectral coherence). In the real world, lasers are not 100 % coherent, but can emit light from several modes at different wavelengths at the same time with not necessarily the same polarization, and their beam can diverge beyond the diffraction limit. However, most lasers used in LiDARs are built to be single mode and diffraction limited. Besides the traditional electronic population inversion lasing method, it is possible to generate laser light through other processes such as relativistic free electron beams and by modifying the vibrational and rotational modes of oscillation of molecules. Lasers can produce light not only in the visible spectrum but also in other regions of the spectrum including the infrared, the ultraviolet, and the X-ray regions.

Lasers can be classified based on the lasing medium as solid-state, liquid, and gas lasers. Examples of solid-state lasers include those based on crystalline paramagnetic ions, glass, solid dyes, semiconductors, polymers, and excimers. Liquid lasers can be based on organic dyes, rare earth liquids, polymers, and excimers. Gas lasers include neutral atoms, ionized gases, and molecular gases (Weber 2001). One of the most common lasers used in LiDAR technology is based on the solid-state crystal: neodymium-doped yttrium aluminum garnet (Nd:YAG), which lases at 1,064 nm.

Based on their modes of operation, lasers can be classified into continuous wave lasers if its output power is constant over time (although the intensity of the beam can be modulated) and pulsed lasers if the optical energy is released in sudden bursts. Laser pulses packing a relatively high amount of energy, compared to continuous operation, can be obtained through the Q-switching technique. Pulses obtained through Q-switching are typically in the range of hundreds of picoseconds to tens of nanoseconds in length. Extremely short pulses in the picosecond to the femtosecond range containing very little energy can be created using the mode-locking technique.

High Signal-to-Noise Ratio (SNR) and Photon-Counting Detectors

The optical backscattered signal resulting from the interaction between the radiation and the target needs to be detected by the LiDAR system. For this purpose, many different types of photodetectors can be employed. These photodetectors include

PN, PIN, and avalanche photodiodes and photomultiplier tubes. The selection of the photodetector is a crucial aspect in the design of a LiDAR system (Kaufmann 2005), and factors that must be taken into account in this process are the wavelength and the magnitude (signal strength) and magnitude range (dynamic range) of the radiation to be detected and the speed at which it needs to be detected. Generic characteristics of photodetectors include its wavelength band of operation (spectral response), its sensitivity (how much electric signal is produced per unit of detected radiation), its noise characteristics (how much electric signal is produced even when no radiation is incident on the detector), response speed (ability to detect distinct events separated by short times), active area, number of elements (single element vs. array of detecting elements), and its operating voltage and power consumption.

Independent of the type of photon detector used, there are two main modes of operation depending on the magnitude of the detected signal: high signal-to-noise ratio (SNR) or analog detection and low SNR, also called photon counting or digital detection (Hamamatsu Corporation 2005). In high SNR LiDAR systems, the magnitude of the detected signal is many times larger than the general background noise, including scattered solar radiation and artificial lighting, and the detector thermal noise. High SNR is typical of short-range, high-power systems such as mapping and elastic backscattering LiDARs. In the low SNR domain, the magnitude of the detected signal is very close to the noise level, and in some cases, the detector responds to the excitation of single photon events, and this is why it is also called photon counting. Photon counting is used in extremely long-range systems such as SLR and LLR, for systems where the interaction between the radiation and matter is particularly weak such as in Raman LiDAR or high atmosphere Rayleigh scattering and resonant fluorescence LiDARs (Abshire et al. 2005; Whiteway et al. 2008), water penetrating (bathymetric) LiDAR, and low-power multichannel systems (Cossio et al. 2010).

The LiDAR Equation

The LiDAR equation is a mathematical formulation that provides an estimate of the received optical signal strength by a system as a function of instrument parameters, atmospheric phenomena, and detection range. The LiDAR equation is used to design systems and to evaluate the performance of existing systems, and it is inverted to determine atmospheric properties from real observations. There are many versions of the equation depending on the type of system it describes. In its most generic form, it is (Wandinger 2005)

$$P_r(R) = K_s \times G(R) \times T^2(R) \times \beta(R)$$

where $P_r(R)$ is the received power as a function of the range, K_s is a constant factor dependent upon system parameters such as transmit power and optical efficiency, $G(R)$ is a factor that depends on the geometry of the observation as function of the range, $T(R)$ is the propagation medium transmission factor, and $\beta(R)$ is a factor that describes the target backscattering properties. Each of these factors can be expanded and/or adjusted to account for the specifics of each system and application.

For instance, the LiDAR equation for elastic backscattering atmospheric LiDAR, where the targets are atmospheric constituents (atoms or molecules), can be expanded as (Wandinger 2005)

$$P_r(R) = \left[\frac{P_0 \eta c \tau A}{2}\right] \times \left[\frac{O(R)}{R^2}\right] \times \left[e^{-\int_0^r \alpha \times dR}\right]^2 \times \beta(R)$$

where P_0 is the emitted laser power (pulse energy/pulse length), η is the optical efficiency of the system, c is the speed of light in the transmission medium, τ is the laser pulse width, A is the receiving telescope area, $O(R)$ is the fractional overlap area collected by the receiver, and α is the extinction coefficient. In this case, both the atmospheric transmission and scattering coefficient are the properties under study. The scattering coefficient indicates the probability that a photon will be backscattered. The atmospheric transmittance is the exponential integration of the extinction coefficient which is proportional to the amount of scattering material in the atmosphere; it can also be considered as the effective cross-sectional area of particulates per unit volume. The combined expression $c\tau A$ is considered the scattering volume, which when multiplied by the scattering coefficient $\beta(R)$ yields the scattering cross section.

For an altimetry or mapping LiDAR, the equation can be expanded as (Bufton 1989)

$$P_r(R) = [P_0 \eta A] \times \left[\frac{1}{R^2}\right] \times \left[e^{-\int_0^r \alpha \times dR}\right]^2 \times \left[\frac{\rho}{\Omega}\right]$$

where ρ/Ω is the target backscatter or reflectance per solid angle.

These equations can be expanded even further to account for each interaction that affects the laser beam along its two-way travel from the transmitter to the receiver and as stated before need to be adjusted for the particular type of LiDAR system and application.

Comparison of LiDAR to Other Forms of Remote Sensing

Having described LiDAR technology and principles of operation, it is convenient to compare this active optical detection technique against other forms of remote sensing. It is important to remember that every remote sensing technique has its strengths and limitations, and it is crucial to understand the relative advantages and intrinsic limitations of different techniques to determine which is the most appropriate for a given application. The next two sections compare active versus passive remote sensing techniques and LiDAR versus radar.

Advantages and Disadvantages of Active Remote Sensing

Having control of the illumination source creates several advantages for active remote sensing (LiDAR and Radar) over passive techniques. The first advantage is that active systems are independent of day/night conditions. This is particularly true for Radar systems. However, certain types of LiDAR units work better under night conditions, and some can only work at night. Long-wavelength radars ($>$10 cm) are also independent of weather conditions and can work through clouds and rain.

With passive remote sensing techniques such as multispectral and hyperspectral imaging, most of what can be inferred from the target has to do with the amplitude of the detected signal (relative or absolute reflectance). With active systems, there is full knowledge and sometimes control of the parameters of the illumination signal: amplitude, frequency, phase, and polarization. This control allows researchers to study the effect that the target has on all the parameters of the emitted radiation enabling a more complete characterization of the target. The use of phase information makes it possible to accurately measure sub-wavelength scale changes in ranges, which is applied in deformation mapping using InSAR or millimeter-level ranging with LiDAR. Measuring the change in polarization (depolarization) enables the geometric characterization of the target; it is used in atmospheric LiDAR to determine if the scatterers are spherical or not and in polarimetric SAR to determine the orientation and location of the scattering sources.

Finally, measurements of perceived changes in frequency or wavelength allow the use of Doppler techniques to determine the relative speed of the target moving along the line of sight (LoS) of the LiDAR or Radar. A parameter of the illuminating signal for which there is almost full control is the power (limited by the maximum power output of the source) which can generally be adjusted to a level that optimizes the signal-to-noise ratio (SNR) of the detected return, thereby reducing the sensitivity to background and detector noise compared to passive remote sensing techniques.

Despite the many advantages of the active remote sensing technique, there are some disadvantages with respect to the passive techniques. The main disadvantage is that active sources can only sample relatively small areas at a given time, and to increase the spatial resolution, it is often necessary to reduce the extent of the study area. An additional disadvantage is that active sensors provide very little spectral information, limited to a few wavelengths compared to the hundreds of channels that can be studied with a hyperspectral system.

LiDAR Versus Radar

To compare LiDAR and Radar remote sensing, a good starting point is their respective operational wavelengths. Most operational radars work in the wavelengths between 2 and 30 cm (10–2 m), while LiDARs operate between 300 and 2,000 nm (10–9 m). On average, this is a five order-of-magnitude difference, and this has many implications for remote sensing applications. The first implication has to do with the interaction between radiation and matter. As explained earlier, scattering is a process determined by the relative size of the particles and the wavelength. In the case of atmospheric constituents, their size is comparable to the wavelengths in the optical range, and this is why it is possible to study atmospheric scattering with LiDAR. It is also possible to measure Doppler shifts and broadening from optical radiation scattered by moving atmospheric particles, which in turns allows for the remote determination of wind velocities and temperature profiles using LiDAR. The Radar wavelengths, on the other hand, are much larger than atmospheric particles and are not affected by atmospheric atomic and molecular constituents. However, low-wavelength (<10 cm) Doppler radar is sensitive to much larger water drops and ice crystals.

Besides the scattering interaction, there is also the possibility of absorption and atmospheric extinction which is the depletion of transmitted radiation, caused by the combination of scattering and emission. Atmospheric transmission is complementary to extinction. The Earth's atmosphere is practically transparent to radio waves, but it is relatively opaque in certain optical bands. This is of crucial importance for remote sensing applications from satellite platforms for which the electromagnetic radiation to be detected needs to travel through the Earth's atmosphere. Therefore, the bands of operation of spaceborne sensors are selected taking into consideration the transparency of the atmosphere. The atmosphere's transparency in the radio wavelengths allows Radar to operate under most weather conditions, which combined with its day and night operability provides a significant advantage over other forms of remote sensing. However, absorption is not entirely an undesirable phenomenon. Absorption at specific wavelengths due to atmospheric molecules is the principle used by differential absorption LiDAR (DIAL) to detect and measure the concentrations of molecules such as ozone and water vapor in the atmosphere.

A final aspect to consider in the comparison between radar and LiDAR is the divergence or spread of a Radar or laser beam. The divergence also relates to the angular resolution of a remote sensing system. Divergence is determined by diffraction at the output aperture from which optical or radio energy is emitted. The Rayleigh criterion provides an estimate of the angular resolution of optical imaging systems or the beam divergence of active systems as

$$\mathrm{Sin}\,(\theta) = 1.220\frac{\lambda}{D}$$

where θ is the angular resolution or beam divergence in radians, λ is the radiation's wavelength, and D is the diameter of the aperture (lens or antenna). Considering an average optical wavelength of 1,064 nm and a modest aperture of 1 cm, the diffraction-limited divergence of a laser beam is then 0.13 μrad. For a radio wave at an average wavelength of 10 cm and with an antenna 10 m in diameter, the divergence of the radio beam is 12.2 μrad, almost 100 times wider than the laser beam. In order to have the same divergence as the optical beam, the antenna would have to be almost 940 m in diameter. To overcome this limitation, the synthetic aperture radar (SAR) technique was developed to electronically synthesize a virtual antenna many times larger than the physical antenna, based on the platform motion. Smaller divergence of laser beams implies smaller footprints and better angular and spatial resolutions for LiDARs as compared to Radar.

The contrast of higher resolution due to smaller footprints is that LiDARs generally provides smaller spatial coverage. In addition, current spaceborne LiDAR systems for atmospheric and mapping applications operate in single beam profiling mode, which means that the sampling is performed along a single line with no scanning capabilities. On the other hand, spaceborne Radar systems have multiple beams and the capability to electronically steer the beams in a direction perpendicular to the direction of flight. Larger footprint and scanning capabilities of radar systems allow for larger spatial coverage and a better temporal resolution.

Satellite LiDAR Applications

Geodetic and Geodynamic Applications

Geodesy is the study of the shape, size, orientation, motion, and gravity of the Earth; it also includes the establishment of coordinate reference systems used to uniquely describe the location of any point on the Earth. Geodesy is the discipline that enables many of current satellite applications such as satellite-aided navigation (GPS, GLONASS, and Galileo) and satellite remote sensing mapping by establishing the geodetic frame of reference on which these systems operate.

The first geodetic observation is credited to Eratosthenes, a Greek philosopher who lived in the third century BC and who was able to conclude that the Earth had a spheroid shape and was able to estimate its size. Over the centuries, geodetic instruments and techniques have evolved, but the need to measure angles, distances, and time to determine geographic coordinates and the Earth's parameters has not changed. This need for accurate distance and time measurements led geodesists to develop the electronic distance measurement (EDM) technologies, one of which evolved into modern-day ranging LiDAR. Also, for centuries, geodesists have been performing astronomical observations to derive coordinates and distances between remote stations. They realized that this could also be done by observing man-made airborne objects, and so as technology matured, they started using balloons, airplanes, rockets, and eventually satellites as targets. So it is not surprising that the first spaceborne application of LiDAR technology was developed for geodetic studies.

This was achieved by leaving the active LiDAR equipment (laser transmitter and optical detector) on the ground (Fig. 6) and installing passive elements (retroreflector

Fig. 6 NASA MOBLAS-7 mobile SLR system circa 1980 (Courtesy of NASA)

Fig. 7 Laser retroreflector array on COMPASS satellites (Courtesy of Shanghai Astronomical Observatory, Chinese Academy of Sciences)

arrays) on satellites (Fig. 7). This architecture has many advantages, the main one being that technology can improve continuously on the ground segment and need not stop once the satellite is integrated and launched. Also the spacecraft infrastructure, being passive, does not require power or maintenance and typically has extremely long lifetime. The long lifetimes and large number of satellites carrying retroreflectors have allowed the accumulation of over four decades of ranging data.

The first geodetic satellite tracked by LiDAR was the ANNA-1B launched on October 31, 1962. ANNA-1B carried equipment to test three different satellite tracking techniques; one of them was the use of high-intensity optical beacons

(Simons 1964). The beacons operated on command and produced a sequence of five flashes separated by 5.6 s. These flashes were recorded using long-exposure photography; simultaneous observations from different stations allowed the determination of the satellite position (Harris and Berbert 1966). The first geodetic satellite that carried a retroreflector array was the Beacon Explorer-B (designated as the Explorer 22) (Degnan et al. 1994). The Explorer-B was launched on October 9, 1964; it was a 116-pound satellite that in addition to the retroreflector also carried a radio beacon. The satellite was tracked from stations around the world using both radio and LiDAR technology, although radio equipment was much cheaper than the optical Radar, and because the satellite was magnetically stabilized, the retroreflectors were oriented in such a way that it was only possible to track the satellite from stations on the northern hemisphere.

The first laser tracking of the Explorer 22 was carried out on October 31, 1964, by a team from NASA's Goddard Space Flight Center (GSFC). This was the origin of a geodetic LiDAR technique named satellite laser ranging (SLR). The Explorer 22 was soon joined by more satellites carrying corner cube retroreflectors including more satellites of the Explorer series, Explorer 22 (launched on April 29, 1965), Explorer 29 also known as GEOS1 (launched in November 6, 1965), and the Explorer 36 or GEOS 2 (launched on January 11, 1968). The Centre National d'Etudes Spatiales (CNES) from France also contributed to SLR by launching a pair of geodetic satellites, the Diadème-1 D1C (February 08, 1967) and the Diadème-2 D1D (February 15, 1967), equipped with dual-frequency Doppler transmitters and retroreflector arrays. The first international SLR campaign occurred in the spring of 1967 with the participation of five laser stations, three operated by CNES and located in France, Algeria, and Greece, one station operated by NASA in Maryland, and one operated by the Smithsonian Astrophysics Observatory (SAO) in New Mexico. Data from this campaign was used to compare SLR to traditional optical observations, and an improvement by a factor of 4 in the accuracy of determined positions was estimated; however, most important was the development of SAO standard Earth's gravity model (Degnan et al. 1994).

This first international SLR campaign with stations spread across the world helps illustrate the mode of operation of this geodetic LiDAR technique. As shown in Fig. 8, a single satellite can be tracked simultaneously from stations separated by a few meters up to thousands of kilometer, and using triangulation, it is possible to determine the baselines between the stations. Observations from SLR stations are enhanced by colocation with other global space geodetic techniques such as very long baseline interferometry (VLBI), global navigation satellite systems (GNSS), and Doppler orbitography and radiopositioning integrated by satellite (DORIS).

The early geodetic satellites were not optimal for geodesy and relativistic applications because they were launched into low orbits and because they carried a variety of instruments which enlarged their cross section and lowered their density. The satellite's low orbit and low density limited the visibility times and increased their susceptibility to gravitational perturbations, while the large cross section made them susceptible to atmospheric drag, radiation pressure, and other nonconservative

Fig. 8 Simultaneous SLR from three stations at the Goddard Geophysical and Astronomical Observatory (Image courtesy of NASA)

forces. To overcome these limitations, the ultimate Earth satellite, the Moon, was equipped with retroreflectors. As early as 1962, J.E. Faller had proposed the idea of placing a retroreflector on the surface of the Moon, and in 1965, the lunar ranging experiment (LURE) multi-institutional team was formed. Between 1969 and 1973, a total of five retroreflectors were placed on the Moon, three of them by manned Apollo missions (11, 14, and 15) and two French-built retroreflectors carried by the Russian lunar rovers Lunokhod 1 and 2 (Luna 17 and 21 mission) (Bender et al. 1973). These lunar retroreflectors made it possible to range to and track the Moon from stations around the world using a LiDAR technique called lunar laser ranging (LLR).

As a complement to the lunar retroreflectors, several satellites designed exclusively for geodesy using SLR have been launched into relatively high and very stable orbits. These “cannon ball” satellites have high densities and small surface area covered almost entirely by retroreflectors. The first was the French-built Starlette launched in 1975, followed by the American Laser Geodynamics Satellite (LAGEOS-1) launched in 1976. Other SLR-only satellites include the Japanese Ajisai (launched in 1986), the Soviet Etalon-1 and 2 (launched in 1989), the LAGEOS-II (built by the Agenzia Spaziale Italiana and launched in 1992), and the French satellite Stella (launched in 1993).

To this date, more than 130 satellites have been tracked from more than 70 laser stations around the world (Fig. 9) (The International Laser Ranging Service). The massive amount of data collected for almost half a century from SLR and LLR has allowed the accurate determination of the ground station coordinates to the

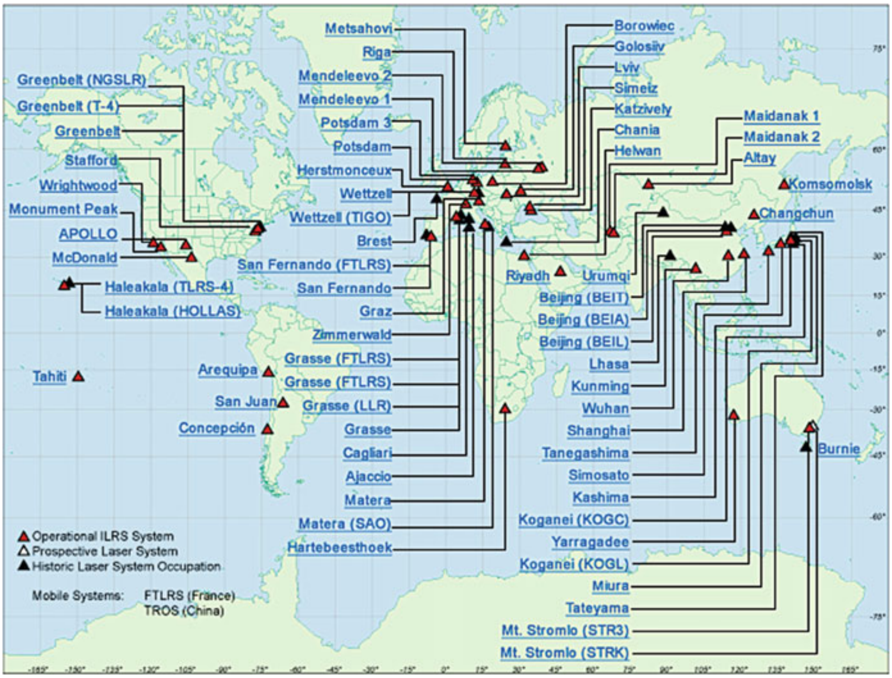


Fig. 9 Stations of the international laser ranging service (Courtesy of ILRS/NASA)

millimeter level and the satellite orbits to the centimeter level. These techniques combined with other space geodetic techniques such as VLBI and GNSS have been applied to scientific issues such as the modeling or establishment of the Earth's gravity field, reference frame, and orientation parameters, to prove geodynamic theories such as plate tectonics, glacial rebound, and crustal deformation, to test principles of general relativity, and to determine Earth–lunar and solar system celestial mechanics parameters (The International Laser Ranging Service; Degnan et al. 1994). Also the establishment of the terrestrial reference frame (TRF) and Earth orientation parameters (EOP) along with the accurate determination of satellite orbits is crucial for satellite applications such as navigation and Earth observation. Some of these applications are described next.

Observations and Modeling of the Terrestrial Gravity Field

The Earth's gravity field is a 3D vector field that specifies the acceleration that an object will experience at a given point at or above the Earth's surface. Its main component or mean gravity, 9.8 m/s^2, is the equivalent gravity of a uniform mass distribution and a spherical shape. The next-order deviation from this simplified model is due to the Earth's rotation and oblate shape. Smaller-order variations are due to mass distribution heterogeneity. In addition to spatial variations, there are temporary variations due to mass redistribution through and among the atmosphere, cryosphere, hydrosphere, and solid Earth.

To study the gravity field, the gravitational potential is modeled by a spherical harmonic series of the form (Heiskanen and Moritz 1967)

$$U = \frac{GM}{r}\sum_{n=0}^{\infty}\sum_{m=0}^{n}\left(\frac{r_0}{r}\right)^n \overline{P}_{nm}(\sin\phi) \times \left[\overline{C}_{nm}\cos(m\lambda) + \overline{S}_{nm}\sin(m\lambda)\right]$$

where n is the degree and m is the order, $\overline{P}_{nm}$ is the fully normalized Legendre polynomial and associated functions, r_0 is the reference radius, ϕ is the latitude and λ is the longitude, and $\overline{C}_{nm}$ and $\overline{S}_{nm}$ are the series coefficients determined from observational data from a variety of sources. Similar spherical harmonics can be used to describe the shape of planetary bodies.

Before dedicated gravity satellite missions such as CHAMP (2000), GRACE (2002), and GOCE (2004), global gravity observational data were obtained by tracking satellites using SLR (Degnan et al. 1994). A satellite orbit is determined primarily by the Earth's gravity field and affected by nonconservative forces such as drag (atmospheric, thermal, neutral density, and charged particles) and radiation pressure. If the effects of the nonconservative forces can be accounted for, then the differences between the predicted and determined orbit of a satellite can be attributed to inaccuracies in the gravity model. Data from SLR, in situ, airborne and shipborne gravimetry, and satellite altimetry have been used to produce gravity models until this last decade. However, data from SLR provide the longest baseline to study temporal variations of the low-order zonal harmonic components of the gravity field (Degnan et al. 1994).

Terrestrial Reference Frame (TRF) and Earth Orientation Parameters (EOP)

Satellite applications require a foundation of permanently operating reference stations to collect the observations required to provide their mapping, positioning, and timing services. This network of stations serves as a terrestrial reference frame which defines the origin (center of mass) and orientation of the Earth. Earth orientation parameters – universal time (UT1), length of day (LOD), and coordinates of the pole and celestial pole offsets – describe the irregularities of the Earth's rotation and the orientation of the axis of rotation relative to inertial space and celestial reference system. Observations with space geodetic techniques, including SLR, LLR, GPS, and VLBI, provide the required data to define the Earth's center of mass, UT1, LOD, and polar motion. VLBI is the only technique capable of accurately determining changes in the orientations of the earth with respect to the crust and to a celestial reference frame composed of natural radio sources (quasars) – the best current approximation of a true inertial reference.

Precision Orbit Determination for Navigation and Earth Observation Missions

Precision orbit determination (POD) is an important aspect of satellite operations, and for some satellites, such as navigation and remote sensing satellites, it is of crucial importance. It is also a technique that is in cyclical improvement. In order to obtain a precise orbit, an accurate gravity model is required. Over periods of years, gravity models are improved, based on observations of satellite orbits obtained from optical, radar, and SLR tracking. The improved gravity model in turn allows for better orbital determination, and so the cycle continues. In the early years of the space era, satellites were tracked from the ground using optical photographic cameras and basic Doppler radar techniques with accuracies of approximately 10 m for satellites in a 1,000 km altitude orbit (Vetter 2007). The introduction of SLR in 1964 provided and alternate method for satellite tracking with an improved accuracy of a few meters. The ability to track satellites has continued to improve over the years to the millimeter-level accuracy obtainable today (McGarry et al. 2005).

SLR is a more precise technique than radar because it can obtain accurate ranges to retroreflector arrays, whose position with respect to the satellite center of mass is well known, whereas radar obtains a range to the center of the satellite radar cross section, whose position relative to the center of mass is known to a lower level of accuracy. Currently, satellites with orbital altitudes below 20,000 km can be continuously tracked using GNSS (or other systems such as NASA’s TDRS) with centimeter-level precision or better. However, for GNSS satellites, to provide positioning, timing, and navigation accurately is necessary to have accurate knowledge of their own orbits. GNSS satellites are tracked by a variety of means including optical and radar. Most of GLONASS satellites, the two current Galileo spacecrafts (GIOVE-A and GIOVE-B), one of the Chinese COMPASS, the Japanese QZS-1, and one GPS satellite (GPS-36) carry retroreflector arrays to be tracked by SLR (GPS-35 decommissioned in April 2009 also carried an array) (The International Laser Ranging Service). Other satellites whose orbit needs to be accurately determined for the fulfillment of their scientific objectives are therefore tracked by SLR and include gravity mappers GOCE and GRACE; radar and LiDAR altimeters Cryosat, Jason 1 and 2, and ICESat (decommissioned); and remote sensing satellites Envisat, ERS-2, TerraSAR-X, and TanDEM-X (The International Laser Ranging Service).

Laser Altimetry and Topographic Mapping

Laser altimetry was the first application of spaceborne LiDAR on which the active equipment was carried by the spacecraft. Laser altimetry originated as an alternative to more traditional Radar altimeter. This was because the large divergence of radio beams makes its footprint on the surface of the planet many times larger than the

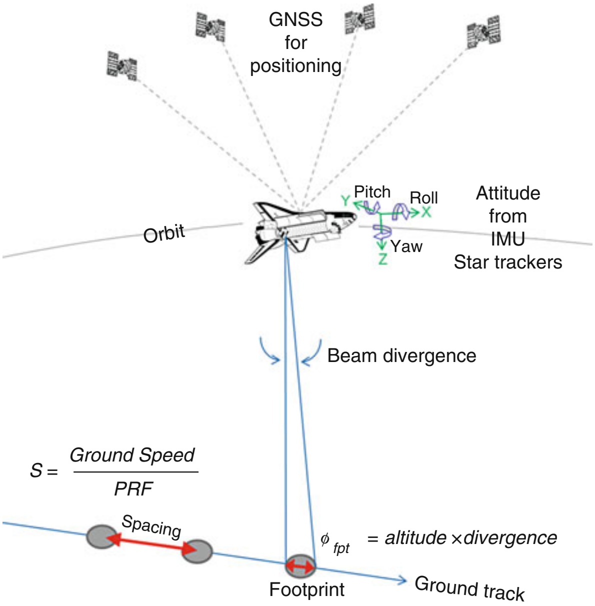


Fig. 10 Principles of operation of satellite LiDAR altimetry

footprint of a narrower laser beam. In altimetry, a smaller footprint results in a more accurate and representative estimate of height (Bufton 1989). As illustrated in Fig. 10, in nadir-looking satellite LiDAR altimetry, the laser footprint is dependent on the satellite orbital altitude and laser beam divergence, while the spacing between footprints (spatial resolution) depends on the orbital velocity and the laser pulse repetition rate (PRF). The accuracy of the derived elevation depends on the precise determination of the spacecraft orbit and attitude.

The first spaceborne altimetry systems were not deployed on Earth observation missions but rather on missions to the Moon and Mars. This was because the Earth's atmosphere presented a huge challenge as most of the laser energy is scattered by atmospheric constituents on a two-way trip from outside the atmosphere to the ground and back. Table 2 presents a historical evolution of spaceborne LiDAR altimeters and their main technical characteristics. The first laser altimeter system was deployed with the Apollo 15 mission to the Moon in 1971. The altimeter was part of the orbital science investigation and was designed to take an altitude reading for each photograph taken with a mapping metric camera (every 20–28 s), although the altimeter was also able to range independently of the camera (at least every 20 s) (Alley et al. 1969). The metric camera, the altimeter, and two other cameras (panoramic and stellar) were located in the scientific instrument module (SIM) within the Apollo service module. The Apollo laser altimeter was based on a Q-switched ruby laser and a photomultiplier tube detector; the system was also deployed on the Apollo 16 and 17 missions in 1972. At its highest sampling rate of 0.05 Hz, the altimeter sampled the lunar surface height every 30–43 km with a footprint of roughly 30 m in diameter. The main problem with this instrument was its

Table 2 Evolution of technical characteristics of spaceborne laser altimeters

Year	System	λ(nm)	PRF (Hz)	Pulse width (ns)	Pulse energy (mJ)	Beam diver. (mrad)	Range (km)	Detector
1971	Apollo 15 laser altimeter	694	<0.05	10	200	300	110	PMT
1992	Mars Orbiter Laser Altimeter 1	1,064	10	8	48	420	780	Si APD
1994	Clementine LiDAR	1,064	0.6	10	171	500	640	Si APD
1994	Lidar In-space Technology Experiment[a]	1, 064, 532, 335	10	27	470	1,800	260	APD
								PMT
								PMT
1996	Shuttle Laser Altimeter 1	1,064	10	8	40	350	305	Si APD
1996	NEAR Laser Range finder	1,064	8	12	15	235	50	Si APD
2003	Geoscience Laser Altimeter System	1,064, 532	40	5	75,32	110	600	Si APD
2004	Mercury Laser Altimeter	1,064	8	6	20	80	800	Si APD
2009	Lunar Orbiter Laser Altimeter	1,064, 532	28	6	2.7	100	50	Si APD

[a] LITE was designed primarily as an atmospheric LiDAR although it performed ranging to land and ocean surface. Difference in the design of atmospheric LiDARs and altimeters can be observed in terms of the pulse widths and beam divergence which tend to be larger in atmospheric LiDARs

short lifetime; during the Apollo 15 mission, the altimeter showed anomalous operation and stopped working in lunar orbit #38. As a result, only two complete and two partial surface profiles had useful data (Robertson and Kaula 1972). For the Apollo 16 mission, the sampling rate was reduced, and the instrument lifetime was extended to lunar orbit #63, some 2,372 laser pulses, of which 69 % had valid data, yielding five complete lunar surface profiles (Wollenhaupt et al. 1972). For the last Apollo lunar mission, the laser was modified to increase its lifetime, and the altimeter lasted during the entire mission. The laser fired 4,026 pulses and yielded 16 complete lunar surface profiles (Wollenhaupt et al. 1973). Data from all the missions combined yielded 7,080 height points, and from these, a lunar mean radius was determined, and a spherical harmonic representation of the lunar shape was produced completely to the 12th order and degree. However, the coverage was limited to ±26° lunar latitude.

LiDAR altimetry returned to the Moon in 1994 onboard the Clementine mission. This instrument had a mass of only 2.4 kg (Smith et al. 1997) (compared to the 22.5 kg of the Apollo altimeter (Robertson and Kaula 1972)), yet it fired around 650,000 laser pulses. Because the system was designed as a military ranging system and not an altimeter, only 19 % of the fired pulses caused reflections that were detected, and of these, only 72,548 were filtered out as valid surface returns (Smith et al. 1997). These data covered the lunar surface between 79°S and 81°N latitude, with a minimum along-track resolution of 20 km and an across-track resolution of roughly 60 km. From these data, a spherical harmonic representation of the lunar shape complete to the 72nd order and degree was produced (Smith et al. 1997). Most recently, the Lunar Reconnaissance Orbiter (LRO) carrying the Lunar Orbiter Laser Altimeter (LOLA) (Ramos-Izquierdo et al. 2009) has been mapping the Moon since September 2009, and as of June 19, 2010, LOLA had collected over two billion elevation measurements using its multichannel technology (Smith et al. 2010).

Besides the Moon, the shape and topography of three other extraterrestrial solar system bodies have been mapped: Mars, Mercury, and the asteroid 433 Eros. The first attempt to use LiDAR to map the Martian topography was the Mars Orbiter Laser Altimeter 1 (MOLA-1) launched onboard the Mars Observer launched in 1992 (Smith et al. 2001; Garvin et al. 1998). Unfortunately, the Mars Observer was lost on August 21, 1993, a few days before the orbit insertion maneuver. The second attempt was by MOLA-2 onboard the Mars Global Surveyor; MOLA-2 performed regular mapping operations between February 28, 1999, and June 30, 2001, and within this time frame, approximately 640 million points were collected of the Martian surface (Smith et al. 2001; NASA). Figure 11 shows some samples of Mars topography from MOLA-2 data.

Eros was mapped by the laser range finder (Colea et al. 1996) onboard the NEAR-Shoemaker spacecraft launched in 1996, which entered orbit around Eros on February 14, 2000, and landed on the surface of the asteroid on February 12, 2001, and the mission was terminated on February 28, 2001. During its mapping mission, the laser range finder collected around 11 million measurements and allowed the best determination of shape, gravity, and rotational state of any asteroid

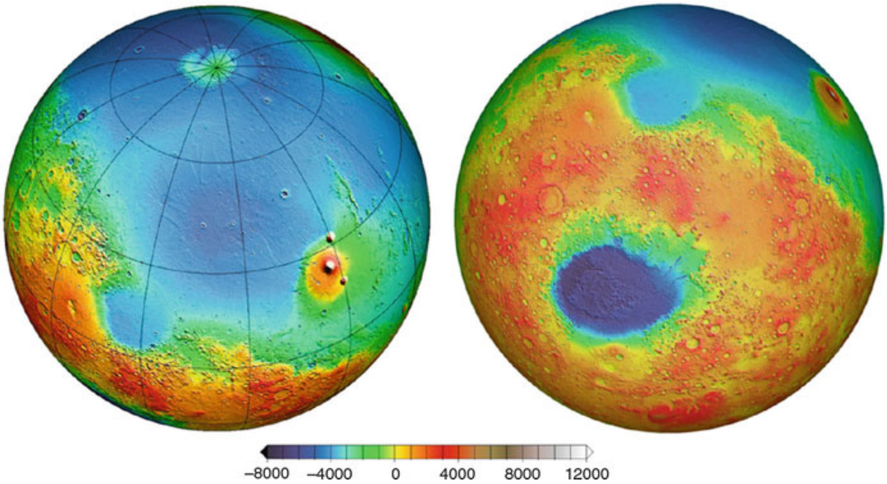


Fig. 11 Mars topography from MGS – MOLA-2 (Image courtesy of NASA)

to date (Zuber et al. 2000; Miller et al. 2002). The most recent extraterrestrial body whose surface has been studied with LiDAR is Mercury. The Mercury surface, space environment, geochemistry, and ranging (MESSENGER) mission was launched on August 3, 2004, carrying the Mercury Laser Altimeter (MLA) (Cavanaugh et al. 2007). After launching from Earth, MESSENGER has to perform six reversed gravity assists to obtain an orbital orientation and velocity suitable for its orbital insertion around Mercury in March 2011. These gravity assists are the result of flybys of planetary bodies, one with Earth (2005), two with Venus (2006 and 2007), and three with Mercury (January and October 2008, September 2009). MLA has been activated on the three Mercury flybys, and results have been reported for second flyby. During the flyby, a 3,200-km-long profile along the equatorial region was collected (Zuber et al. 2008). The laser footprint at the surface ranged between 23 and 134 m, while the spacing between the footprints varied from 725 to 888 m. Even this modest data profile has improved our knowledge of the shape and topography of the planet and has provided a preview of Mercurian crater morphology.

With regard to planet Earth, there are a few reports that indicate the existence of an altimetry LiDAR system named LORA, which was used to obtain precise altitude of photographs taken from a large format camera onboard a Soviet satellite (Werner et al. 1995, 1996). This LiDAR was reported to be operational as early as 1984; however, it has been hard to obtain independent confirmation of these reports. The first confirmed LiDAR returns from the surface of the Earth were obtained in September 1994 during the STS-64 mission. The LiDAR In-space Technology Experiment (LITE) was flown into space in the cargo bay of the Space Shuttle Discovery (Winker et al. 1996). However, LITE was designed primarily as an experimental atmospheric LiDAR and will be discussed in greater length in the next section.

The first LiDAR altimeter designed for Earth observation was the Shuttle Laser Altimeter (SLA) (Garvin et al. 1998). SLA was designed to fit in two hitchhiker canisters mounted on as special bridge structure carried in the Shuttle cargo bay as part of the small self-contained payload program (SSCP) most commonly known as the Getaway Special (GAS). This compact design allowed the SLA to be carried on any shuttle mission on which there was room for the GAS bridge. The SLA design was based on MOLA-1 and was constructed using MOLA spares. One of the GAS canisters housed the optical receiver that consisted of a 38 cm Cassegrain telescope and at its prime focus a silicon avalanche photodiode detector (Si APD). It also contained a coaxial transmitter based on a diode-pumped, Q-switched, Nd:YAG laser. The second canister contained the flight computer, power electronics, temperature sensors, and ancillary equipment. An upgrade from the MOLA architecture was the inclusion of a waveform recorder which digitized each received pulse in 4 ns samples quantized at 8 bits. The digitizer allows the determination of a redundant time of flight obtained from the time interval meter (TIM) to characterize the structure of the surface that caused the backscattering. SLA was flown twice: the first time was during the Endeavour STS-72 mission in January 1996 (Garvin et al. 1998) and the second during the Discovery STS-85 mission in August 1997 (Carabajal et al. 1999). During the STS-72 mission, SLA-01 collected about 82 h of nadir-looking altimetry data, roughly totaling three million observations. The Endeavour orbit for STS-72 had an altitude of 300 km, an inclination of 28.45°, and an average orbital velocity of 7 km/s. The orbit inclination and the nadir-looking orientation of SLA constrained the ranging acquisition in the midlatitudes between 28.45°N and 28.45°S, the laser footprint size determined from the altitude and beam divergence was $\sim$100 m, and the spacing between footprints determined by the combination of the velocity and PRF was $\sim$700 m. After preprocessing and filtering, roughly 475,000 valid returns were obtained from land and 1.1 million from the ocean surface (Garvin et al. 1998).

For the second flight of SLA onboard the Discovery STS-85 mission, the hardware was upgraded to include a variable gain amplifier (VGA) that allowed the detector to adjust to the high dynamic range of the laser returns observed during SLA-01 that had caused the saturation of the waveform recorder. Similar to SLA-01, SLA-02 collected almost 83 h of data, firing close to three million points (Carabajal et al. 1999). The main difference was that the orbital inclination of STS-85 was 57° which allowed altimetry sampling up to high latitudes. After preprocessing and filtering, roughly 590,000 valid returns were obtained from land and 1.5 million from the ocean surface. There were plans for two more flights of SLA to keep improving the system by reducing the beam footprint and increasing the PRF. A third flight was planned for late 1998 in partial support of the Shuttle Radar Topography Mission (SRTM). However, no additional flights of SLA past SLA-02 were executed. Figure 12 shows the ground tracks of collected data from the SLA-01 and SLA−02 experiments. Data from the SLA mission were compared against other ground and sea surface elevation databases (Behn and Zuber 2000; Harding et al. 1999) and were also used to perform accuracy assessments of the later collected STRM dataset (Suna et al. 2003).

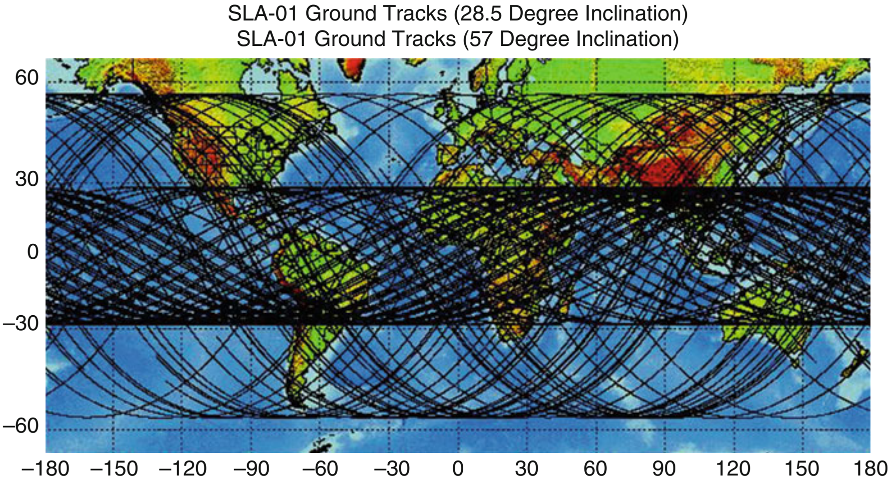


Fig. 12 Ground tracks for the SLA-01 and SLA-02 collections

The lessons learned from the two SLA missions were incorporated into the most recent and advanced spaceborne LiDAR altimeter for Earth observation to date: the Geoscience Laser Altimeter System (GLAS). GLAS was deployed on a dedicated platform: the Ice, Cloud, and land Elevation Satellite, ICESat (Abshire et al. 2005; Schutz et al. 2005). ICESat was launched on January 13, 2002, into a 600 km altitude orbit with a 94° inclination. This orbit has a nadir repetition cycle (within 1 km) of 183 days (or 2,753 revolutions), and ground track spacing within the cycle is 15 km at the equator and 2.5 km at ±80° latitude.

The GLAS transmitter was powered by three diode-pumped, Q-switched Nd: YAG lasers which operate one at a time (Abshire et al. 2005; Schutz et al. 2005). The lasers produced 5 ns pulses at 40 Hz and 1,064 nm, part of the 1,064 nm pulse was passed through a nonlinear frequency-doubler crystal to obtain a 532 nm pulse. The transmitted pulse energy was 75 mJ at the infrared wavelength and 35 mJ at the green wavelength with a beam divergence of 110 μrad. The orbital and laser characteristics yielded a footprint of 65 m on the surface with successive spots spacing of 172 m. The backscattered radiation was collected by a 100 cm diameter beryllium telescope; the 1,064 nm component is used to detect strong backscattering in analog mode from clouds, water, ice, and land surfaces, while the 532 nm component was use in photon-counting mode to detect scattering from thin high-altitude clouds. The 1,064 nm signal was filtered through an 800 pm spectral filter and detected by a Si APD (there were actually two APDs for redundancy). The APD output was digitized separately at 1 GHz and 2 MHz rates; the 1 GHz rate enables a range resolution of 15 cm for accurate surface determination, while the 2 MHz yields a 77 m resolution for the detection of thick clouds and aerosols. The 532 nm component was filtered twice through 370 and 30 pm spectral filters to limit background light, and the resultant beam was split into eight beamlets that were individually detected by eight Si APD detectors operating in Geiger mode (Abshire et al. 2005; Schutz et al. 2005).

To obtain an accurate geolocation of the laser returns, besides the accurate determination of the two-way time of flight, it is necessary to determine the position and attitude of the instrument and the orientation of the fired laser shot. Precise orbit determination (POD) was performed via GPS tracking using two redundant dual-frequency blackjack receivers connected to two separate antennas on the zenith deck of the spacecraft (Schutz et al. 2005). On the nadir deck, a corner cube reflector array allowed the satellite to be tracked using SLR for an accuracy assessment of the GPS-derived orbit (Schutz et al. 2005). There were two attitude determination systems onboard the spacecraft, one for the satellite and one for the sensor optical bench. GLAS's optical bench attitude was determined to better than 10 μrad with reference to inertial space through a stellar reference system (SRS) based on data acquired from a 10 Hz zenith looking star camera and a precision gyroscope (Schutz et al. 2005). In addition, the far-field pattern of the laser beam for each laser pulse was imaged, and its orientation was determined with respect to the optical bench and inertial space (Schutz et al. 2005). GLAS was designed to perform nadir pointing ranging; however, the spacecraft could be commanded so GLAS could point ±5° off-nadir to acquire targets of opportunity. Figures 13 and 14 show photos of the ICESat satellite integration, which highlight crucial elements of GLAS and the subsystems that enabled precise orbit and attitude determination.

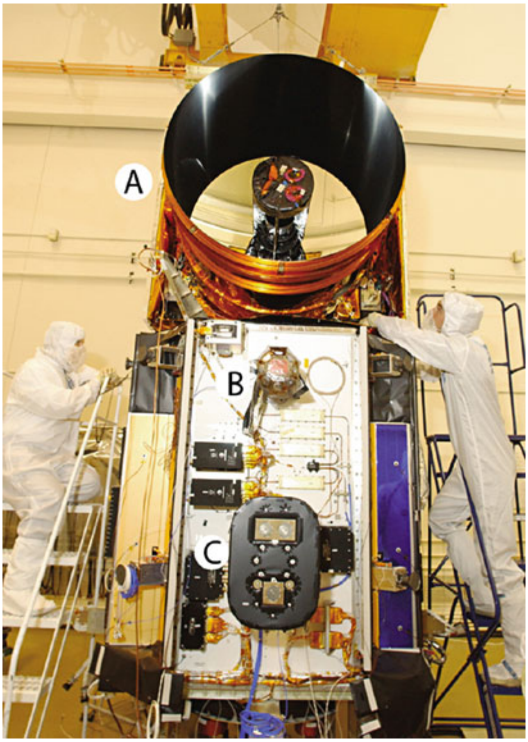


Fig. 13 ICESat's nadir deck showing (**a**) receiving telescope, (**b**) retroreflector array, and (**c**) telemetry antenna (Image courtesy NASA)

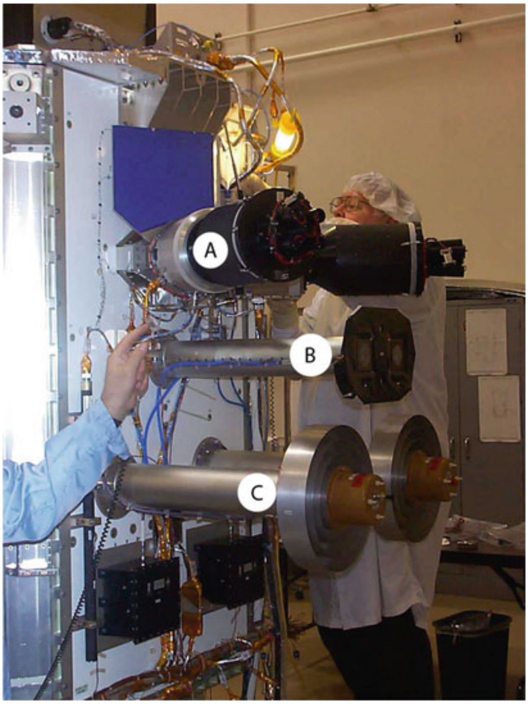


Fig. 14 ICESat's Zenith deck showing (**a**) the satellite star trackers, (**b**) telemetry antenna, and (**c**) GPS antennas (Image courtesy NASA)

The GLAS lasers were expected to last for 3 continual years of operation. Unfortunately, laser 1 failed prematurely after 37 days. This failure prompted a change in the collection strategy for the mission from a continual collection with an 8-day repeat cycle to a campaign collection mode with a 33-day repeat cycle, resulting in less temporal and spatial resolution but allowing the measurement of polar ice height over the extended 7-year period. The last GLAS laser ceased operation on October 11, 2009, and was decommissioned in August 14, 2010 (Abshire et al. 2005). In its almost 8 years in space, GLAS fired almost two billion laser pulses (Abdalati et al. 2010). The primary objective of the ICESat mission was the accurate determination of interannual and long-term changes of polar ice volume and mass balance; however, additional applications included the monitoring of land topography, hydrology, vegetation canopy height, cloud heights, and atmospheric aerosol distributions (Abshire et al. 2005). Figure 15 illustrates the use of GLAS data collected between 2003 and 2007 to generate maps of Antarctic and Greenland's ice sheet elevation change rates. The images indicate the dynamic thinning of ice sheets in certain areas and the accumulation of ice and snow in others.

To continue the critical measurement of the polar ice sheets, an improved ICESat-2 mission is currently being developed and scheduled for launch in 2016 (Abdalati et al. 2010). To obtain a denser spatial sampling than that of ICESat-1, a multi-beam approach (Figs. 16 and 17) combined with a higher PRF is under study (Yua et al. 2010). The baseline design consists of a micropulse laser with a PRF of a

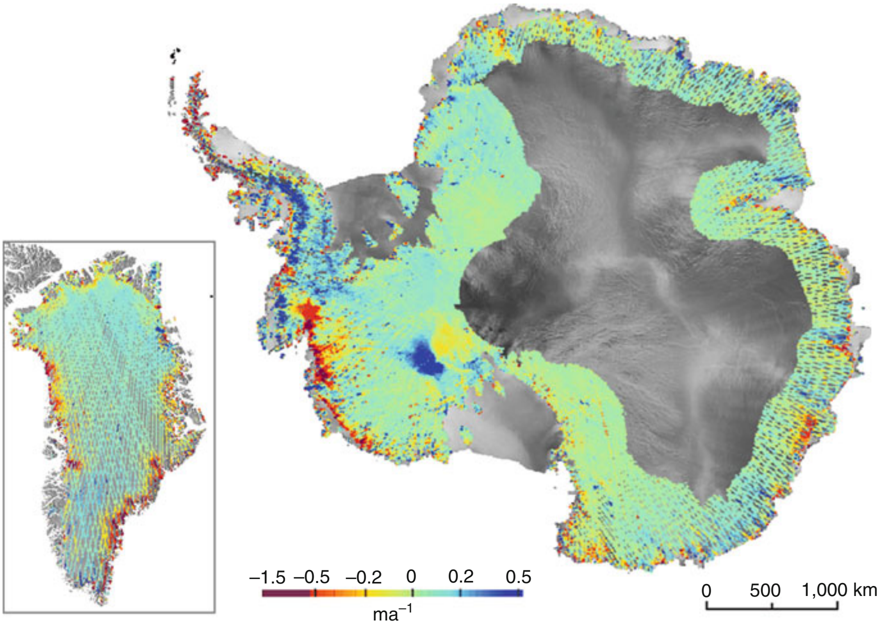


Fig. 15 ICESat data showing changes in elevation (m/year) in the Greenland and Antarctica ice sheets (Image courtesy of NASA)

Fig. 16 Multi-beam LiDAR transmitter concept for the ICESat 2 mission (Image courtesy of NASA)

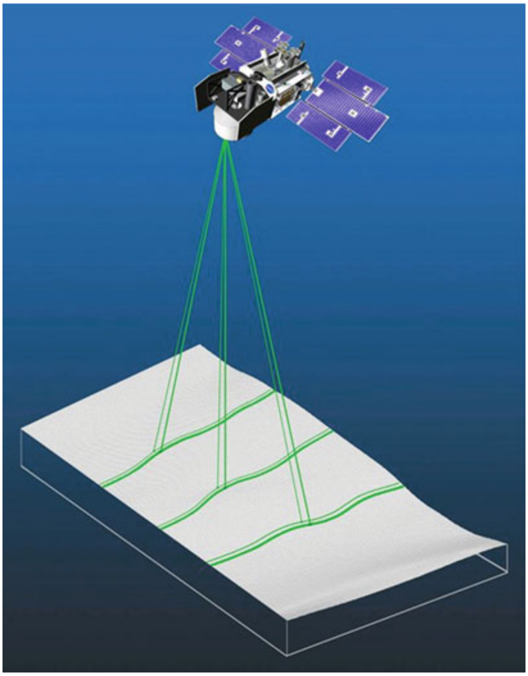
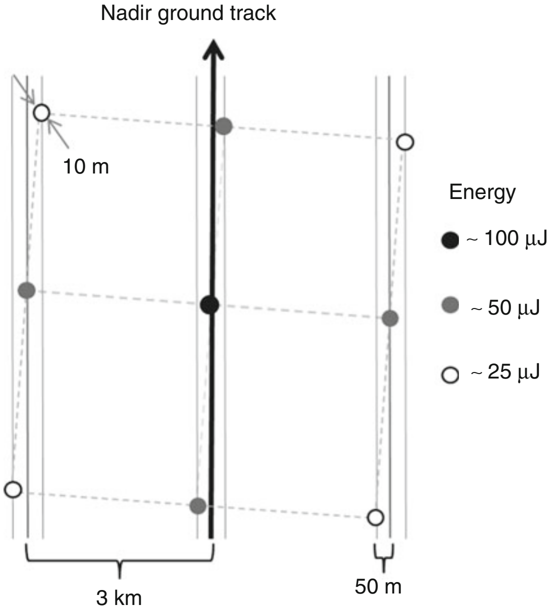


Fig. 17 Layout of the multi-footprint concept for ICESat 2 obtained from the DOE



10 kHz, 0.1 mJ of energy per pulse, and a pulse width of ~1 ns. A diffractive optical element (DOE) splits the beam into nine beamlets with different energy and arranged in a 3 × 3 slanted array (Fig. 17). The footprint of each beam is expected to be 10 m in diameter, and because the array is slanted with respect to the flight line, the projection on the ground produces nine parallel tracks grouped in threes. Each group will be spaced 3 km apart, and within the group, the spot separation in the across-track direction will be 50 m (Fig. 17).

Besides ICESat-2, the NRC decadal survey recommends two additional LiDAR altimetry missions to be launched before 2020. The most immediate mission is the Deformation, Ecosystem Structure, and Dynamics of Ice (DESDynI) (National Research Council 2007) which will attempt to exploit the synergy between L-band polarimetric InSAR and multi-beam LiDAR altimeter. The second mission recommended for the last half of the decade is LiDAR surface topography (LIST). The objective of LIST will be to produce a global elevation dataset with a horizontal resolution of 5 m with at least 10 cm vertical precision (National Research Council 2007).

Atmospheric Studies

Obtaining global datasets on atmospheric composition, structure, and circulation is of crucial importance for the development of global climatic models. These datasets are obtained with a myriad of instruments using both direct detection and remote sensing. It is also important to use both a bottom-to-top and top-to-bottom approach.

In situ Radar and LiDAR provide the bottom-to-top measurements which are able to detect phenomena in the lower denser layers of the atmosphere, but because of the higher density of the lower layers, they are not able to obtain measurements of the thinner upper layers. Spaceborne sensors provide the top-to-bottom view detecting phenomena in the higher and thinner layers of the atmosphere and generating a much needed global coverage not attainable any other way. Spaceborne atmospheric LiDARs have been employed mainly to study the Earth's atmosphere and in particular cases the Martian atmosphere. All other planetary atmospheres in the solar system are too dense to be probed in the optical wavelengths.

Although there are some unconfirmed reports that as early as 1984 a Soviet reconnaissance satellite carried a LiDAR to obtain precise altitude of photographs taken from a large format camera and was used for early atmospheric observations (Werner et al. 1996; Werner et al. 1995), the first confirmed spaceborne LiDAR built primarily for atmospheric studies flew into space in September 1994. The LiDAR In-space Technology Experiment (LITE) was flown into space in the cargo bay of the Space Shuttle Discovery during the STS-64 mission (Winker et al. 1996). LITE was designed and built based on the experience accumulated over two decades by NASA's Langley Research Center designing, building, and operating ground-based and airborne atmospheric LiDARs. LITE was designed mainly to detect and measure clouds and aerosols in the troposphere and stratosphere, determine the height of the planetary boundary layer (PBL), and derive temperature and density profiles in the stratosphere at heights between 25 and 40 km. It was also capable of detecting returns from land and sea surfaces, however, without the precision of an altimetry system.

As shown in Fig. 18, LITE was designed to fly in the cargo bay of the space shuttle integrated into a Spacelab 3 m pallet. The laser transmitter system was based on two redundant flashlamp-pumped, Q-switched, Nd:YAG lasers (Winker et al. 1996). Part of the energy of the 1,064 nm fundamental wavelength was passed through nonlinear frequency-doubling crystals to obtain 532 and 355 nm beams. The pulse repetition frequency (PRF) was 10 Hz, and each pulse had a width of 27 ns; energy per pulse was 470, 530, and 170 mJoules, with a divergence of 1.8, 1.1, and 0.9 mrad for the 1,064, 532, and 355 nm wavelengths, respectively. The laser beams were steered through a two-axis gimbaled prism to maintain optical alignment with the field of view of the receiver. The receiver was based on a 1-m diameter Ritchey–Chrétien telescope, with a rotating wheel with multiple aperture stop settings to configure the instrument for day or night collections. Dichroic beam splitters separate the return signal into the three spectral components, and part of the 532 return signal was used to determine and control the boresight alignment between the transmitter and receiver. The three beams were directed through narrowband spectral filters before their respective detectors, photomultiplier tubes (PMT) for the 355 and 532 nm components, and an avalanche photodiode (APD) for the 1,064 nm component. The output from the detectors was digitized with 12-bit amplitude at 10 MHz (550 μs).

The STS-64 carrying LITE was launched into a 260 km altitude orbit with a 57° inclination and 7.4 km/s orbital velocity. These orbital characteristics combined with

Fig. 18 The LITE LiDAR onboard the Space Shuttle Discovery (Image courtesy of NASA)

optical transmitter specifications yielded footprints 470 and 290 m in diameter for the 1,064 and 532 nm beams, respectively, and footprints were spaced every 740 m. During the 11-day mission, LITE was operated roughly 5° off-nadir to avoid saturation from high specular reflections and acquired a total of 53.6 h of quick view data (43.5 high-rate profiles) collection; almost two million laser pulses were fired (1.16 from the first laser, 0.77 from the second) (Winker et al. 1996). These collections provided the first ever high-resolution transects of the atmospheric constituents and cloud structures. LITE data was validated against ground-based and airborne measurements.

After LITE, there were several short-lived spaceborne atmospheric LiDAR experiments. Including the Balkan-1 onboard the Spektr module of the Russian MIR space station (launched on May 20, 1995) (Werner et al. 1995), the French designed and built l'Atmosphere par LIdar Sur SAliout (ALISSA) onboard the Priroda module (launched April 23, 1996) also of the MIR space station (Chanin et al. 1999) and the Balkan-2 onboard the ALMAZ-1B Earth observation satellite (launched on January 1, 1997) (Matvienko et al. 1994). Currently, the joint NASA and CNES Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite carry the only operational spaceborne terrestrial atmospheric LiDAR: the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP).

Launched on April 28, 2006, CALIPSO is managed by NASA's Langley Research Center due to the center's overall expertise on atmospheric LiDAR systems. An interesting aspect of the CALIPSO mission is that it is part of the afternoon or "A-Train" satellite constellation which also includes the Aqua, Aura, PARASOL, and CloudSat satellites. All the satellites follow the same Sunsynchronous orbit (705 km altitude, 98° inclination) and are separated from each other by a few seconds to minutes. The sensor suite carried by the satellites in the constellation enables the first global, near simultaneous measurements of aerosols, clouds, temperature, relative humidity, and radiative fluxes. CloudSat with its cloud profiling radar (CPR) leads CALIPSO by 10–15 s, which allows for the simultaneous profiling of the same cloud systems with both the radar and LiDAR. CALIPSO attitude is controlled such that the LiDAR points 0.3° ahead of nadir in the alongtrack direction, to avoid saturating the detector with strong specular returns from calm water bodies. Based on the spacecraft orbital parameters and the transmitter characteristics, the footprint on the ground is 70 m in diameter, and adjacent spots are separated 333 m in the along-track direction.

Figures 1 and 2 (see the "High-Level Technical Overview of LiDAR" section) illustrate CALIOP's system design; its transmitter is based on two redundant, diode-pumped, Q-switched Nd:YAG lasers with a PRF of 20.16 Hz, with a nominal energy per pulse of 220 mJ (Winker et al. 2004). Part of the energy of the 1,064 nm pulses is passed through a frequency-doubling crystal to produce a 532 nm component. The energy of the transmitted pulses at both wavelengths is nominally 110 mJ and is measured before passing through a beam expander that limits the divergence to 100 μrad. The laser polarization is also controlled to be linearly polarized with a purity greater than 99 %. Backscattered photons are collected by a 1-m beryllium mirror telescope; a field stop at the telescope focus limits the receiver field of view and provides a spatial filter limiting background noise. A dichroic beam splitter separates the 1,064 and the 532 nm components; the 1,064 nm stream is filtered through a narrow band spectral filter and then is directly detected by an avalanche photodiode (APD). The 532 nm stream is passed by a double spectral and etalon filter to limit the background noise. The pure 532 nm component is then passed through a polarization beam splitter to separate the perpendicular and parallel polarization components and from there directed to separate photomultiplier tubes (PMT). For each channel, the output of the detector is amplified by two parallel amplifiers and 14bit digitizers which provide an effective 22-bit dynamic range. This dynamic range covers the expected magnitude range of the backscattering signals from molecules, aerosols, and cloud surfaces encountered in the atmosphere. Data acquisition starts when the laser pulses are estimated to be 115 km above sea level and finish at above 18.5 km below MSL; the output from the digitizers is sampled and recorded at 10 MHz (15 range bin).

CALIOP's data provides thin transects of the Earth's atmosphere that characterize the vertical distribution of atmospheric aerosols and molecules. It is a valuable complement to other type of meteorological sensors that provide information on the horizontal distribution of clouds and other atmospheric features. Figure 19 shows an example of one of such atmospheric transect as CALIPSO was on an ascending

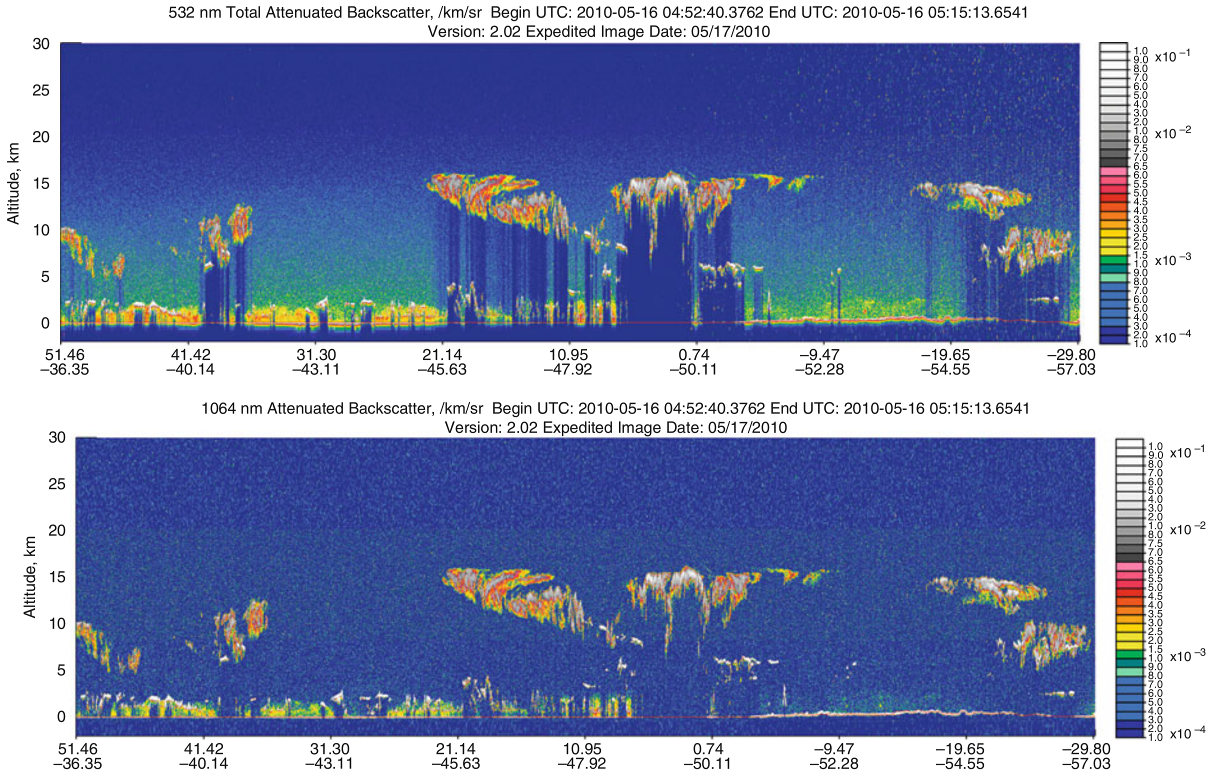


Fig. 19 Example of an atmospheric backscattering profile detected by CALIOP's 532 nm parallel + perpendicular and 1,064 nm channels (Image courtesy of NASA)

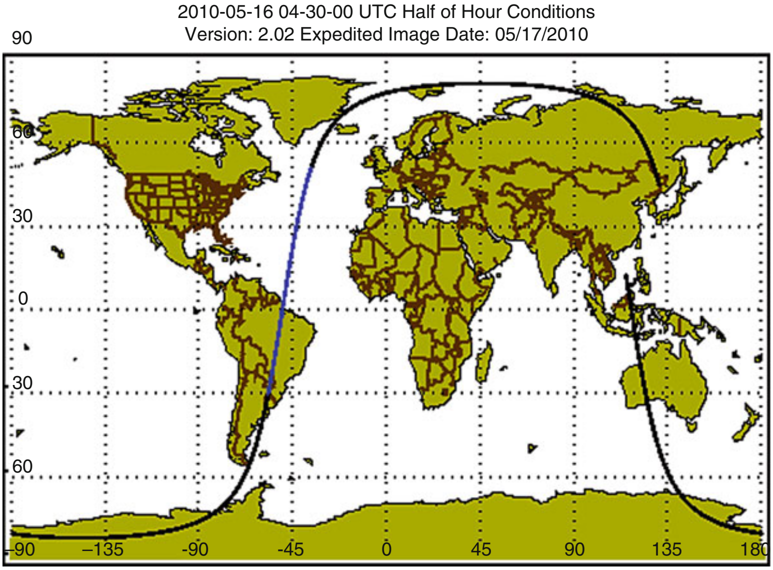


Fig. 20 Ground track for the atmospheric scattering profiles shown in Fig. 19 (Image courtesy of NASA)

pass from South America up to the North Atlantic (Fig. 20). Figure 19 illustrate the difference in detected scattering at the 532 and 1,064 nm channels. The 532 nm is the most sensitive channel due to its shorter wavelength. Figure 21 illustrates the complementary value of atmospheric LiDAR to other forms of passive remote sensing; overlaid over an AQUA MODIS image is the ground track of the CALIOP profile. Complementing the horizontal cloud structure from the MODIS image, CALIOP data shows the vertical cloud structure, including an ash plume produced by the eruption of the Eyjafjallajökull volcano in May 2010.

An interesting implementation of atmospheric LiDAR occurred in 2008, when a ground-based atmospheric LiDAR was deployed and made successful measurements of the Martian atmosphere for 152 days. The LiDAR was part of the meteorological station (MET) onboard the Phoenix Mars Lander that was launched from the Earth on August 4, 2007, landed on Mars on May 25, 2008, and collected and transmitted scientific information until October 29, 2008 (a total of 152 Martian days) (Whiteway et al. 2011). What is outstanding about this LiDAR system is the degree of miniaturization that was achieved. The entire unit had a total mass of 6 kg. The transmitter was based on a single diode-pumped, Q-switched Nd:YAG laser with a PRF of 100 Hz and a pulse width of 10 ns (Whiteway et al. 2008). Part of the energy of the 1,064 nm pulses was passed through a frequency-doubling crystal to produce a 532 nm component. The pulse energy was 0.3 mJ at 1064 nm and 0.4 mJ at the 532 nm. The divergence of the laser beams was 250 μrad. The backscattered photons were collected by a 10-cm diameter reflective telescope, separated into the two spectral components by a dichroic mirror. The 1,064 beam was filtered through a

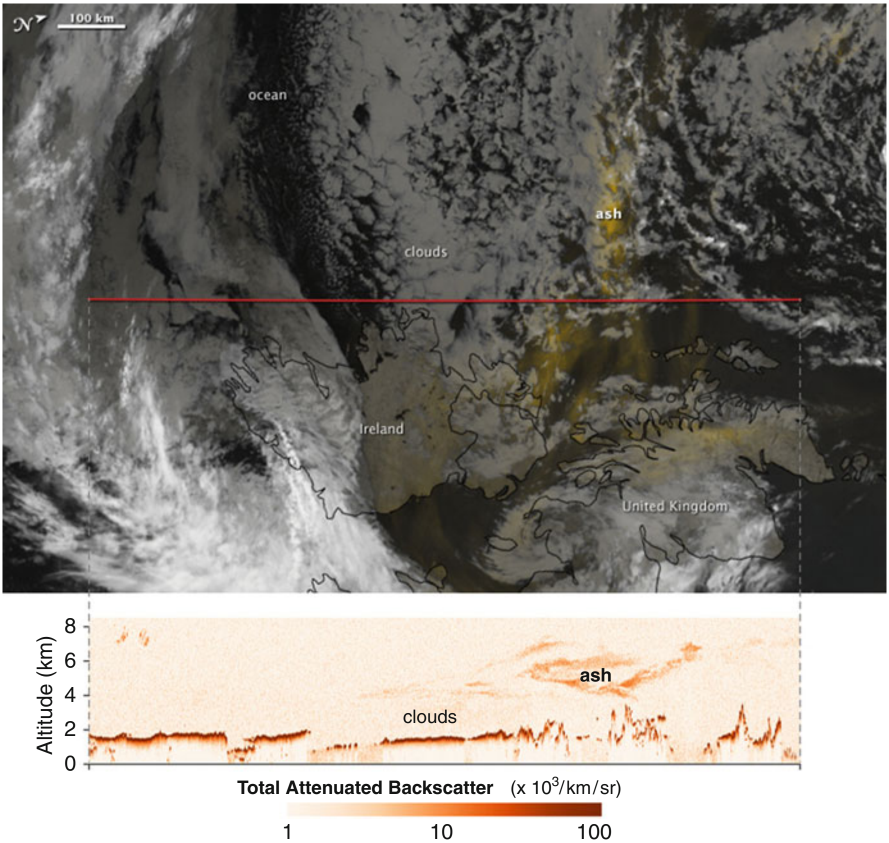
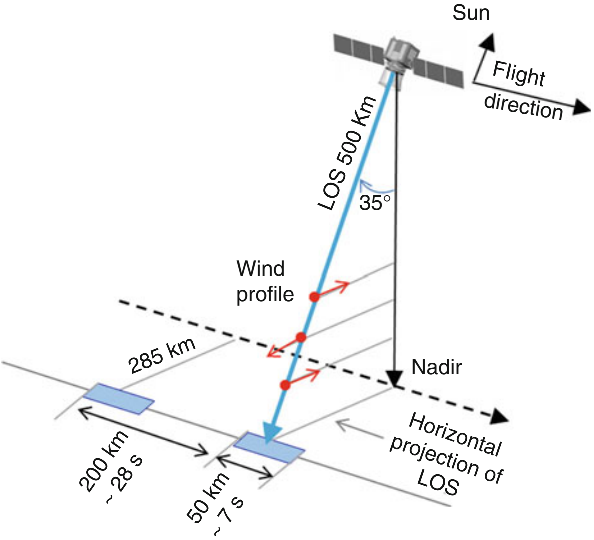


Fig. 21 Horizontal cloud structure from AQUA MODIS and vertical profile from CALIOP showing the ash plume from the Eyjafjallajökull volcano (Image courtesy of NASA)

2 nm interference filter and detected by a Si APD working in analog mode. The 532 beam was passed through a 1 nm interference filter, limited by a field stop and detected by a PMT, whose signal was collected in both analog and photon-counting modes. The analog output was recorded with a 14-bit amplitude at a 30 MHz sampling frequency (333 µs per bin). Analog detection was used for backscattering below 10 km, while photon counting was used to detect weak signals from backscattering up to 20 km.

A future satellite, ADM-Aeolus scheduled for launch in 2013, will carry the first atmospheric LiDAR to be used for the remote determination of global wind speed profiles. Along with temperature, pressure, and humidity, wind velocities are the basic variables used to describe the state of the atmosphere, and the knowledge of global circulation is crucial for the improvement of global climate models. The Atmospheric Laser Doppler Instrument (ALADIN) onboard Aeolus is designed and constructed as a direct detection Doppler LiDAR. The operation principle of the

Fig. 22 Concept for the future ALADIN Doppler wind LiDAR



instrument, illustrated in Fig. 22, consists of detecting Mie and Rayleighscattering by aerosols and atmospheric molecules and using a high-resolution spectrometer to measure the wavelength shift of the backscattered radiation with respect to that emitted by the laser transmitter (Ansmann et al. 2007). The wavelength shift is proportional to the relative velocity along the line of sight (LOS) between the satellite and the scattering particles. By taking into account the spacecraft motion and the Earth's rotation, it is possible to isolate the wind velocity. The satellite is planned to orbit at a 400 km altitude (7.21 km/s ground speed), and the ALADIN will point 35° off-nadir in the across-track direction. The Doppler shift and thus the wind speed are to be determined at different ranges (heights) along the LOS, and the wind horizontal component perpendicular to the satellite ground track will be projected from the slanted vector. Mission requirements call for the wind measurements to be averaged across 50 km cells, and average measurements are to be obtained about 200 km apart. To achieve these, the LiDAR will operate in burst modes, transmitting continuous burst for 7 s every 28 s (Ansmann et al. 2007).

ALADIN has a monostatic design, i.e., the transmit and receive paths go through the same telescope (Ansmann et al. 2007). The telescope is an afocal Cassegrain design with a diameter of 1.5 m and its field of view of only 12 μrad, which produce a footprint of 12–15 m at the end of the 500 km LOS. The optical transmitter is based on a single-mode, diode-pumped, Q-switched, frequency-tripled Nd:YAG laser. The output laser pulse in the ultraviolet range has a wavelength of 355 nm with a pulse width of 30 ns, energy per pulse of 120 mJ, and a planned PRF of 100 Hz. The spectrally pure laser pulses are passed through linear and circular polarizers before being expanded through the telescope. The backscattered photons are collected by the telescope and passed through the polarizers; only the parallel polarized components are accepted and passed through a field stop and 1 nm spectral filter to limit the effect of background illumination. Once the return beam is spatially and spectrally filtered, it is directed to the spectrometer system which is comprised of a Fizeau interferometer, which detects the spectrally narrow Mie backscattered peak (channel 1) and two Fabry–Perot etalons (channels 2 and 3), to detect the wide Rayleigh–-Brillouin backscatter spectrum. The output of the spectrometers is detected by two accumulation charged-coupled devices (ACCD).

Besides the described spaceborne atmospheric LiDARs, both NASA and ESA are currently contemplating future mission that would incorporate atmospheric LiDAR instruments. Currently in the design phase, the joint ESA JAXA Earth Clouds, Aerosols, and Radiation Explorer (EarthCARE) is aimed at improving our understanding of the interactions between cloud, radiative, and aerosol processes. EarthCARE proposes a suite of atmospheric instruments which includes a cloud profiling radar (CPR), multispectral imager (MSI), a broadband radiometer (BBR), and an atmospheric backscattering and depolarization LiDAR (ATLID). ALTLID is envisioned to be an ultraviolet high spectral resolution backscattering LiDAR (Le Hors et al. 2008), much like an upgraded version of CALIOP. On the Decadal Survey, the NRC recommended to NASA the design and implementation of three missions that incorporate atmospheric LiDARs (National Research Council 2007). The most immediate is the Active Sensing of CO_2 Emissions over Nights, Days, and Seasons (ASCENDS), which is envisioned to incorporate a multiwavelength LiDAR system. The Aerosol-Cloud-Ecosystems (ACE) mission, with a primary goal to reduce uncertainty about climate forcing in aerosol–cloud interactions and ocean ecosystem carbon dioxide (CO_2) uptake, will incorporate an atmospheric backscattering LiDAR, a multiangle polarimeter, and a Doppler radar. Finally a demonstration mission is recommended, the 3D-Winds, which should incorporate a Doppler LiDAR to map tropospheric winds for weather forecasting and pollution transport modeling.

Guidance, Navigation, Control, and Inspection

The most recent application of LiDAR in spaceborne platforms is for on-orbit operations such as ranging for rendezvous and docking, active imaging for inspection and servicing, and robot vision for autonomous operation. The first use of LiDAR technology for semiautonomous/autonomous vehicle operation was for the Mars Pathfinder microrover “Sojourner,” which landed on Mars on July 4, 1997, and operated until September 27, 1997, when the communication with the Lander was suddenly lost. The Sojourner microrover was equipped with a stereo-pair imaging system for rover navigation. To aid the camera in proximity operations, a laser triangulation system was included which consisted of five semiconductor diode laser stripe projectors (JPL 1997). Using preflight calibration tables, the system was able to determine distances from the rover to the projected laser stripes based on the pixel position on which the laser spots were detected.

The increasing need for active imaging for on-orbit inspection and servicing has created another application of LiDAR technology. This need was extremely evident

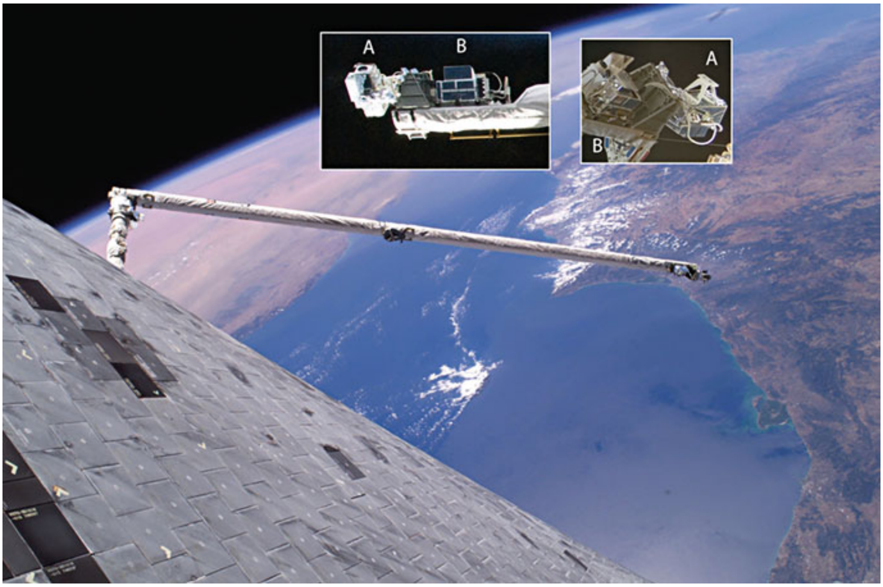


Fig. 23 The shuttle orbiter boom and sensor system (OBSS), *inset* images show close-ups of the laser dynamic range imager (LDRI) (**a**) and the laser camera system (LCS) (**b**) (Images courtesy of NASA)

after the tragic loss of the Columbia Shuttle during atmospheric reentry on February 1, 2003. Among the many improvements to the systems and procedures of the space shuttle program during the return-to-flight effort was to include on every future mission tools for on-orbit inspection. The inspection of the critical shuttle areas such as the wing leading edge and thermal protection tiling is now performed with a suite of passive and active imaging systems mounted on the end of 15 m long boom that serves as an extension to the shuttle remote manipulator system (SRMS). The orbiter boom and sensor system (OBSS) includes three sensors: two of them, the laser dynamic range imager (LDRI) and the intensified television camera (ITVC), are mounted on a pan and tilt platform, while the third sensor, the laser camera system (LCS), is rigidly mounted on the side of the boom (Fig. 23) (NASA 2005). LDRI and LCS are active imaging sensors that use nonconventional LiDAR technology for the collection of 3D data.

LDRI was developed by Sandia National Laboratories and uses a combination of phase difference ranging and video to derive 3D information. LDRI has a laser transmitter based on a continuous wave (CW) diode laser emitting light at 805 nm with a maximum power of 12 W (Smithpeter et al. 2000). The CW amplitude (intensity) is modulated at 3.125 or 140 MHz. In contrast to most LiDAR systems, where the divergence of a laser beam is restricted, the light from the LDRI is expanded and then passed through a diffuser plate to produce a floodlight effect; different plates can yield beam spreads of 10–60°, with a normal used value of 40°. This expanded beam is used to illuminate the target; the backscattered photons are collected by a refractive lens and passed through a narrow 30 nm spectral filter to limit the contribution from external illumination and then focused on the cathode of an image intensifier tube. The optical gain of the intensifier tube is modulated with the same signal used to modulate the laser output. The output from the image intensifier is coupled by a fiber optic taper to a CDD detector which is read by a conventional 640 × 480 analog video recorder operating at 30 frames per second. To perform ranging of a given target area, the area is illuminated by the variable intensity laser for a given time on which several video frames are recorded. Assuming that each pixel of the frame is imaging the same target area and that the range remains constant throughout the different collected frames, it is possible to derive the phase difference for each pixel between the emitted and backscattered radiation by comparing the changes in intensity from several frame captures. Knowing the phase difference, it is possible to determine the range between the sensor and the target on a pixel by pixel basis, thus generating intensity and spatial datasets of the illuminated area.

LCS was developed by the Neptec Design Group of Canada and first flew into space on August 10, 2001, as a detailed test objective (DTO) during the STS-105 mission of the Space Shuttle Discovery (STS-105 Shuttle Press Kit 2001). LCS is a hybrid video and imaging LiDAR sensor; the LiDAR sensor is based on the triangulation ranging principle capable of imaging a 30° × 30° field of regard (FOR) from a range between 1 and 10 m (Dupuis et al. 2008; Deslauriers et al. 2005). The transmitter of LCS is based on a continuous wavelength-shifted Nd:YAG laser emitting at 1,500 nm. Scanning mirror/galvanometers are used to steer the laser beam in two dimensions over the FOR to illuminate the target. The reflected photons are captured by a refractive lens, filtered by a narrow band-pass spectral filter, and detected by a linear detector array (LDA). By determining the array coordinates of the pixel that detects the highest intensity signal and knowing the baseline distance and the galvanometer angles, it is possible to determine the range to the target using the triangulation principle with high precision (3 mm at 5 m) and at fast acquisition rates. LCS has been upgraded by Neptec to have a hybrid LiDAR design which combines a triangulation LiDAR operating at 1,400 nm with a time-of-flight (TOF) LiDAR operating at 1,540 nm which shares the same scanning mechanisms (Dupuis et al. 2008; English et al. 2005). This upgraded sensor also includes a thermal imager, and it is designated as TriDAR. TriDAR exploits in a synergistic approach the advantages of the TOF and triangulation ranging mechanism, combining the long-range capabilities with coarse precision of the TOF (range <3 km, <25 mm precision) with the sub-cm accuracy in the short range of the triangulation units. TriDAR first flew into space as a DTO onboard Discovery during STS-128 in August–September 2009, to demonstrate its capabilities to perform autonomous acquisition and tracking of the ISS. It also performed real-time docking measurements during the STS-31 mission in April 2010.

An additional space-based ranging and imaging LiDAR was carried by the Air Force XSS-11 satellite which operated between 2005 and 2007. The rendezvous laser system (RLS) sensor, also referred to as Spaceborne Scanning Lidar System (SSLS), was designed and manufactured by Optech and MDA as a system to allow

XSS-11 to perform autonomous rendezvous and proximity maneuvers (Nimelman et al. 2006; Dupuis et al. 2008). RLS was a time-of-flight scanning LiDAR with a 20° × 20° field of view, a laser beam divergence of 500 μrad, a maximum range of 5 km with a resolution of 1 cm, and an accuracy of 5 cm. During its 22-month operations, XSS-11 used RLS to perform rendezvous and proximity operation around its expended Minotaur launch vehicle and with several US-owned dead or inactive resident space objects.

Conclusion

The entire books have been written on the subject of LiDAR remote sensing from specific points of view. This chapter is meant to provide a broad overview of LiDAR technology, highlighting the most common applications from spaceborne platforms. It describes the versatility of LiDAR, not only as a remote sensing technique but also as a method of enabling and supporting other remote sensing techniques and satellite applications. LiDAR, despite originating roughly at the same time as Radar, is not yet as mature as Radar or other forms of remote sensing. However, exponential development of its enabling technologies (lasers, photodetectors, positioning, and attitude sensor) as well as LiDAR data processing algorithms over the last two decades is speeding its maturation process. As is the case with any other technology, further technical developments will enable new applications, even when there is much room for the development of LiDAR on its own, and a great deal of progress is also expected from a synergistic approach of combining it with other forms of active and passive remote sensing techniques.

Cross-References

- ▶ Astronaut Photography: Handheld Camera Imagery from Low Earth Orbit
- ▶ Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing
- ▶ Electro-Optical and Hyperspectral Remote Sensing
- ▶ Introduction and History of Space Remote Sensing
- ▶ Operational Applications of Radar Images

References

W. Abdalati, H.J. Zwally, R. Bindschadler, B. Csatho, S.L. Farrell, H.A. Fricker, D. Harding, R. Kwok, M. Lefsky, T. Markus, A. Marshak, T. Neumann, S. Palm, B. Schutz, B. Smith, J. Spinhirne, C. Webb, The ICESat-2 laser altimetry mission. Proc. IEEE **98**(5), 735–751 (2010)

J.B. Abshire, X. Sun, H. Riris, J.M. Sirota, J.F. McGarry, S. Palm, D. Yi, P. Liiva, Geoscience laser altimeter system (GLAS) on the ICESat mission: on-orbit measurement performance. Geophys. Res. Lett. **32**, L21S02 (2005)

C.O. Alley, P.L. Bender, R.F. Chang, D.G. Currie, R.H. Dicke, J.E. Faller, W.M. Kaula, G.J.F. MacDonald, J.D. Mulholland, H.H. Plotkin, S.K. Poultney, D.T. Wilkingson, I. Winer, W. Carrion, T. Johnson, P. Spadin, L. Robinson, E. Joseph Wampler, D. Wiebrr, E. Silverberg, C. Steggerda, J. Mullendore, J. Bayner, W. Williams, B. Warner, H. Richardson, B. Bopp, Laser ranging retroreflector. Section 7, of Apollo 11 Preliminary Science Report. NASA SP 214 (1969)

A. Ansmann, U. Wandinger, O. Le Rille, D. Lajas, A.G. Straume, Particle backscatter and extinction profiling with the spaceborne high-spectral-resolution Doppler lidar ALADIN: methodology and simulations. Appl. Opt. **46**(26), 6606–6622 (2007)

M.D. Behn, M.T. Zuber, A comparison of ocean topography derived from the Shuttle Laser Altimeter-01 and TOPEX/POSEIDON. IEEE Trans. Geosci. Remote Sens. **38**(3), 1425–1438 (2000)

P.L. Bender, D.G. Currie, R.H. Dickey, D.H. Eckhardt, J.E. Faller, W.M. Kaula, J.D. Mulholland, H.H. Plotkin, S.K. Poultney, E.C. Silverberg, D.T. Wilkinson, J.G. Williams, C.O. Alley, The lunar ranging experiment. Science **182**(4109), 229–238 (1973). New Series

J.L. Bufton, Laser altimetry measurements from aircraft and spacecraft. Proc. IEEE **77**(3), 463–477 (1989)

C. Carabajal, D.J. Hardin, S.B. Luthcke, W. Fong, S.C. Rowton, J.J. Frawley, Processing of shuttle laser altimeter range and return pulse data in support of SLA-02, in *Proceedings of the ISPRS Workshop Mapping Surface Structure and Topography by Airborne and Spaceborne Lasers*, Portland, 1999

W.E. Carter, The lunar laser ranging pointing problem. Unpublished doctoral dissertation, University of Arizona, Tucson, 1973

W.E. Carter, R.L. Shrestha, K.C. Slatton, Geodetic laser scanning. Phys. Today **60**(12), 41–49 (2007)

J.F. Cavanaugh, J.C. Smith, X. Sun, A.E. Bartels, L. Ramos-Izquierdo, D.J. Krebs, J.F. McGarry, R. Trunzo, A.M. Novo-Gradac, J.L. Britt, J. Karsh, R.B. Katz, A.T. Lukermire, R. Szymkiewicz, D.L. Berry, J.P. Swinski, G.A. Neumann, M.T. Zuber, D. Smith, The Mercury laser altimeter instrument for the MESSENGER mission. Space Sci. Rev. **131**(1–4), 451–479 (2007)

M.L. Chanin, A. Hauchecorne, C. Malique, D. Nedeljkovic, J.E. Blamont, M. Desbois, G. Tulinov, V. Melnikov, Premiers résultats du lidar Alissa embarqué à bord de la station Mir. C.R. Acad. Sci. Ser. IIA Earth Planet Sci. **328**(6), 359–366 (1999)

T.D. Colea, A.F. Chenga, M. Zuberb, D. Smith, The laser rangefinder on the near earth asteroid rendezvous spacecraft, in *Acta Astronautica. Second IAA International Conference on Low-Cost Planetary Missions*, Laurel, vol. 39, Issue no 1–4, pp. 303–313, July–Aug 1996

T.K. Cossio, K.C. Slatton, W.E. Carter, K.Y. Shrestha, D. Harding, Predicting small target detection performance of low-SNR airborne LiDAR. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. **3**(4), 672–688 (2010)

J.J. Degnan, *30 Years of SLR (invited paper), Proceedings of the 9th International Workshop on Laser Ranging Instrumentation, Australian Government Publishing Service*, Canberra, p. 8, 1994, http://ilrs.gsfc.nasa.gov/docs/ThirtyYearsOfSatelliteLaserRanging.pdf

A. Deslauriers, I. Showalter, A. Montpool, R. Taylor, I. Christie, Shuttle TPS inspection using triangulation scanning technology, in *Spaceborne sensors II. Proceedings of the SPIE,* Orlando, Florida, USA, vol. 5798, pp. 26–33, 2005

J.O. Dickey, P.L. Bender, J.E. Faller, X.X. Newhall, R.L. Ricklefs, J.G. Ries, P.J. Shelus, C. Veillet, A.L. Whipple, J.R. Wiant, J.G. Williams, C.F. Yoder, Lunar laser ranging: a continuing legacy of the Apollo program. Science **265**(5171), 482–490 (1994). New Series

A. Donnellan, P. Rosen, J. Ranson, H. Zebker, Deformation, ecosystem structure, and dynamics of ice (DESDynI), in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, Honolulu, 2008

E. Dupuis, J.C. Piedboeuf, E. Martin, Canadian activities in intelligent robotic systems: an overview, in *Proceedings of International Symposium on Artificial Intelligence, Robotics and Automation in Space*, Hollywood, Feb 2008

Y. Durand, A. Hélière, P. Bensi, J.-L. Bézy, R. Meynart, Lidars in ESA's earth explorer missions, in *14th Coherent Laser Radar Conference*, Snowmass, 2007

C. English, S. Zhu, C. Smith, S. Ruel, I. Christie, Tridar: a hybrid sensor for exploiting the complimentary nature of triangulation and LIDAR technologies, in *The 8th International Symposium on Artificial Intelligence, Robotics and Automation in Space*, ed. by B. Battrick. ESA SP-603 (European Space Agency, München, 2005)

J.C. Fernandez-Diaz, Scientific applications of the mobile terrestrial laser scanner (M-TLS) system, M.S. thesis, Department of Civil Engineering, University of Florida, Gainesville, 2007, http://purl.fcla.edu/fcla/etd/UFE0021101. Accessed Feb 2011

G. Fiocco, L.D. Smullin, Detection of scattering layers in the upper atmosphere (60–140 km) by optical radar. Nature **199**, 1275–1276 (1963)

J. Garvin, J. Bufton, J. Blair, D. Harding, S. Luthcke, J. Frawley, D. Rowlands, Observations of the Earth's topography from the shuttle laser altimeter (SLA): laser-pulse echo-recovery measurements of terrestrial surfaces. Phys. Chem. Earth **23**(9–10), 1053–1068 (1998)

Hamamatsu Corporation, Photon counting using photomultiplier tubes (2005), http://sales.hamamatsu.com/assets/applications/ETD/PhotonCounting_TPHO9001E04.pdf

D.J. Harding, D.B. Gesch, C.C. Carabajal, S.B. Luthcke, Application of the shuttle laser altimeter in an accuracy assessment of GTOP30, a global 1-kilometer digital elevation model, in *Proceedings of the ISPRS Workshop Mapping Surface Structure and Topography by Airborne and Spaceborne Lasers*, Portland, Nov 1999

D.W. Harris, J.H. Berbert, NASA/MOTS optical observations of the ANNA 1B satellite, NASA Technical Note D-3174, Jan 1966

W.A. Heiskanen, H. Moritz, *Physical Geodesy* (Freeman, San Francisco, 1967)

E.O. Hulburt, Observations of a searchlight beam to an altitude of 28 kilometers. J. Opt. Soc. Am. **27**, 377–382 (1937)

A. Javan, W.R. Bennett, D.R. Herrott, Population inversion and continuous optical maser oscillation in a gas discharge containing a He–Ne mixture. Phys. Rev. Lett. **6**, 106–110 (1961)

E.A. Johnson, R.C. Meyer, R.E. Hopkins, W.H. Mock, The measurement of light scattered by the upper atmosphere from a search-light beam. J. Opt. Soc. Am. **29**, 512–517 (1939)

JPL, Rover camera instrument description (1997), http://starbase.jpl.nasa.gov/mpfr-m-rvreng-2_3-edr_rdr-v1.0/mprv_0001/document/rcinst.htm. Accessed Feb 2011

K. Kaufmann, Choosing your detector, OE Mag., Mar 2005

W.M. Kaula, G. Schubert, R.E. Lingenfelter, W.L. Sjogren, W.R. Wollenhaupt, Apollo laser altimetry and inferences as to lunar structure, in *Lunar Science Conference*, Houston, 18 Mar 1974, Proceedings, vol. 3, (A75-39540 19-91) (Pergamon Press, New York, 1974), pp. 3049–3058

L. Le Hors, Y. Toulemont, A. Hélière, Design and development of the backscatter Lidar Atlid for earthcare, in *Proceedings of the International Conference on Space Optics*, Toulouse, 2008

G.G. Matvienko, V.E. Zuev, V.S. Shamanaev, G.P. Kokhanenko, A.M. Sutormin, A.I. Buranskii, S.E. Belousov, A.A. Tikhomirov, Lidar BALKAN-2 for the space platform ALMAZ-1B, Lidar techniques for remote sensing, in *Proceedings of SPIE*, vol. 2310, 1994, http://spie.org/x648.html?product_id=195859

F.J. McClung, R.W. Hellwarth, Giant optical pulsations from Ruby. J. Appl. Phys. **33**(3), 828–829 (1962)

J. McGarry, T. Zagwodzki, A brief history of satellite laser ranging: 1964 – present. Published by the Crustal Dynamics Data Information System (CDDIS), NASA Goddard Space Flight Center, Greenbelt (2005), http://cddis.gsfc.nasa.gov/ngslr/docs/gsfcslr_jm0504.pdf. Accessed Feb 2011

J.K. Miller, A.S. Konopliv, P.G. Antreasian, J.J. Bordi, S. Chesley, C.E. Helfrich, W.M. Owen, T.C. Wang, B.G. Williams, D.K. Yeomans, D.J. Scheeres, Determination of shape, gravity and rotational state of Asteroid 433 Eros. Icarus **155**, 3–17 (2002)

NASA, Mars topography, http://mola.gsfc.nasa.gov/topography.html

NASA, The space shuttle's return to flight, mission STS-114 press kit (2005)

National Research Council, Earth science and applications from space: national imperatives for the next decade and beyond, in *Committee on Earth Science and Applications from Space: A Community Assessment and Strategy for the Future* (The National Academies Press, Washington, DC, 2007). ISBN 0-309-10387-8

NEPTEC, Tridar, http://www.neptec.com/media/brochures/Canadian/Space-TriDAR.pdf

M. Nimelman, J. Tripp, A. Allen, D.M. Hiemstra, S.A. McDonald, Spaceborne scanning lidar system (SSLS) upgrade path. Proc. SPIE **6201**, 62011V-1–62U11V-10 (2006)

J.C. Piedboeuf, E. Martin, M. Doyon, On-orbit servicing in Canada: advanced developments and demonstrations, in *Proceedings of the 8th ESA Workshop on Advanced Space Technologies for Robotics and Automation ASTRA*, Noordwijk, 2004

L. Ramos-Izquierdo, V.S. Scott III, J. Connelly, S. Schmidt, W. Mamakos, J. Guzek, C. Peters, P. Liiva, M. Rodriguez, J. Cavanaugh, H. Riris, Optical system design and integration of the Lunar orbiter laser altimeter. Appl. Opt. **48**, 3035–3049 (2009)

F.I. Robertson, W.M. Kaula, Apollo 15 laser altimeter, Part D, Section 25, Apollo 15 Preliminary Science Report. NASA SP-289, 1972

B.E. Schutz, H.J. Zwally, C.A. Shuman, D. Hancock, J.P. DiMarzio, Overview of the ICESat mission. Geophys. Res. Lett. **32**, L21S01 (2005)

J. Shan, C.K. Toth (eds.), *Topographic Laser Ranging and Scanning: Principles and Processing* (CRC Press, Boca Raton, 2009)

H. Simons, Secret mapping by satellite, New Sci. **21**(381) (1964)

D.E. Smith, M.T. Zuber, G.A. Neumann, F.G. Lemoine, Topography of the Moon from the Clementine lidar. J. Geophys. Res. **102**(E1), 1591–1611 (1997)

D.E. Smith, M.T. Zuber, H.V. Frey, J.B. Garvin, J.W. Head, D.O. Muhleman, G.H. Pettengill, R.J. Phillips, S.C. Solomon, H.J. Zwally, W.B. Banerdt, T.C. Duxbury, M.P. Golombek, F.G. Lemoine, G.A. Neumann, D.D. Rowlands, O. Aharonson, P.G. Ford, A.B. Ivanov, C.L. Johnson, P.J. McGovern, J.B. Abshire, R.S. Afzal, X. Sun, Mars orbiter laser altimeter: experiment summary after the first year of global mapping of Mars. J. Geophys. Res. **106**(E10), 23689–23722 (2001)

D.E. Smith, M.T. Zuber, G.A. Neumann, F.G. Lemoine, E. Mazarico, M.H. Torrence, J.F. McGarry, D.D. Rowlands, J.W. Head III, T.H. Duxbury, O. Aharonson, P.G. Lucey, M.S. Robinson, O.S. Bamouin, J.F. Cavanaugh, X. Sun, P. Liiva, D. Mao, K.C. Smith, A.E. Bartels, Initial observations from the Lunar Orbiter Laser Altimeter (LOLA). Geophys. Res. Lett. **37**, L18204 (2010)

C.L. Smithpeter, R.O. Nellums, S.M. Lebien, G. Studor, Miniature high-resolution laser radar operating at video rates. Proc. SPIE **4035**, 279–286 (2000)

L.D. Smullin, G. Fiocco, Optical echoes from the Moon. Nature **194**, 1267 (1962)

STS-105 Shuttle press kit (2001), http://www.shuttlepresskit.com/sts-105/index.htm

G. Suna, K.J. Ranson, V.I. Kharuk, K. Kovacs, Validation of surface height from shuttle radar topography mission using shuttle laser altimeter. Remote Sens. Environ. **88**(4), 401–411 (2003)

E.H. Synge, A method of investigating the higher atmosphere. Philos. Mag. **9**, 1014–20 (1930)

N. Taylor, *LASER: The Inventor, the Nobel Laureate, and the Thirty-Year Patent War* (Simon & Schuster, New York, 2000)

The International Laser Ranging Service, http://ilrs.gsfc.nasa.gov/. Accessed Feb 2011

N. Thomasa, T. Spohnb, J.-P. Barriotc, W. Benza, G. Beutlerd, U. Christensene, V. Dehantf, C. Fallnichg, D. Giardinih, O. Groussini, K. Gundersona, E. Hauberb, M. Hilchenbache, L. Iessj, P. Lamyk, L.-M. Laral, P. Lognonnem, J.J. Lopez-Morenol, H. Michaelisb, J. Oberstb, D. Resendesn, J.-L. Reynaudk, R. Rodrigol, S. Sasakio, K. Seiferlina, M. Wieczorekm, J. Whitbya, The BepiColombo laser altimeter (BELA): concept and baseline design. Planet. Space Sci. **55**(10), 1398–1413 (2007)

M.A. Tuve, E.A. Johnson, O.R. Wulf, A new experimental method for study of the upper atmosphere. J. Terrest. Magnet. **40**, 452–454 (1935)

J.R. Vetter, Fifty years of orbit determination, development of modern astrodynamic methods. J. Hopkins APL Tech. Dig. **27**(3), 239–252 (2007)

U. Wandinger, Introduction to Lidar, in *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*, ed. by C. Weitkamp (Springer, New York, 2005), pp. 1–18

M.J. Weber (ed.), *Handbook of Lasers* (CRC Press, Baco Raton, 2001). ISBN 978-1-4200-5017-2

C. Weitkamp (ed.), *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere* (Springer, New York, 2005)

C. Werner, G. Kokhanenko, G. Matvienko, V. Shamanaev, Y. Grachjov, I. Znamenskii, U.G. Oppel, Spaceborne laser rangefinder “LORA” used as a cloud lidar. Opt. Rev. **2**(3), 221–224 (1995)

C. Werner, Spaceborne lidar mission, past and future, in *Proceedings Conference on Lasers and Electro-optics Europe, CLEO/Europe*, Hamburg, Sep 1996, p. 212

J. Whiteway, L. Komguem, C. Dickinson, Observations of Mars atmospheric dust and clouds with the Lidar instrument on the phoenix mission, in *Abstract on the Forth International Workshop on the Mars Atmosphere: Modeling and Observations*, Paris, Feb 2011

J. Whiteway, J.M. Daly, A. Carswell, T. Duck, C. Dickinson, L. Komguem, C. Cook, Lidar on the Phoenix mission to Mars. J. Geophys. Res. **113**, E00A08 (2008)

D.M. Winker, R.H. Couch, M.P. McCormick, An overview of LITE: NASA’s Lidar in-space technology experiment. Proc. IEEE **84**(2), 164–180 (1996)

D.M. Winker, W.H. Hunt, C.A. Hostetler, Status and performance of the CALIOP lidar, in *Laser Radar Techniques for Atmospheric Sensing (Proceedings of the SPIE)*, ed. by U.N. Singh, vol. 5575, Maspalomas/Gran Canaria, 2004, pp. 8–15

W.R. Wollenhaupt, W.L. Sjogren, Apollo 16 laser altimeter, Chapter 30, Part A, Apollo 16 Preliminary Science Report SP-315, 1972

W.R. Wollenhaupt, W.L. Sjogren, R.E. Lingenfelter, G. Schubert, W.M. Kaula, Apollo 17 laser altimeter, Chapter 33, Part E, Apollo 17 Preliminary Science Report SP-330, 1973

A.W. Yua, M.A. Stephen, S.X. Li, G.B. Shawa, A. Seasa, E. Dowdyea, E. Troupakib, P. Liivab, D. Pouliosc, K. Mascetti, Space laser transmitter development for ICESat-2 mission. Proc. SPIE **7578**, 757–809 (2010)

M.T. Zuber, D.E. Smith, A.F. Cheng, J.B. Garvin, O. Aharonson, T.D. Cole, P.J. Dunn, Y. Guo, F.G. Lemoine, G.A. Neumann, D.D. Rowlands, M.H. Torrance, The shape of 433 Eros from the NEAR-shoemaker laser rangefinder. Science **289**, 2097 (2000)

M.T. Zuber, D.E. Smith, S.C. Solomon, R.J. Phillips, S.J. Peale, J.W. Head III, S.A. Hauck II, R.L. McNutt Jr., J. Oberst, G.A. Neumann, F.G. Lemoine, X. Sun, O. Barnouin-Jha, J.K. Harmon, Laser altimeter observations from MESSENGER’s first Mercury Flyby. Science **321**, 77 (2008)

Fundamentals of Remote Sensing Imaging and Preliminary Analysis

Siamak Khorram, Stacy A. C. Nelson, Cynthia F. van der Wiele, and Halil Cakir

Contents

S. Khorram (✉)
Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA

Center for Geospatial Analytics, North Carolina State University, Raleigh, NC, USA
e-mail: khorram@berkeley.edu

S.A.C. Nelson
Department of Forestry and Environmental Resources North Carolina State University Raleigh, NC, USA

Center for Geospatial Analytics North Carolina State University Raleigh, NC, USA
e-mail: Stacy_Nelson@ncsu.edu

C.F. van der Wiele
Region IV NEPA Program Office, US Environmental Protection Agency, Research Triangle Park, NC, USA
e-mail: cynthia.vanderwiele@alumni.duke.edu

H. Cakir
Air Quality Analysis Group/AQAD/OAQPS, US Environmental Protection Agency, Research Triangle Park, NC, USA
e-mail: Cakir.Halil@epa.gov

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_46

Abstract

Airborne and satellite digital image acquisition, preprocessing, and data reduction techniques as applied to remotely sensed data for the purpose of extracting useful Earth resources information are discussed in this chapter. The image processing and postprocessing tools are described in the next chapter. The concepts discussed in this chapter include:

- Image acquisition considerations including currently available remotely sensed data
- Image characteristics in terms of spatial, spectral, radiometric, and temporal resolutions
- Preprocessing techniques such as geometric distortion removals, atmospheric correction algorithms, image registration, enhancement, masking, and data transformations
- Data reduction, fusion, and integration techniques
- International policies governing acquisition and distribution of remotely sensed data

Keywords

Remote sensing data acquisition • Data fusion • Digital image processing • Digital image data integration • Electromagnetic spectrum • Hyperspectral imaging • Light detection and ranging (LiDAR) • Multispectral imaging • Radio detection and ranging (RADAR) • Radiometric resolution • Satellite remote sensing • Spatial resolution • Spectral resolution • Temporal resolution • Geospatial data integration

Introduction

Remote sensing involves the collection of data in digital or analog forms (e.g., aerial photographs and videos) by space-based instruments or sensors without any physical contact. Remote sensing can be defined as the acquisition and measurement of data/information of one or more properties of a phenomenon, object, or material by a recording device not in physical contact with the feature(s) under surveillance.

In general, remotely sensed data are collected from airborne platforms (e.g., aircraft and satellites), with reflectance or emission values at various wavelength regions of the electromagnetic spectrum in a variety of spectral, spatial, radiometric, and temporal resolutions. Most remotely sensed data is collected in digital form. However, analog products can be digitized (with certain limitations), that can be then be processed digitally.

Airborne and satellite data are collected in raw or unprocessed form. The raw spectral data or radio signals must be processed or enhanced in order to produce images or other products. This typically involves computer-aided digital processing, although visual image interpretation is still used for some applications. The advantages of digital processing include: (1) the ability to go well beyond the limits of the human

Fig. 1 Digitally processed image of Earth (Courtesy of NASA)

eye, which is up to 32 gray levels; most data are in 1024 gray levels or higher; (2) the speed of processing and extracting information over large areas; (3) the lower cost for processing remotely sensed data; (4) the repeatability of producing the same results when using the same algorithms; and (5) the nature of having all inputs, outputs, and intermediate results in digital form, which makes it possible to combine a variety of data types for solving complex problems and performing geospatial modeling.

The spectacular illustration in Fig. 1, reminiscent of the famous Apollo-era images of Earth taken by lunar astronauts, is of the Western Hemisphere during one of the strongest hurricanes ever recorded in the Eastern Pacific. It was produced through the utilization of digital image processing techniques. This combination of science, engineering, and artistry was generated by researchers in the Laboratory for Atmospheres at NASA's Goddard Space Flight Center, Greenbelt, Maryland, by digital processing of data from three different Earth observing satellite instruments.

Another example of the complexity involved in digital image processing is of the Earth rising above the lunar horizon, as recently captured by NASA's Lunar Reconnaissance Orbiter (LRO). This unique view of Earth from the spacecraft's vantage point in orbit around the moon was created on October 12, 2015. The LRO was rolled 67° to the side, then coordinated with the direction of travel to maximize the width of

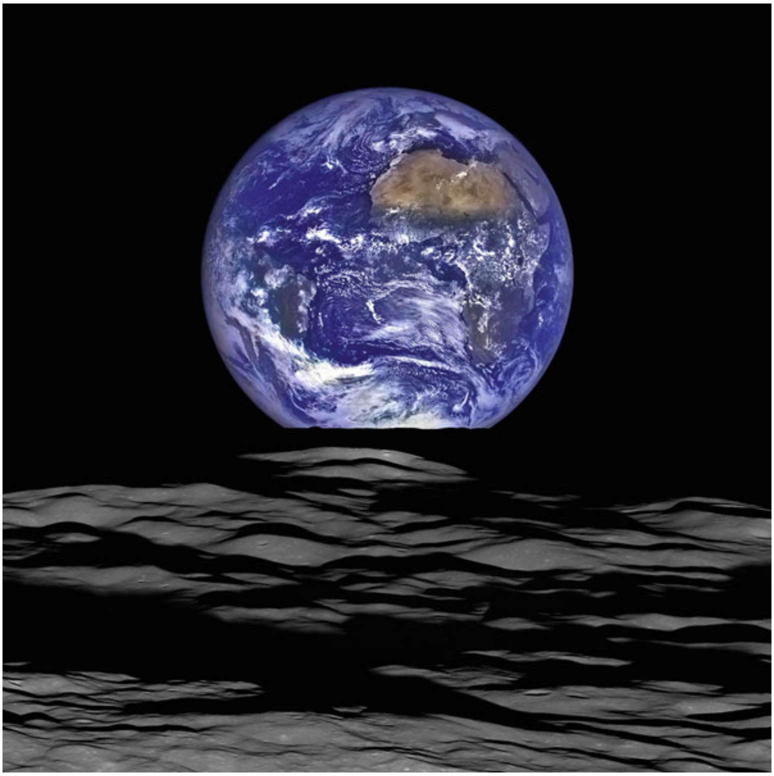


Fig. 2 The Earth rise as captured from NASA's Lunar Reconnaissance Orbiter (LRO) from the spacecraft's vantage point in orbit around the moon (Image courtesy of NOAA's Environmental Visualization Laboratory)

the lunar horizon in the narrow angle camera (NAC) image. The LRO was traveling over 1600 m/s relative to the lunar surface below. Consequently, due to the three motions and the fact that the NAC is a line scanner, the raw image geometry is distorted. In addition, because of the substantial distance between the Earth and the Moon, the geometric correction is different for each (Robinson 2015). As said by Noah Petro, Deputy Project Scientist for LRO at NASA's Goddard Space Flight Center, "The image is simply stunning; the image of the Earth evokes the famous 'Blue Marble' image taken by Astronaut Harrison Schmitt during the Apollo 17 mission, 47 years ago, which also showed Africa prominently in the picture" (Fig. 2).

Many scientists and engineers have contributed significantly to advancing image processing techniques as applied to Earth resources, planetary explorations and studies, and medical imaging and to their scientific definition. These include the various individuals and their relevant books and articles as referenced in end notes (Swain and Davis 1978; Dobson et al. 1995; Jähne 1991; Dai and Khorram 1997, Khorram et al. 2012; Lunetta and Elvidge 2000; Russ 2002; Herold et al. 2003; Jensen 2005; Lillesand et al. 2008; Yuan et al. 2009).

Thematic maps, color composites, and color-coded classified images are some of the most characteristic output products from digital formats. A variety of image enhancement and visualization techniques can be applied to these output products in order to extract useful information. Fortunately, many commercial software packages have been developed specifically for processing satellite imagery.

Data Acquisition and Policies

Satellite and Airborne Data Resolution and Formats

Over the last decade, the number of airborne and spaceborne remote sensing platforms and sensors has increased dramatically. While airborne platforms have emerged to become much more customizable by providing application-specific tools, remote sensing satellites (spaceborne) offer the distinct advantage of synoptic acquisition of large and/or remote areas with repetitive revisit cycles.

Airborne platforms normally involve the capture of features on the Earth's surface by means of photographic cameras (air photos), video, spectroradiometers, radio detection and ranging (RADAR) apparatus, and LiDAR (light detection and ranging). Film- and video-based air photos normally fall into two categories that are based on the orientation of the camera's optical axis with respect to the Earth's surface (Avery and Berlin 1992). The first category is known as vertical air photos. Vertical air photos are photographic images captured from a camera angle that is either vertical (90°) or near vertical (90° ± 3°) to the Earth's surface. Oblique air photos are photographic images captured from camera angles that exceed an angular amount of 20° from being vertical to the Earth's surface. Both vertical and oblique air photos may incorporate photographic images created from color-, black and white-, color infrared-, and video-based film.

Specialized airborne platforms include spectroradiometer applications developed and employed by federal, private, and commercial enterprises. These sensors take advantage of capturing narrow bands of light reflectance within specified portions of the electromagnetic spectrum (EMS). Microwave satellites, either passive or active, record longer EMS wavelengths (radio waves) compared to those of optical sensors. Figure 3 illustrates the electromagnetic spectrum.

The resultant image scale or spatial resolution can be controlled by the aircraft's altitude. Two examples of specialized airborne platforms include the NASA Stennis Space Center's airborne terrestrial applications sensor (ATLAS) and the real-time data acquisition camera system (RDACS) which are both aircraft-deployable remote sensing systems. The ATLAS system is a multispectral instrument, flown onboard a NASA Learjet, that is capable of collecting data on 12 channels ranging from 0.45 to 12.2 μm. The NASA RDACS sensor is a 120-band multispectral airborne system made up of an array of three digital cameras mounted underneath a single engine airplane. Each camera of the RDACS sensor is fitted with a filter that determines the camera wavelength sensitivity.

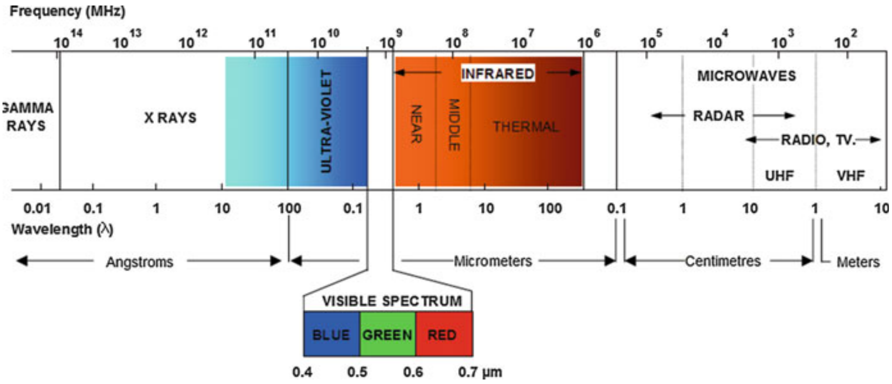


Fig. 3 The electromagnetic spectrum

Airborne as well as spaceborne platforms may also include sensors that operate within the microwave portion of the EMS. This portion of the spectrum incorporates wavelengths from approximately 1 mm to 1 m (Lillesand et al. 2008).

RADAR sensors are active microwave sensors that utilize emitted radio waves which give day or night, all weather (through cloud cover, smoke, and haze) imaging capability. This is an extremely important capability during disasters. Thus, RADAR sensors find wide applications in disaster management.

RADAR and optical sensors are complimentary to each other. In other words, they utilize the electromagnetic spectrum in a way to provide different but not competing information about the feature or the phenomenon observed. For example, in mineral deposit exploration, while hyperspectral imagery exploits the different spectral reflectance properties of the minerals for explorations, RADAR data exploit the structural geology by highlighting faults, folds, synclines and anticlines, and lineaments.

General application areas of RADAR data have been shown to be useful in this field of geology for analysis of topography and topographic changes and the mapping of surface expressions such as fault lines and movements. The variety of the spaceborne platforms includes two main categories: optical/infrared satellite platforms and microwave satellite platforms. Optical/infrared satellites operate with the visible, near-infrared and short-wave infrared portions of the electromagnetic spectrum (EMS). These sensors form images of the Earth's surface by capturing varied amounts of the Sun's radiation that is reflected from objects on the ground. Only a small portion of the electromagnetic spectrum is in the visible range as elaborated in Fig. 4. The visible portion of the electromagnetic spectrum relates to the portion of the spectrum that is visible to the human eye (390–700 *u*m; see Fig. 4).

RADAR data have also been applied to fields such as:

- Hazards and disaster management: flooding, volcanoes, earthquakes, marine oil spill, river ice buildup
- Oceanography: monitor ocean currents, winds, and changes in surface features, seasonal sea ice thickness and coastal processes (these have important

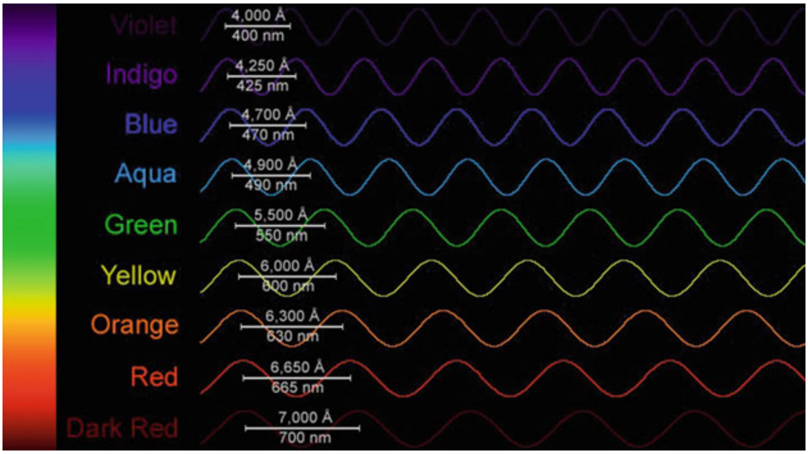


Fig. 4 The visible portion of the electromagnetic spectrum

implications for climate change research, naval navigation, and off-shore oil explorations)
- Atmospheric: measure and vertically "profile" clouds
- Hydrology: soil moisture, drainage network, land-water delineation, monitor changes in surface water discharge and storage to refine our understanding of wetland hydraulics
- Ecology: land cover classification, biomass measurements, monitoring deforestation (especially in rainforests since optical sensors are limited due to heavy cloud cover) provide quantitative information about forest structure and biomass components to estimate fuel loading for wildfire management
- Planetary: infer the geophysical processes that help shape a planet's surface.

Figure 5 illustrates some examples of passive RADAR data applications (courtesy of Shannon Ross, Canadian Center for Remote Sensing). LiDAR is also an active senor that is usually operated from an airborne platform. LiDAR is similar to RADAR in that they both are considered active remote sensing technologies. That is, the images are created from returning energy originally emitted from the senor. However, unlike RADAR, LiDAR takes advantage of ultraviolet, visible, or near-infrared light to image objects.

Light energy (photons) is emitted from the LiDAR sensor in pulses; images of the ground are created by calculating the time of the pulse return, usually in the form of an elevation dataset (Fig. 6).

Spaceborne remote sensing platforms, given the altitude advantages over airborne platforms, are often geared toward applications of mapping, monitoring, and analyzing features or changing conditions occurring on (or just below) the Earth's surface. A global inventory of the current major Earth observing satellite systems is provided in Table 1.

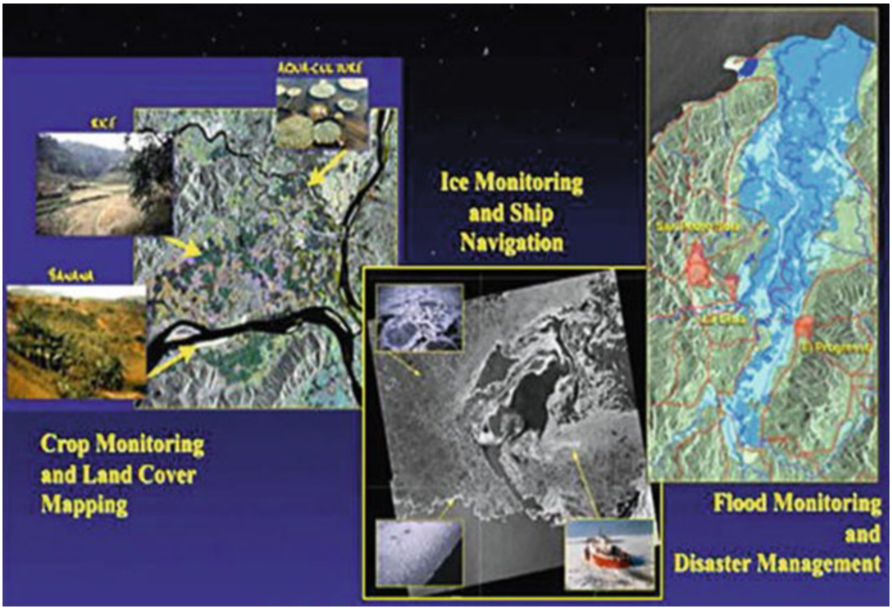


Fig. 5 Examples of some of the common applications of RADAR data from RADARSAT, Canada (Graphics courtesy of the Canadian Space Agency)

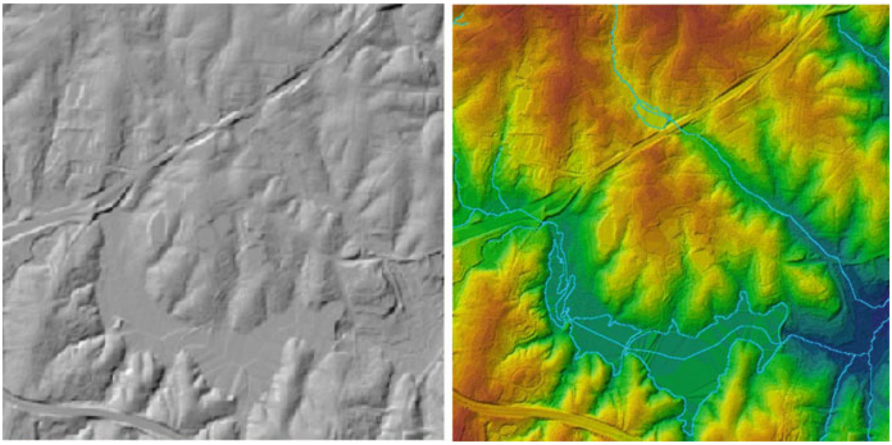


Fig. 6 The image on the left represents a 10-m digital elevation map of the upper Lake Johnson area in southwest Raleigh, NC, USA, a 93 km^2 area. This digital elevation map derived from LiDAR data acquired from the North Carolina State Floodplain Mapping Program. The image of the same area on the right represents a display that has the included breaklines (*light blue lines*) and color added to accentuate the detail

Table 1 Characteristics of selected satellite sensors (Adapted from Khorram et al. 2016, in turn adapted from Rogan and Chen 2004; updated for this publication)

Sensor (Mission)	Organization[1]	Operation period	Swath width (km)	Spatial resolution[2] (m)	Temporal resolution	Radiometric resolution	Spectral resolution (μm)	Spectral bands
MSS (Landsat 1-5)	NASA, USA	1972–1992	185	80 (MS), 240 (TIR)[3]	16–18 days	8-bit	0.5–1.1, 10.4–12.6[3]	4–5[3]
TM (Landsat 4, 5)	NASA, USA	1982–2011	185	30 (MS), 120 (TIR)	16 days	8-bit	0.45–2.35, 10.4–12.5	7
ETM+ (Landsat 7)	NASA, USA	1999–	185	15 (PAN), 30 (MS), 60 (TIR)	16 days	8-bit	0.52–0.9 (PAN), 0.45–2.35, 10.4–12.5	7+ PAN
OLI (Landsat 8)	NASA, USA	2013–	185	5 (PAN), 30 (MS)	16 days	12-bit	0.5–0.68 (PAN), 0.433–0.453 (coastal/aerosol), 0.45–2.3, 1.36–1.39 (cirrus)	8+ PAN
TIRS (Landsat 8)	NASA, USA	2013–	185	100	16 days	12-bit	10.6–11.2, 11.5–12.5	2
MODIS (EOS Terra and Aqua)	NASA, USA	1999–	2300	250 (PAN), 500 (VIR), 1000 (SWIR)	1–2 days	12-bit	0.620–2.155, 3.66–14.385	36
ASTER (EOS Terra)	NASA, USA; METI, Japan	1999–	60	15 (VNIR), 30 (SWIR), 90 (TIR)	4–16 days	8-bit (VNIR/SWIR), 12-bit (TIR)	0.52–0.86, 1.60–2.43, 8.125–11.65	14
Hyperion (EO-1)	NASA, USA	2000–	7.5	30	16 days	12-bit	0.353–2.577	220

(*continued*)

Table 1 (continued)

Sensor (Mission)	Organization[1]	Operation period	Swath width (km)	Spatial resolution[2] (m)	Temporal resolution	Radiometric resolution	Spectral resolution (μm)	Spectral bands
ALI (EO-1)	NASA, USA	2000–	37	10 (PAN), 30 (MS)	16 days	12-bit	0.48–0.69 (PAN), 0.433–2.35	9+ PAN
OMI (EOS Aura)	NIVR, Netherlands; FMI, Finland; NASA, USA	2004–	2600	13,000 × 48,000, 13,000 × 24,000[4]	1 day	12-bit	0.27–0.5	740
CALIOP (CALIPSO)	NASA, USA; CNES, France	2006–	0.1	333	16 days	22-bit[5]	0.532, 1.064	2[6]
VIIRS (Suomi NPP)	NASA, USA; NOAA, USA	2011–	3000	375 or 750 depending on application	12 h	12-bit	0.41–12.5	21+ day/night PAN
Aquarius (SAC-D)	NASA, USA; CONAE, Argentina	2011–	390	150	7 days	–	3 L-band microwave radiometer beams + radar	N/A
AVHRR (NOAA 6–19; Metop-A, -B)	NOAA, USA; EUMETSAT	1978–[7]	2700	1100	12 h	10-bit	0.58–12.5	6[8]
I-M Imager (GOES 12–15)	NESDIS, USA	1975–[9]	8	1000 (VNIR), 4000 (SWIR), 8000 (moisture), 4000 (TIR)	0.25–3 h	10-bit	0.55–12.5	5
HICO (International Space Station)	Office of Naval Research (ONR), USA; NASA, USA	2009–2014	42 × 192[10]	90	≈3 days	14-bit	0.353–1.08 (0.4–0.9)[11]	128 (87)[11]

SAR (RADARSAT-2)	CSA, Canada	2007–	20–500[12]	3–100[12]	24 days[13]		C-band radar	N/A
VEGETATION (Proba-V)	ESA	2013–	2250	100/350[14]	1–2 days[15]	12-bit	0.438–1.634	4
SAR (Sentinel-1A)	ESA	2014–	20–250[12]	5–40[12]	12 days		C-band radar	N/A
PMC (Gaofen-1)	CNSA, China	2013–	69	2 (PAN), 8 (MS)	4 days	10-bit	0.45–0.9 (PAN), 0.45–0.89	4+ PAN
WFI (Gaofen-1)	CNSA, China	2013–	830	16	4-days	10-bit	0.45–0.89	4
HROI (Gaofen-2)	CNSA, China	2014–	48	0.8 (PAN), 3.2 (MS)	?[16]	?[16]	?[16]	?[16]
MUXCAM (CBERS-4)	CBERS, China/Brazil	2014–	120	20	26 days	8-bit	0.45–0.89	4
PANMUX (CBERS-4)	CBERS, China/Brazil	2014–	60	5 (PAN), 10 (MS)	52 days	8-bit	0.51–0.73 (PAN), 0.52–0.89	3+ PAN
IRSCAM (CBERS-4)	CBERS, China/Brazil	2014–	120	40 (NIR/SWIR), 80 (TIR)	26 days	8-bit	0.77–12.5	4
WFICAM (CBERS-4)	CBERS, China/Brazil	2014–	866	64	5 days	10-bit	0.45–0.89	4
LISS-IV (RESOURCESAT-2)	ISRO, India	2011–	70	5.8	24 days	10-bit	0.52–0.86	3
LISS-III (RESOURCESAT-2)	ISRO, India	2011–	141	23.5	24 days	10-bit	0.52–1.70	4

(*continued*)

Table 1 (continued)

Sensor (Mission)	Organization[1]	Operation period	Swath width (km)	Spatial resolution[2] (m)	Temporal resolution	Radiometric resolution	Spectral resolution (μm)	Spectral bands
AWiFS (RESOURCESAT-2)	ISRO, India	2011–	740	56	24 days	12-bit	0.52–1.70	4
TANSO-FTS (GOSAT/Ibuki)	JAXA, Japan	2009–	160[17]	10.5	3 days		0.758–14.3	4
Geoton-L1 (Resurs-P No. 1, No. 2)	Roscosmos, Russia	2013– 2014–	32 (PAN), 38 (MS)	1 (PAN), 3-4 (MS)	3 days	10-bit	0.58–0.8 (PAN), 0.45–0.80	5+ PAN
KShMSA-VR/SR (Resurs-P No. 1, No. 2)	Roscosmos, Russia	2013– 2014–	97 (VR mode), 440 (SR mode)	VR: 12 (PAN), 24 (MS); SR: 60 (PAN), 120 (MS)	3 days	12-bit	0.43–0.7 (PAN), 0.43–0.9	5+ PAN
GSA (Resurs-P No. 1, No. 2)	Roscosmos, Russia	2013– 2014–	25	30	3 days	14-bit	0.4–1.1 nm band width	130
COSI (KOMPSAT-5)	KARI, South Korea	2013–	5–100[12]	1–20[12]	28 days		X-band radar	N/A
AEISS-A (KOMPSAT-3A)	KARI, South Korea	2015–	12	0.55 (PAN), 2.2 (MS)	28 days	14-bit	0.45–0.9 (PAN), 0.45–0.9	4+ PAN
IIS (KOMSAT-3A)	KARI, South Korea	2015–	12	5.5		14-bit	3.3–5.2	1
HRVIR (SPOT 4)	Airbus Group, France[18]	1998–2013	60	10 (PAN), 20 (MS)	26 days[19]	8-bit	0.61–0.68 (PAN), 0.50–1.75	4+ PAN
VEGETATION (SPOT 4, 5)	Airbus Group, France[18]	1998–2013; 2002–2014	2250	1150	26 days[19]	10-bit	0.43–1.75	4

HRG (SPOT 5)	Airbus Group, France[18]	2002–	60	2.5–5 (PAN), 10 (VNIR), 20 (SWIR)	26 days[19]	8-bit	0.48–0.71 (PAN), 0.50–1.75	4+ PAN
SPOT 6, 7	Airbus Group, France / Azercosmos, Azerbaijan[20]	2012– 2014–	60	1.5 (PAN), 6 (MS)	26 days[21]	12-bit	0.45–0.745 (PAN), 0.45–0.89	4+ PAN
Pléiades 1A, 1B	CNES/Airbus Group, France	2011– 2012–	20	0.5 (PAN), 2 (MS)	26 days	12-bit	0.45–0.83 (PAN), 0.43–0.94	4+ PAN
IKONOS	DigitalGlobe, US[22]	1999	11.3	1 (PAN), 4 (MS)	3–5 days	11-bit	0.526–0.929 (PAN), 0.445–0.853	4+ PAN
Quickbird	DigitalGlobe, USA	2001–	18	0.65 (PAN), 2.62 (MS)	2.5–5.6 days	11-bit	0.405–1.053 (PAN), 0.43–0.918	4+ PAN
GeoEye-1	DigitalGlobe, US[22]	2008–	15.2	0.41 (PAN), 1.65 (MS)	<3 days	11-bit	0.45–0.80 (PAN), 0.45–0.92	4+ PAN
WorldView-2	DigitalGlobe, USA	2009–	16.4	0.46 (PAN), 1.85 (MS)	1.1–3.7 days	11-bit	0.45–0.80 (PAN), 0.45–1.04	8+ PAN
WorldView-3	DigitalGlobe, USA	2014–	13.1	0.31 (PAN), 1.24 (MS), 3.7 (SWIR), 30 (CAVIS)[23]	<1 day	11-bit (PAN and MS), 14-bit (SWIR)	0.45–0.80 (PAN), 0.4–1.04, 1.195–2.365,	28+ PAN

(*continued*)

Table 1 (continued)

Sensor (Mission)	Organization[1]	Operation period	Swath width (km)	Spatial resolution[2] (m)	Temporal resolution	Radiometric resolution	Spectral resolution (μm)	Spectral bands
							0.405–2.245 (CAVIS)	
RapidEye	BlackBridge, Germany	2009–	77	5	1 day	12-bit	0.44–0.85	5
HiRAIS (Deimos-2)	Elecnor Deimos, Spain	2014–	12	1 (PAN), 4 (MS)	4 days[24]	10-bit	0.45–0.9 (PAN), 0.42–0.89	4+ PAN

[1] Organization acronyms: *NASA* National Aeronautics and Space Administration, *METI* Ministry of Economy, Trade, and Industry, *NIVR* The Netherlands Agency for Aerospace Programmes, *FMI* Finnish Meteorological Institute, *CNES* Centre National d'Études Spatiales, *NOAA* National Oceanic and Atmospheric Administration, *NESDIS* National Environmental Satellite, Data, and Information Service, *CSA* Canadian Space Agency, *ESA* European Space Agency, *CNSA* Chinese National Space Administration, *CBERS* China-Brazil Earth Resources Satellite Program, *CONAE* Comisión Nacional de Actividades Espaciales, *EUMETSAT* European Organisation for the Exploitation of Meteorological Satellites, *JAXA* Japanese Aerospace Exploration Agency, *ISRO* Indian Space Research Organization, *KARI* Korean Aerospace Research Institute

[2] Acronyms used in describing sensor channels/configurations: *PAN* panchromatic, *MS* multispectral, *VNIR* visible and near-infrared, *SWIR* short-wave infrared, *TIR* thermal infrared

[3] The MSS sensor on Landsat 3 had a fifth spectral band for thermal infrared. The MSS sensors on other Landsat missions had four-band configurations

[4] The OMI sensor aboard EOS Aura has one channel recording at 0.27–0.31 μm (in the ultraviolet spectral range) and another recording at 0.306–0.5 μm (ultraviolet and visible spectral range); the pixel size of the former channel is larger (13 × 48 km vs. 13 × 24 km)

[5] Each receiver channel on the CALIOP sensor, a spaceborne LiDAR system, has dual 14-bit digitizers that jointly provide a 22-bit dynamic range

[6] The CALIOP sensor produces simultaneous laser pulses at two wavelengths: 0.532 and 1.064 μm

[7] The NOAA satellite program began in 1978. Currently, NOAA-19 is designated as the program's operational satellite. The AVHRR sensor is also carried as a "legacy" instrument on the Metop-A and Metop-B satellites

[8] While the AVHRR sensor has six spectral channels, only five are transmitted to the ground at any time. The bands designated 3A and 3B are transmitted only during daytime and nighttime, respectively

[9] Although the GOES satellite program has been active since 1975, only GOES-13 through GOES-15 are currently operational. GOES-14 was launched in 2009 but remains in on-orbit storage mode

[10] Operated as a demonstration instrument, the HICO sensor captured only a single 42 × 92 km scene per orbit. Over its 5-year period of operation, HICO collected >10,000 scenes worldwide

[11] Collected HICO data have 128 bands spanning the 0.353–1.08 μm range of wavelengths, but the data from outside the 0.4–0.9 μm range are less accurate and therefore, only 87 bands are included in the output files made available to users

[12] The radar systems in this table (e.g., SAR/RADARSAT-2, SAR/Sentinel-1A, and COSI/KOMPSAT-5) operate in a variety of scan modes with different swath widths and spatial resolutions

[13] RADARSAT-2 has a repeat cycle of 24 days, but its left- and right-looking modes may provide more rapid revisits

[14] The spatial resolution of the VEGETATION sensor on Proba-V is 100 m at nadir but 350 m over the full field of view (i.e., at the scene edges)

[15] The VEGETATION sensor provides daily coverage for the following latitudes: 35–75° N and 35–56° S. It offers full global coverage every 2 days

[16] Few details about Gaofen-2 are available, but it is likely similar to Gaofen-1 in terms of spectral resolution and revisit time

[17] The FTS sensor on the GOSAT satellite records 10.5 × 10.5 km images that are spaced 150 km apart in a grid

[18] Spot Image, a company created by the French Space Agency (CNES), developed and operated the SPOT satellites. Spot Image became a subsidiary of Astrium, an EADS (European Aeronautic Defense and Space) company in 2008. EADS was reorganized as the Airbus Group in 2014

[19] The SPOT satellites have the capacity to record data off-nadir (i.e., to record data in areas that are not directly below the satellite). This may reduce revisit time to 2–3 days (1 day for the wide-swath VEGETATION sensor), but the images will have different observation geometry, which may affect image processing

[20] Unlike the first five SPOT satellites, SPOT 6 and 7 were developed by the Airbus Group without financial support from the French government. In 2014, SPOT 7 was sold to Azerbaijan's space agency Azercosmos, who renamed it Azersky. Per agreement, Azercosmos has commercial rights to imagery from both satellites for the southern Asia region, while Airbus Group retains the rights elsewhere

[21] The SPOT 6 and 7 satellites are a constellation, operating on the same orbit but phased 180° apart. With the satellites' capacity to record data off-nadir, the constellation can provide 1-day revisit time

[22] GeoEye, the company that originally operated IKONOS and GeoEye-1, merged with DigitalGlobe in 2013

[23] CAVIS = Clouds, Aerosols, Vapors, Ice, and Snow

[24] The 4-day revisit time of the HiRAIS sensor on Deimos-2 is enabled by a ± 45° off-nadir viewing capability

Airborne platforms provide a diversity of image formats (i.e., air photos, video, spectral images, radar images). In fact, air photos represent the longest historic record of monitoring and capturing images of the Earth's surface from distance. However, the majority of spectral remote sensing images acquired from spaceborne platforms are not photographs but rather digital images that are recorded within very specific ranges (or bands) of the electromagnetic spectrum. The EMS is made up of wavelengths of light or natural energy that extends from very short wavelengths (gamma rays and X-rays), through the section of visible light that humans use to see (blue, green, red), to the much longer wavelengths of infrared and radio waves (see Fig. 3).

Spectral images that are captured by the sensor are made up of picture elements known as *pixels*. Within an image, pixels that make up images are oriented in horizontal rows and vertical columns. Each pixel contains the reflectance data of specific ranges (or bands) of the electromagnetic spectrum that the sensor was designed to capture. Each pixel corresponds to the particular location on the Earth's surface from where the image was collected. Additionally, the ground spatial resolution of the sensor corresponds to the individual pixel size that each image is comprised of. For example, the spatial resolution of the multispectral scanner (ETM+) aboard the Landsat-7 satellite is 30 m. This means that of the thousands of pixels that make up a 185 km (across track) by 180 km (along track) Landsat-7 scene, each pixel will have a spatial resolution corresponding to a 30 m by 30 m area on the ground.

Hyperspectral sensors also referred as "imaging spectroscopy" are passive sensors that take images of the spectral radiance using a wider range of the electromagnetic spectrum than the multispectral sensors.

Hyperspectral images may contain hundreds of bands as illustrated in Fig. 7. For example, while the MODIS (or moderate resolution imaging spectroradiometer) sensor, aboard the Terra (EOS AM) and Aqua (EOS PM) satellites, is acquiring data in 36 spectral bands, [1] the AVIRIS (airborne visible infrared imaging spectrometer) instrument collects data from 224 continuous spectral channels (bands) ranging from ultraviolet to near-infrared.[2]

Hyperspectral imagery is used for a broad array of applications ranging from mineralogy to air quality. Some of these applications include uses in the atmospheric sciences, such as the examination of cloud properties, air quality, and the variability in aerosol composition. Terrestrial applications include mineralogy, geology, ecology, agriculture, as well as the monitoring snow/ice extent, and climate change. Aquatic system applications include monitoring coral reefs, mapping chlorophyll, phytoplankton, dissolved organic materials, and suspended sediment concentrations. Hazard and disaster management include biomass burning, flooding, hurricanes, and oil spills as illustrated in Fig. 9.

---

[1] Hyperspectral sensing and imaging, http://modis.gsfc.nasa.gov/about

[2] Airborne visible infrared imaging spectrometer, http://aviris.jpl.nasa.gov/

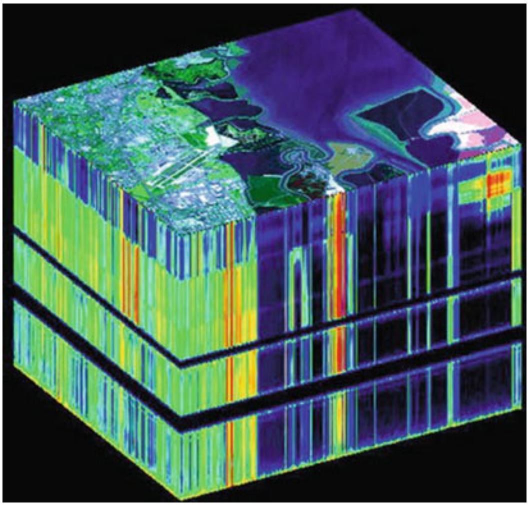


Fig. 7 An example of AVIRIS data that was acquired on August 20, 1992, over Moffett Field, California, at the southern end of the San Francisco Bay (Recent AVIRIS Imaging http://aviris.jpl.nasa.gov/html/aviris.cube.html) (Image courtesy of NASA JPL and NASA Ames)

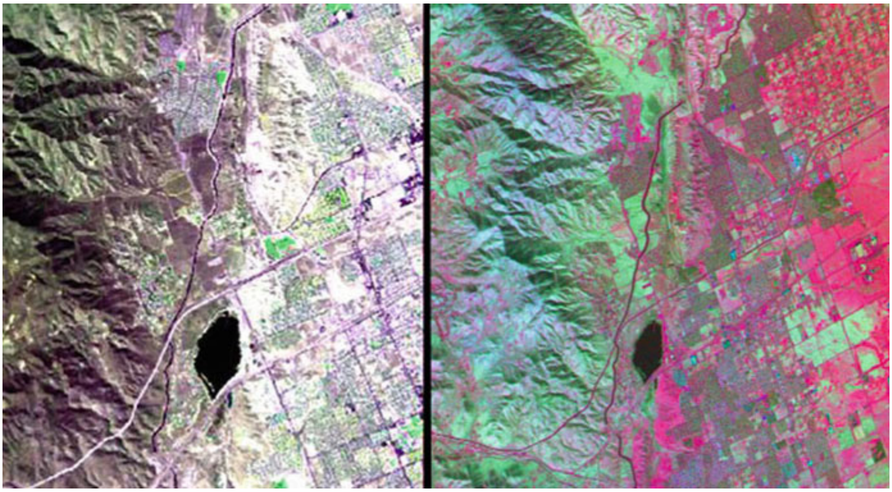


Fig. 8 NASA's Hyperspectral Infrared Imager (HyspIRI) airborne overflight of California's San Andreas Fault, captured March 29, 2013. The image on the *left* represents a color composite (*red*, *green*, *blue*) airborne visible/infrared imaging spectrometer (AVIRIS) image of the fault. The image on the *right* demonstrates the application of temperature information that was collected simultaneously by the NASA Jet Propulsion Laboratory's (JPL) MASTER instrument. The image on the *right*, *red* areas represent urban areas that are composed of minerals with high silica, while *darker* (cooler) areas represent water and more heavily vegetated areas (Courtesy of NASA Jet Propulsion Laboratory)

Hyperspectral sensors are able to allow for small increments of wavelengths between each band allow detection and mapping of features, objects, and phenomena, which often is not possible by conventional airborne or satellite operational scanners. Figure 8 depicts an example of hyperspectral data over a terrestrial system.

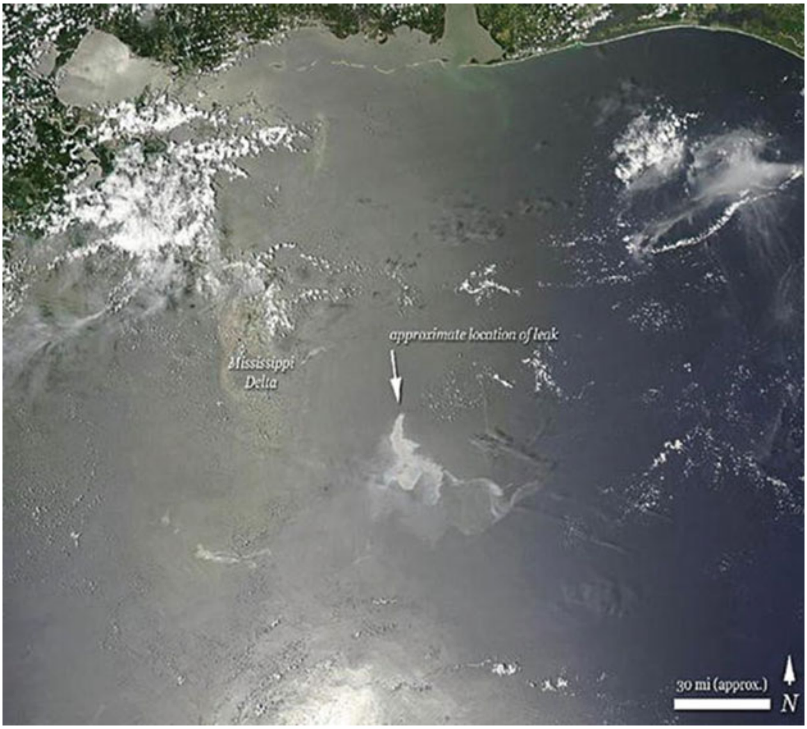


Fig. 9 Illustrating the 2010 BP oil spill in the Gulf of Mexico from MODIS data (Image courtesy of NASA)

Figure 9 illustrates an example of MODIS application for detecting the oil slick on July 14, 2010, from BP's Deep Horizon well. The largest oil slick is located in the center of the image, but a few isolated ribbons of oil are visible to the east. The lighter-colored waters around the river delta are full of sediment and do not reflect oil slicks.

Image Processing Tools

Image Display

Images are typically displayed in true color composites (TCC), false color composites (FCC), or classified form. Raw data displays are most frequently provided as true color composite and the false color composite. True color composite displays the blue band in blue color, the green band in green color, and the red band in red color, while false color composite displays any three bands in blue, green, and red colors. The standard FCC provides an infrared band displayed in red color, the red band displayed in green color, and the green band displayed in blue color.

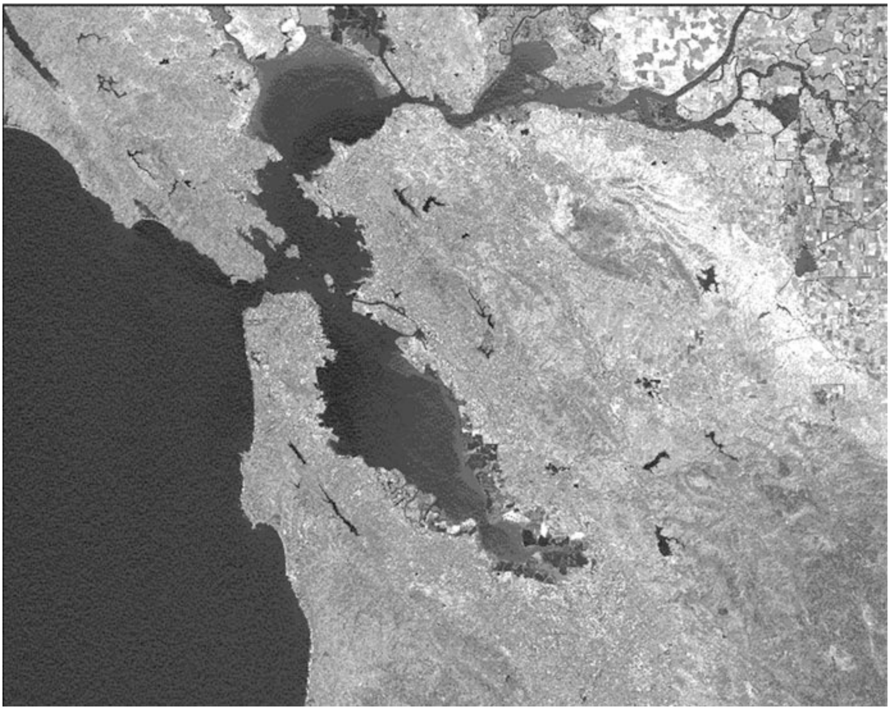


Fig. 10 Panchromatic display of San Francisco Bay Area from Landsat ETM data

Raw data displays are most frequently provided as True Color Composite (TCC) and the False Color Composite (FCC). True Color Composite displays the Blue band in blue color, the Green band in green color, and the Red band in red color while False Color Composite displays any three bands in blue, green, and red colors. The standard FCC provides an Infrared band displayed in red color, the Red band displayed in green color, and the Green band displayed in blue color.

Examples of panchromatic, true color composite, and standard false color composite displays of the San Francisco Bay Area acquired from Landsat 7 ETM (Enhanced Thematic Mapper) are shown in the following three figures (Figs. 10, 11, and 12 respectively).

Image Preprocessing

Remotely sensed data often contain various types of distortions due to less than optimal atmospheric conditions, rotation of the Earth, satellite or aircraft motion, curvature of the Earth, and the exact location of a given point within an image. The remotely sensed data, particularly in higher spatial resolutions, involve enormous amounts of data that often needs to be reduced before image processing. Various

Fig. 11 True color composite display of San Francisco Bay Area from Landsat ETM data

techniques designed to remove systematic and nonsystematic effects and distortions as well as most commonly accepted data reduction methodologies are discussed in this section. Image processing and postprocessing are described in the following chapter.

Atmospheric Correction

Clouds, suspended particles, or other materials in the atmosphere at the time of data acquisition change the digital numbers on an image. While most terrestrial applications of remotely sensed data have looked for virtually cloud-free days and/or multiple scenes, and thus have essentially avoided or ignored this issue, haze, clouds, and other atmospheric effects over coastal and near-shore ocean areas are pervasive. Four primary methods have been developed to remove or minimize the atmospheric effects on an image.

Absolute radiometric correction of atmospheric attenuation takes into account the solar zenith angle at the time of satellite overpass, the total optical thickness caused by molecular scattering, the atmospheric transmittance for a given angle of incidence, the spectral irradiance at the top of the atmosphere, and the Rayleigh and Mie scattering laws (Turner and Spencer 1972; see also Forster 1984).

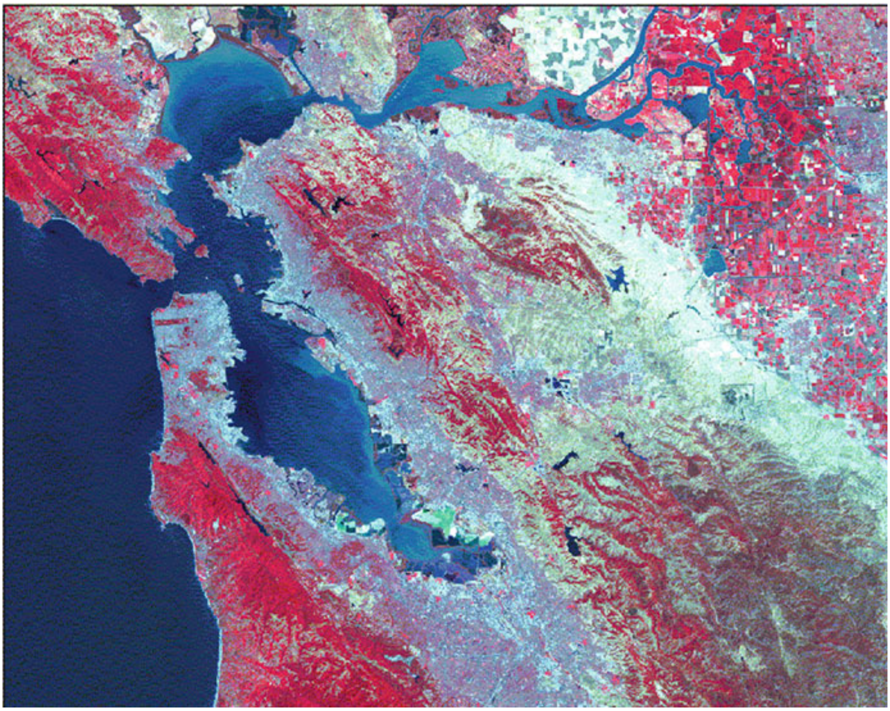


Fig. 12 Standard false color composite display of San Francisco Bay Area from Landsat ETM data

Relative radiometric correction of atmospheric attenuation normalizes the intensities among different bands within a scene to remove detector-related problems and then corrects the intensities through a comparison with a standard reference surface on the same date and same scene.

Single-image normalization uses a histogram adjustment for any shift in the histogram that may have been caused by atmospheric effects. This method is based on the fact that infrared data are largely free of atmospheric scattering effects as compared to the visible region. Thus, the histogram shifts due to haze can be used to adjust for the atmospheric effects. This method involves a subtractive bias established for each band and is very simple to use.

Multiple image normalization uses regression analysis for a number of dates. This method is primarily used to make sure the spectral values from one date are comparable to other dates, which implicitly takes the atmospheric corrections into account. This method is primarily used for change detection purposes and is also fairly simple to use.

Geometric Corrections

Sources of errors causing geometric distortions in remotely sensed data are divided into systematic and nonsystematic. Systematic distortions are due to image motion caused by forward movement of the spacecraft, variations in the mirror scanning rate, panoramic distortions, variations in platform velocity, and distortions due to the curvature of the Earth. Nonsystematic distortions are due to variations in satellite altitude and attitude. Much of the systematic errors are removed in commercially available remotely sensed data. The most common techniques for removing the remaining systematic and nonsystematic distortions are image-to-map rectification and image-to-image registration through the selection of a large number of ground control points.

Radiometric Corrections

The type of radiometric distortions varies among the different sensors. Typically, the Sun elevation corrections and Earth–Sun distance corrections are applied to satellite data for removing the effects of the seasonal position of the Sun relative to the Earth and to normalize for the seasonal variations in the distance between the Earth and the Sun. Noise removal algorithms can be applied to remove unwanted disturbance due to sensing, signal quantization, and recording. Several destriping algorithms are available to remove the striping and banding effects in satellite data. The line drops can be corrected by replacing the spectral values in the missing band with the average of the line(s) above and below them. The nonsystematic variations in gray levels from pixel to pixel (i.e., bit errors) can be corrected by replacing these values with neighboring values that exceed the threshold values established by analysts. Figures 13 and 14 depict the removal of radiometric differences before and after normalization.

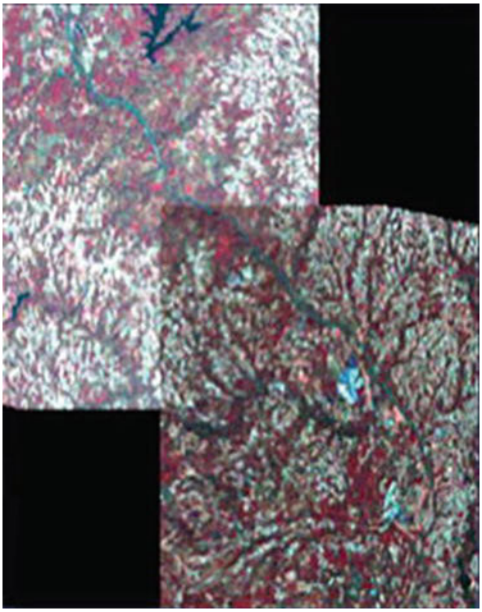


Fig. 13 Before normalization of radiometric differences (Graphic courtesy of Siamak Khorram)

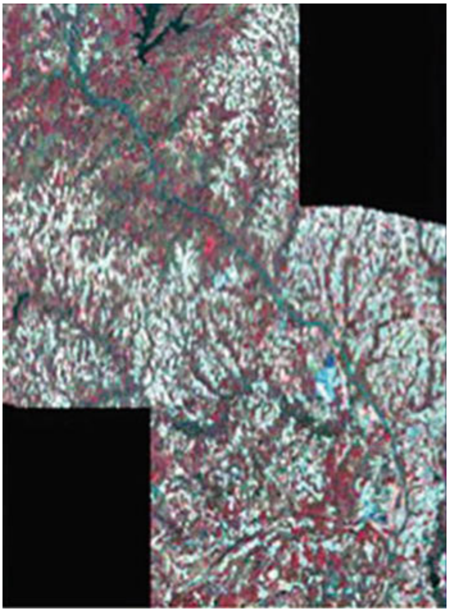


Fig. 14 After normalization of the radiometric differences (Graphic courtesy of Siamak Khorram)

Occasionally, due to a malfunction in the scanners or other issues, images contain a large number of line drops that need to be corrected. Often the digital values for these line drops are zero. A common way of correcting this issue is to replace the zero values with the average of the values for the scan lines above and below the bad scan line. A linear averaging is often sufficient for drop lines that are not more than a few pixels up or down. When the number of drop lines are more than a few pixels (usually 3–5 depending on the local scene dynamic), the linear averaging does not seem to be appropriate. In this case, either a more complicated modeling approach is considered or these large drop lines should be left uncorrected. Figure 15a depicts the uncorrected image and Fig. 15b depicts the image after radiometric correction with one large line which was left uncorrected based on airborne ocean color scanner data (Khorram 1982).

Spectral distortions may also be evident in the raw imagery that may limit the ability to develop accurate classifications. For example, the left side of Fig. 15 shows imagery of two individual multispectral scanner (MSS) bands that were acquired from Landsat-4. Linear distortion features are noticeable across each image, while for general land use/land cover classification purposes these distortions may not have a large impact on developing overall general classification categories for a single image. However, to ensure the best possible use of training sites between multiple images, including images that may not have these types of distortions, it is generally good practice to remove or reduce these types of distortions.

In the example depicted in Fig. 16, the linear distortions occur in a uniformed orientation pattern, most likely due to the deteriorating age of the sensor. Uniformed distortions are generally less of a challenge to remove than distortions that occur

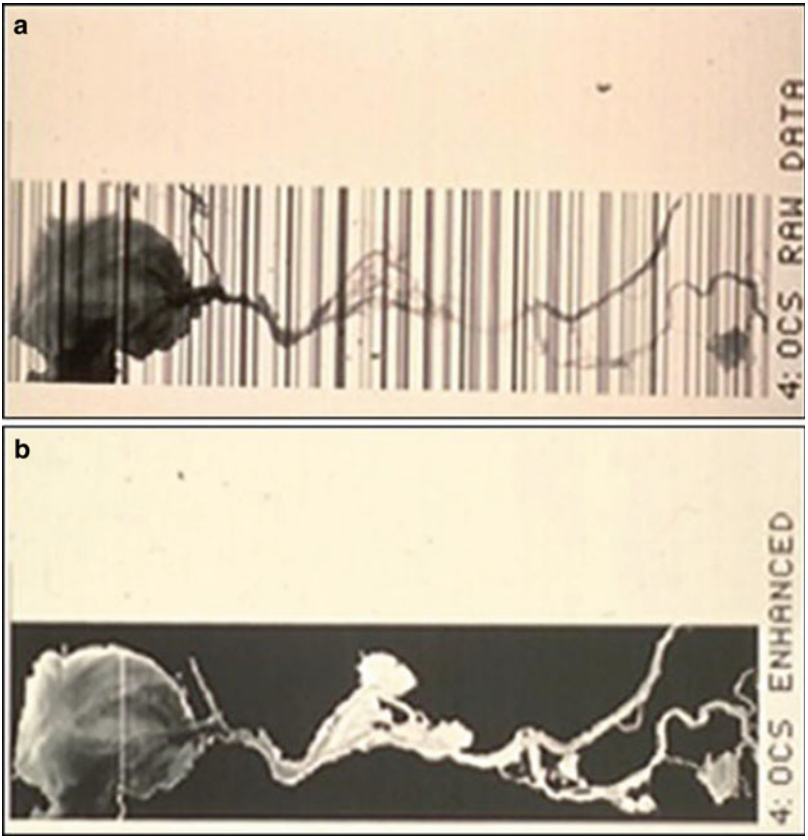


Fig. 15 (**a**) Remotely sensed image data containing a large number of *dropped lines*. (**b**) Remotely sensed image data where *dropped lines* (zero values) have been replaced with the average of the values of the *scan lines* above and below the *bad scan lines* (Graphic Courtesy of Siamak Khorram)

randomly across an image. By dividing band 1 by band 2 to create a new ratioed image, the combined band data were able to reduce the original distortions apparent in the individual scenes. The newly created "ratioed image" may then be substituted in the band combination used for the classification procedure (Nelson et al. 2002). This type of radiometric correction, as well as any corrections that improve the interpretability of an image, is also known as a type of image enhancement procedure.

Registration and Coordinate Systems

Image registration or image alignment is the process of transforming multiple sets of images, or other spatially referenced data that will be used together with the image

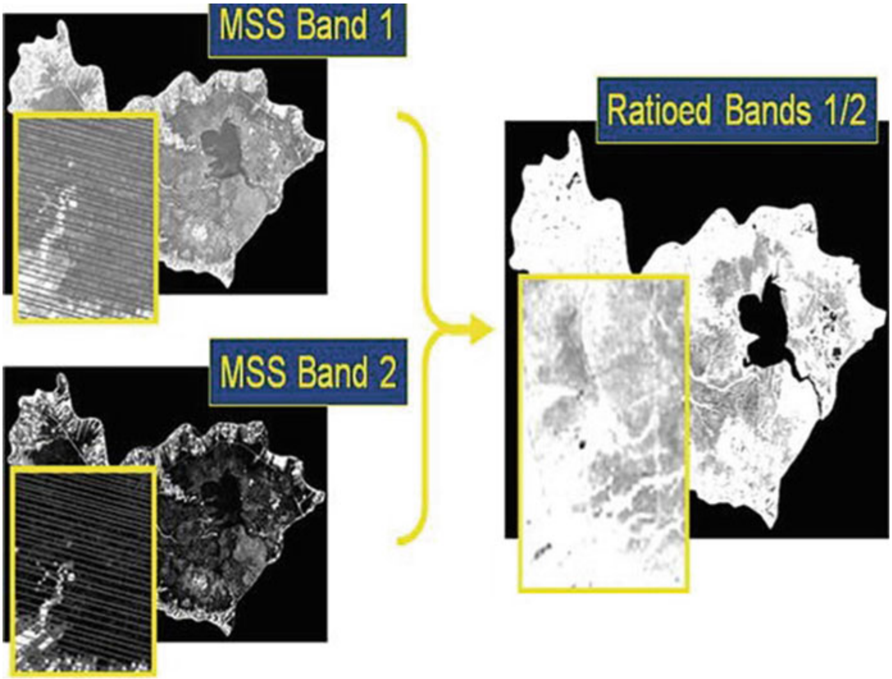


Fig. 16 Images to the *left* represent two individual raw Landsat MSS band scenes that contain linear distortions across the imagery, most likely due to the age of the sensor. The *right* image represents a new scene created from the "ratioed bands" to reduce visible distortions (Images Courtesy of NOAA)

data, into a single coordinate system of interest. A coordinate system (i.e., angular or geographical) is a system which uses a series of numbers, or coordinates, to uniquely determine locations on the Earth's surface. Coordinate systems also provide a reference system used to measure horizontal and vertical distances on a planimetric map (i.e., flat map). The coordinate system is usually defined by a map projection, a spheroid of reference, a datum, one or more standard parallels, a central meridian, and possible shifts in the x- and y-directions to locate X, Y positions. The map projections are used to transform a position on a curved Earth's surface (identified by latitude and longitude) into Cartesian coordinates (X, Y).

Image data may be acquired at different times or from different data sources or sensors. All of these data may be registered in different coordinate systems that prevent corresponding locations on each image to overlay on top of each other. For this data useful in subsequent analyses involving multiple images or spatial datasets, the multiple images must be coregistered in a process that brings all data into a single coordinate system. For example, positional image registration errors (misregistration errors) have been shown to significantly limit change detection accuracies (Dai and Khorram 1999).

The majority of current change detection techniques depend on the accuracy of geometric registration of the two images being used as the resultant change analysis is generally performed on a pixel-by-pixel basis (Townshend et al. 1992). If accurate registration between images is not achieved, false differences will be detected due to the misalignment of features on the ground that would be evaluated as real changes at the same location from one time and another. To ensure maximum spatial fidelity between various types of image data, it is often recommended to coregister all data, as well as georeference the data using ground control points (GCPs). A collection of a suitable number of GCPs (generally 50 points per class) will allow for an adequate feature matching and boundary-matching approach that will allow for a statistically valid computation of the root mean square error (RMSE). The RMSE is the absolute fit of the model to the data or the difference between values predicted by the transformation model and the actual values observed.

Data Reduction and Fusion Techniques

Data reduction and fusion are important components of image processing. Data reduction aims to reduce redundancy in digital images thus saving crucial storage space and processing time. Usually dubbed as image compression, data reduction techniques find applications in multimedia, communications, remote sensing, and digital image databases. One of the most widely known image compression method is the JPEG 2000 standard which is based on wavelet analysis techniques. Image compression can be lossless or loss-y meaning that some information might be lost depending upon the technique used.

Data fusion techniques, on the other hand, try to assimilate information from multiple sources into a more useful form depending upon the objectives and the need of the user. If the need is, for example, to have high-resolution color images, colored (multispectral) but low spatial resolution satellite imagery can be fused with high-spatial-resolution panchromatic image to have sharp multispectral images. This fusion process is known as pan sharpening.

The most widely used tools for data reduction and fusion are described in this section. Figure 17 illustrates their application.

Principal Components Analysis

Principal component analysis (PCA) is a multivariate statistical technique that originated from Pearson's early work (Pearson 1901). PCA is a mathematical way of determining the linear transformation of a sample of points in N-dimensional space that exhibits the properties of the sample most clearly along the coordinate axes (Gauch 1982).

Along the new axes, the sample variances are extremes (maxima and minima) that are uncorrelated. An analysis in terms of principal components can show (linear) interdependence in data (Legendre and Legendre 1998).

PCA is calculated either using variance-covariances or correlations. (If a correlation matrix is used, it is also known as "factor analysis" or "standardized principal components.") Given an image with n-number of bands, n-number of PCs can be calculated. The first PC explains most of the variability of the original data set.

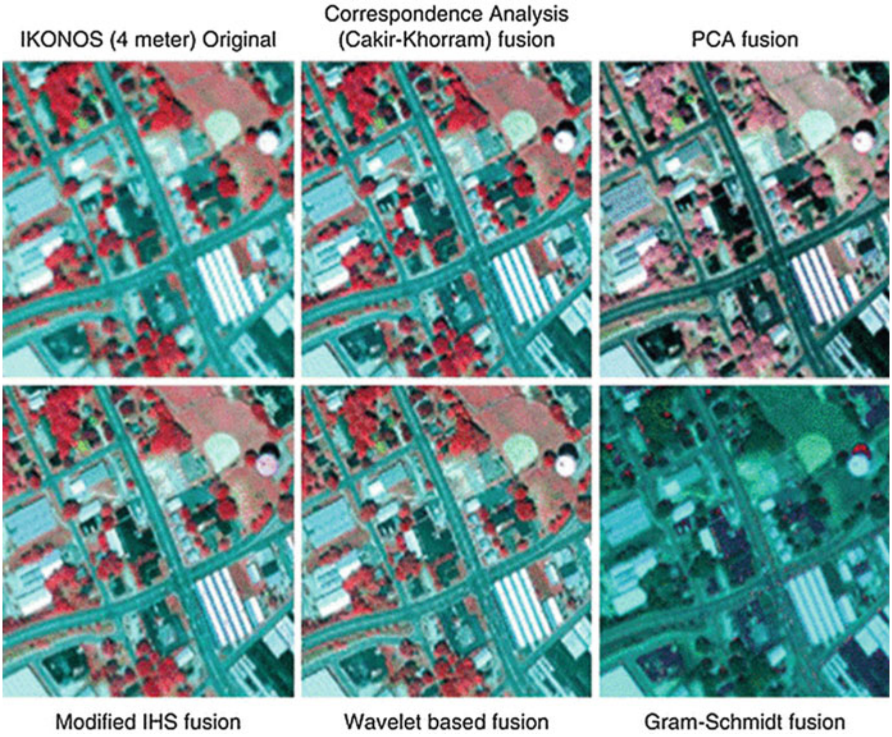


Fig. 17 An illustration of various data fusion techniques to a false color composite (FCC) of IKONOS satellite data (Graphic courtesy of IKONOS)

The second, third, fourth, and so on, components contain decreasing amounts of variance found in the data set. Thus, eliminating the components that have less information, associated with smallest eigenvalues, reduces the abundance of data to be analyzed. This is extremely helpful when classifying correlated sets of multispectral data.

The success of PCA depends on the data analyzed. By definition, PCA can be applied best if the data are Gaussian and grouped into a single cluster. Also, the success of PCA depends on the extent of multicolinearity within the data to be analyzed.

In remote sensing, PCA is mostly applied to correlated data sets to improve interpretability of the images (Short 2003). PCA is useful for image encoding, image data compressing, image enhancement, digital change detection, multitemporal dimensionality, and image fusion (Pohl and Genderen 1998).

Correspondence Analysis

A more recent multivariate method, correspondence analysis (CA) was developed independently by several authors. An algebraic derivation of CA is often accredited to Hirschfield (1935) who developed a formulation of the correlation between the rows and columns of a two-way contingency table (Beh 2004). For the development, history, and more information of the technique, readers can refer to the various citations in the end notes (Greenacre 1984; Benzécri 1992; Beh 2004). The term "correspondence analysis" is a translation of the French "analyse factorielle des correspondances."

Correspondence analysis can be applied to data tables other than contingency tables as long as the elements of a table to be analyzed are dimensionally homogenous (i.e., same physical units, so that they can be added) and nonnegative (so that they can be transformed into probabilities or proportions). The difference between PCA and CA is that CA preserves the chi-square (χ^2) distance when computing the association between bands or variables (Carr and Matanawi 1999). In PCA, the distance among objects, in both the multidimensional space of original descriptors and the reduced space, is calculated using Euclidean distances. Most of the time, the last CA component is omitted from the analyzing procedure because the last eigenvalue is insignificant or small.

Although it is very well known to ecologists correspondence analysis is rarely explored in the remote sensing community. Carr and Matanawi (1999) introduced CA for principal component transformations of multispectral and hyperspectral imagery. Later, it was successfully applied to detection of change and data fusion (Cakir et al. 2006).

In image processing, CA, similar to PCA, can be used for image encoding, dimensionality reduction and image compression, data fusion, multitemporal analysis and change detection, image classification, etc. While PCA normally highlights similarities in image contents, CA highlights differences.

Figure 18 illustrates a comparison between the two widely used data fusion and reduction techniques, the principal components analysis, and the correspondence analysis. In this figure, one must compare the color fidelity of the two techniques to the original image, which is an indication of preserving the spectral integrity of the image after fusion. Here, the Cakir-Khorram correspondence analysis technique is superior to the principal components analysis.

In Figure 19, an image of an area adjacent to the Pyramids in Egypt is shown illustrating before and after applying the Cakir-Khorram correspondence analysis data fusion technique.

Intensity, Hue, and Saturation Analysis

Intensity, hue, and saturation (IHS) refer to the parameters of human color perception. Intensity refers to the total brightness of a color. Hue refers to the dominant or average wavelength of light contributing to a color. Saturation specifies the purity of a color relative to gray. Vivid colors are highly saturated, while pale pastel colors have low saturation. IHS transformation has been primarily used to fuse panchromatic data with three-band multispectral data in order to enhance spatial resolution of the multispectral data. A numerical procedure, which was developed to convert a three-band RGB (red-green-blue) display into its fundamental physiological (IHS) elements of human color perception, is performed to transform the multispectral input data into IHS space. Several algorithms have been developed for the IHS

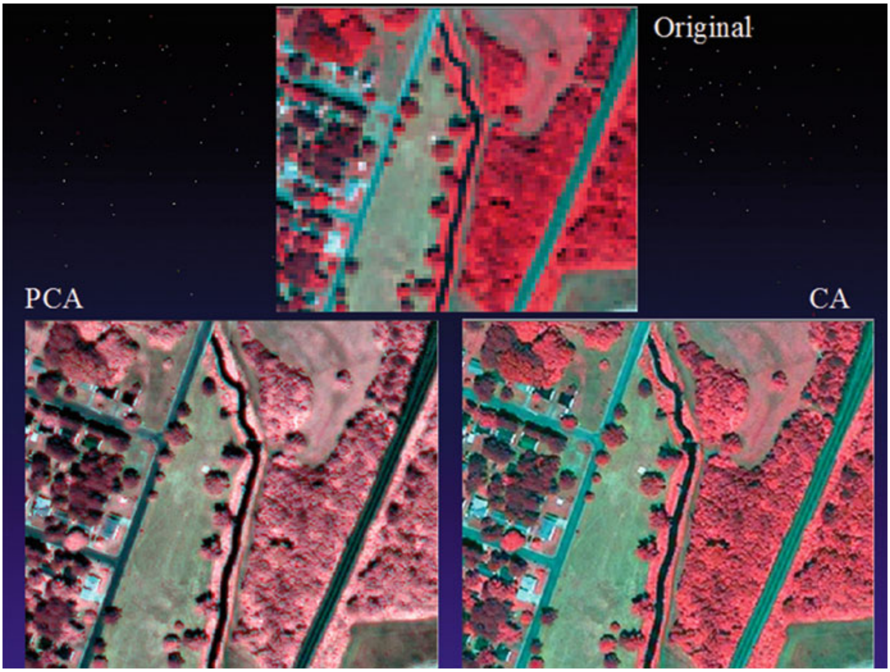


Fig. 18 The application of the PCA and CA techniques as compared to the original image. The true color composite on the *left* and the standard false color composite on the *right*

Fig. 19 Pan sharpening through fusion of multispectral and panchromatic Quickbird images. A true color composite (TCC) of the original multispectral image is shown on the *left*, while a TCC of the pan-sharpened images is shown on the *right*

transformation. Generally they are similar to each other with small differences in calculating the intensity component.

In data fusion (pan sharpening) of multispectral images, utilization of IHS transformation involves the substitution of one of the components with another data. Usually, it is the intensity component that is replaced by the panchromatic image. It is assumed that the intensity component is highly correlated to the panchromatic data. Panchromatic data can be contrast stretched to match the intensity component before replacement in order to reduce the spectral distortion on the final product. An inverse transformation is then performed and the data are transformed back to the original RGB space. The IHS method is sufficient for SPOT imagery but not for sensors like the Landsat TM sensor that has six or more reflection bands. The correlation analysis of original multispectral image data and their counterparts in IHS composites indicates the need to consider carefully the potential influence alternative implementations of IHS procedures might have on the spectral characteristics of the resulting multiresolution products.

Wavelet Method

Wavelets are an extension of Fourier analysis. Thus, to understand the wavelets and wavelet transform, it is important to look at Fourier analysis and its limitations. Fourier transform decomposes or separates a waveform or function into sinusoids of different frequency which sum to the original waveform. It recognizes or distinguishes the different frequency sinusoids and their respective amplitudes. It transforms a function f that depends on time (or on space) into a new function, $(\hat{f})$, or "f hat," which depends on frequency. This new function is called the Fourier transform of the original function or, when the original function is periodic, it is called a Fourier series.

A function and its Fourier transform are two faces of the same information. The function displays time (or space) information and hides information about frequencies. Fourier transform displays information about frequencies, but information about time or space is hidden in the phases. Wavelets can provide efficient localization in both time and frequency or scale when the signal is represented as a function of time, that is, the major difference between Fourier transform and wavelet transform.

Grossmann and Morlet among others have showed that when wavelets are used to represent a signal, the "energy" of the signal is unchanged. This means that one can transform a signal into wavelet form and then get back exactly the same signal back again. It also means that a small change in the wavelet representation produces a correspondingly small change in the signal (Grossmann and Morlet 1984).

Some typical wavelet applications in remote sensing include: speckle reduction, image compression, texture analysis and classification, automatic geometric registration of images, and merging multiresolution satellite images (See the following articles: Horgan 1998; Singh 1999; Singh and Harrison 1985; Zhu and Yang 1998; Djamdji et al. 1993; Zhou et al. 1998; Yocky 1996; Núñez et al. 1999; Cakir et al. 1999; Ranchin and Wald 2000). Utilization of wavelets for data fusion in remote sensing is best explained with the ARSIS method (Ranchin and Wald (1993); Also see Ranchin and Wald (2000)). ARSIS, the French acronym for *Amélioration de la Résolution Spatiale par Injection de Structures*, enables the fusion of images having different spatial and spectral resolutions. That is, it allows the ability to insert details from high-spatial resolution data (e.g., SPOT panchromatic image) to high-spectral resolution data (e.g., SPOT multispectral). With this method, details extracted from a panchromatic image by multiresolution wavelet decomposition are used to estimate details that are missing in successive resolutions (the spatial resolution difference between panchromatic image and multispectral image) in multispectral imagery.

Those processing methods that represent the closest colors to the original image color on the upper left side (IKONOS MS 4-m) indicate the best performance. In this case, the Cakir-Khorram (1-m) method has produced the best results. The data fusion techniques are applied to produce a 1-m multispectral data by fusing 4-m multispectral data with 1-m panchromatic data.

International Policies Governing Remotely Sensed Data

Each day, millions of individual Earth observations are collected, allowing the ability to examine, monitor, and model climate change, ecosystem health, atmospheric composition, seismic activity, terrestrial and marine resources, agricultural monitoring, weather patterns, humanitarian violations, and hundreds of other phenomena. As of 2015, there are now more than 330 Earth-observation (EO) satellites launched by 30 countries, compared with just three in 1980, as well as an increasing number of countries which have their own image-receiving stations for remote sensing systems, due to the reduction in acquisition costs (de Montluc 2009; UCS 2015).

Historically speaking, remote sensing was originally developed for military applications, with nearly 75 % of satellites employed for military reconnaissance and surveillance purposes (Keeley and Huebert 2004; Jakhu 2004). The launch of the civilian American satellite, Landsat-1, in 1972, however, facilitated widespread Earth observation by researchers.

In 1959, shortly after launching the first satellite, the United Nations General Assembly established the Committee on the Peaceful Uses of Outer Space (COPUOS) as the only international forum for the development of international space law. Since its inception, the committee has concluded five international legal instruments and five sets of legal principles governing space-related activities.

International law and policies generally focus on three primary areas:

1. The right to acquire remotely sensed imagery/the right to launch remote sensing satellites
2. The right to disseminate remotely sensed imagery without prior consent of the sensed state
3. The right to obtain remotely sensed satellite imagery from a particular state

When Earth observation was in its infancy, the prevailing attitude was one of global scientific cooperation, particularly with regard to meteorological data sharing to warn of extreme weather events. Moreover, the earliest systems operated at large-scale (spatial, temporal, and spectral) resolution. As the worldwide availability of high-quality unclassified satellite imagery has dramatically increased, however, there has been increasing concern that this could be used for unsanctioned military or terrorist purposes. Consequently, international and national policies on the acquisition and distribution of remotely sensed satellite imagery have changed over time.

After extensive discussions in the COPUOS, the United Nations General Assembly unanimously adopted the resolution, *Principles Relating to Remote Sensing of the Earth from Outer Space*, in 1986. Although the principles are not a binding source of international law per se, this landmark resolution is regarded by many space law researchers as the primary international legal document that addresses issues of remote sensing (Harris 2003; Jakhu 2004; Macauley 2005; Rao and Murthi 2006).

As such, the principles are regarded a codification of customary law and have acquired the "evidence of a general practice accepted as law" according to Article 38 (1)(b), Statute of International Court of Justice (Gabrynowicz (1993); also see Williams (2006)). The 15 principles place international customary obligations on states and form the basis for remote sensing activities globally, regulating cooperation between sensing and sensed states. In particular, the principles require that sensed states should have access to data without any discriminatory practices and at "reasonable cost." Further, states are urged to consult with sensed states for mutual cooperation and benefit in the use of data, and states that are engaged in remote sensing should make technical assistance accessible to other interested states on "mutually agreed terms."

In addition to the UN principles, a number of other factors have contributed to a presumption of open access. Nondiscriminatory access policies have been adopted by major remote sensing nations (e.g., Japan, the United States, and Canada) and the data policies of some remote sensing missions (e.g., ENVISAT, RADARSAT) specifically incorporate nondiscriminatory access (Harris 2003). At present, there are at least 65 statutes worldwide that govern access to information, of which at least 50 establish a right of access to information, rather than a mere limited right of access to documents (United Nations Resolution 41/65 1986). The right of access to environmental information is, in certain circumstances, guaranteed by the European Convention on Human Rights.

International and national laws and policies are dynamic and ever-changing in response to changes in politics, technologies, as well as to real or perceived risks to national security. While it is difficult to know what new changes will be put into action over the near future, there are some indications of several policy directions. Internationally, two relatively new organizations involved with remote sensing are the **Committee on Earth Observation Satellites** (CEOS) and the voluntary partnership of governmental and intergovernmental organizations, **Group on Earth Observations** (GEO). CEOS works toward coordinating international earth observation systems and activities to meet the common good of member states, with special attention paid to developing countries. In response to an appeal for more international cooperation and coordination in data sharing regarding atmospheric, land, and water data, GEO – a collaborative of over 130 governments and international organizations – was formally created by resolution at the third Earth Observation Summit (EOS), held in Brussels in February 2005. In creating GEO on a voluntary and legally nonbinding basis, the founding governments and international organizations represented resolved that GEO would establish the Global Earth Observation System of Systems (GEOSS) by 2015 using the 10-year [2005–2015] Implementation Plan. The expected benefits for nine "Societal Benefit Areas[3]" plan rely on cooperatively sharing Earth observation data by GEO members and participating organizations through the GEOSS. A major initiative of the plan is the establishment of the GEOSS Data Collection of Open Resources for Everyone (Data-CORE). Many challenges remain to be resolved including figuring out who pays for infrastructure, training, and administration; whether to control data access; how to include the private sector; and whether problems of collective action will continue to hamper the effort.

Conclusion

With regard to national laws and policies, as high-resolution (i.e., submeter resolution) imagery continues to be widely available to the public through a multitude of sources, the boundary between open (public) access and restricted (military) access has vanished. Consequently, shutter control is not a viable national security policy; there are numerous sources for acquiring high-quality imagery. We have shifted from an era in which a small handful of developed countries had access to high-resolution imagery to one in which virtually everyone will have this kind of access. Transparency offers both enormous benefits and challenges. As yet, while governments throughout the world are preparing for this new era of access and transparency; policies are predominantly ad hoc, reactive, and not based on a working knowledge of geospatial technology. A harmonized international framework of international legal norms that goes beyond the scope of the UN Remote Sensing Principles will be necessary to resolve these challenges.

Cross-References

- ▶ Developments in Hyperspectral Sensing
- ▶ Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing
- ▶ Electro-Optical and Hyperspectral Remote Sensing

---

[3] The GEOSS 10-year Implementation Plan includes nine "Societal Benefit Areas": disasters, health, energy, climate, water, weather, ecosystems, agriculture, and biodiversity.

▶ Geographic Information Systems and Geomatics
▶ Introduction and History of Space Remote Sensing
▶ Lidar Remote Sensing
▶ Remote Sensing Data Applications

References

T.E. Avery, G.L.L. Berlin, *Fundamentals of Remote Sensing and Air Photo Interpretation*, 5th edn. (Prentice Hall, New York, 1992). 472 p

E.J. Beh, Simple correspondence analysis: a bibliographic review. Int. Stat. Rev. **72**(2), 257–284 (2004)

J.P. Benzécri, *Correspondence Analysis Handbook* (Marcel Dekker, New York, 1992)

H.I. Cakir, S. Khorram, X.L. Dai, P. de Fraipont, Merging SPOT XS and SAR imagery using the wavelet transform method to improve classification accuracy, in *Geoscience and Remote Sensing Symposium, IGARSS'99 Proceedings, IEEE International*, Hamburg **1**, 71–73 (1999)

H.I. Cakir, S. Khorram, S.A. Nelson, Correspondence analysis for detecting land cover change. Remote Sens. Environ. **102**, 306–317 (2006)

J.R. Carr, K. Matanawi, Correspondence analysis for principal components transformation of multispectral and hyperspectral digital images. Photogramm. Eng. Remote Sens. **65**(8), 909–914 (1999)

X. Dai, S. Khorram, Quantification of the impact of misregistration on digital change detection accuracy, in *Proceedings of IEEE/IGARSS'97 International Geoscience and Remote Sensing Symposium,* Singapore, 1997

X. Dai, S. Khorram, Data fusion using artificial neural networks: a case study on multi-temporal change analysis. Comput. Environ. Urban Syst. **23**, 19–31 (1999)

B. de Montluc, The new international political and strategic context for space policies. Space Policy **25**, 20–28 (2009)

J.P. Djamdji, A. Bijaoui, R. Manieri, Geometrical registration of images: the multiresolution approach. Photogramm. Eng. Remote Sens. **59**, 645–653 (1993)

J.R. Dobson, E.A. Bright, R.L. Ferguson, D.W. Field, L.L. Wood, K.D. Haddad, H. Iredale, J.R. Jensen, V. Klemas, R.J. Orth, J.P. Thomas, *NOAA Coastal Change Analysis Program (C-CAP): Guidance for Regional Implementation* (National Oceanic & Atmospheric Administration, NMFS, Washington, DC, 1995). 92 p

B.C. Forster, Derivation of atmospheric correction procedures for Landsat MSS with particular reference to urban data. Int. J. Remote Sens. **5**, 799–817 (1984)

J.I. Gabrynowicz, The promise and problems of the Land Remote Sensing Policy Act of 1992. Space Policy **9**, 319–328 (1993)

H.G. Gauch Jr., *Multivariate analysis in community structure* (Cambridge University Press, Cambridge, 1982)

M.J. Greenacre, *Theory and Application of Correspondence Analysis* (Academic, London, 1984)

A. Grossmann, J. Morlet, Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM J. Math. Anal. **15**, 723–736 (1984)

R. Harris, Current policy issues in remote sensing: report by the International Policy Advisory Committee of ISPRS. Space Policy **19**, 293–296 (2003)

M. Herold, S. Guenther, K.C. Clarke, Mapping urban areas in the Santa Barbara south coast using IKONOS and eCognition. eCognition Appl. Note. Munchen: Definiens ImgbH **4**(1), 20 p (2003)

H.O. Hirschfield, A connection between correlation and contingency, in Mathematical Proceedings of the Cambridge Philosophical Society, Oxford **31**, 520–4 (1935)

G. Horgan, Wavelets for SAR image smoothing. Photogramm. Eng. Remote Sens. **64**(12), 1171–1177 (1998)

B. Jähne, *Digital Image Processing* (Springer, New York, 1991), pp. 219–230

R. Jakhu, International law governing the acquisition and dissemination of satellite imagery, in *Commercial Satellite Imagery and United Nations Peacekeeping: A View from Above* (Ashgate, Burlington, 2004). 259 p

J.R. Jensen, *Introductory Digital Image Processing*, 3rd edn. (Pearson Prentice Hall, Upper Saddle River, 2005). 316 p

J.F. Keeley, R.N. Huebert, *Commercial Satellite Imagery and United Nations Peacekeeping: A View from Above* (Ashgate, Burlington, 2004)

S. Khorram, Coastwatch – water quality mapping of the entire San Francisco Bay and delta from Landsat multispectral scanner data, in *Space Sciences Laboratory*, ed. by: R.N. Colwell, PI, Series 23, Issue 6 (University of California, Berkeley, 1982), 34 p

S. Khorram, F. Koch, C.F. van der Wiele, S.A.C. Nelson, *Remote Sensing* (Springer, New York, 2012). doi:10.1007/9781-4614-3103-9. 134 p. ISBN 978-1- 4614-3102-2

S. Khorram, C.F. van der Wiele, F.H. Koch, S.A.C. Nelson, M.D. Potts, *Principles of Applied Remote Sensing* (Springer, New York, 2016). 307 p. ISBN 978-3-319-22559-3

P. Legendre, L. Legendre, *Numerical Ecology: Second English Edition* (Elsevier Science, Amsterdam, 1998). 823 p

T. Lillesand, R. Kiefer, J. Chipman, *Remote Sensing and Image Interpretation*, 6th edn. (Wiley, New York, 2008). 763 p

R.S. Lunetta, C. Elvidge, *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications* (Taylor & Francis, New York, 2000). 340 p

M.K. Macauley, Is the vision of the earth observation summit realizable? Space Policy **21**, 29–39 (2005)

S.A.C. Nelson, P.A. Soranno, J. Qi, Land cover change in the upper Barataria Basin estuary, Louisiana, from 1972–1992: increases in wetland area. Environ. Manag. **29**(5), 716–727 (2002)

J. Núñez, X. Otazu, O. Fors, A. Prades, V. Palà, R. Arbiol, Multiresolution-based image fusion with additive wavelet decomposition. IEEE Trans. Geosci. Remote Sens. **37**(3), 1204–1211 (1999)

K. Pearson, On lines and planes of closest fit to systems of points in space. Philos. Mag. **2**, 559–579 (1901)

C. Pohl, V. Genderen, Multisensor image fusion in remote sensing: concepts, methods and applications. Int. J. Remote Sens. **19**(5), 823–854 (1998)

T. Ranchin, L. Wald, The wavelet transform for the analysis of remotely sensed images. Int. J. Remote Sens. **14**, 615–619 (1993)

T. Ranchin, L. Wald, Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation. Photogramm. Eng. Remote Sens. **66**(1), 49–61 (2000)

M. Rao, K.R.S. Murthi, Keeping up with remote sensing and GI advances – policy and legal perspectives. Space Policy **22**, 262–273 (2006)

M. Robinson, Looking over the limb. LROC, http://lroc.sese.asu.edu/posts/895 (2015)

J. Rogan, D. Chen, Remote sensing technology for mapping and monitoring land cover and land-use change. Progress in Planning **61**, 301–325 (2004)

J.C. Russ, *The Image Processing Handbook* (CRC Press, Boca Raton, 2002). 744 p

N.M. Short, The remote sensing tutorial. Published by NASA via Internet, (2003), http://rst.gsfc.nasa.gov/

V.K. Singh, Discrete wavelet transform based image compression. Int. J. Remote Sens. **20**, 3399–3405 (1999)

A. Singh, A. Harrison, Standardized principal components. Int. J. Remote Sens. **6**, 883–896 (1985)

P.H. Swain, S.M. Davis, *Remote Sensing: The Quantitative Approach* (McGraw-Hill, New York, 1978), pp. 166–174

J.R.G. Townshend, C.O. Justice, C. Gurney, J. McManus, The impact of misregistration on change detection. IEEE Trans. Geosci. Remote Sens. **30**, 1054–1060 (1992)

R.E. Turner, M.M. Spencer, Atmospheric model for correction of spacecraft Data, in *Proceedings of the Eighth International Symposium on Remote Sensing of the Environment*, Ann Arbor, vol. 11, 1972, pp. 895–893

Union of Concerned Scientists (UCS), UCS Satellite Database (2015), http://ucsusa.org/satellites

United Nations Resolution 41/65. Principles relating to remote sensing of the Earth from outer space. Adopted without a vote (1986), http://www.oosa.unvienna.org/oosa/en/SpaceLaw/gares/html/gares_41_0065.html

M. Williams, *Legal Aspects of the Privatization and Commercialization of Space Activities, Remote Sensing, and National Space Legislation*. Second report. (International Law Association, Toronto, 2006), p. 2

D.A. Yocky, Multiresolution wavelet decomposition image merger of Landsat thematic mapper and SPOT panchromatic data. Photogramm. Eng. Remote Sens. **62**(9), 1067–1074 (1996)

H. Yuan, C.F. Van Der Wiele, S. Khorram, An automated artificial neural network system for land use/land cover classification from Landsat TM imagery. Remote Sens. **1**(3), 243–265 (2009)

J. Zhou, D.L. Civco, J.A. Silander, A wavelet transform method to merge Landsat TM and SPOT panchromatic data. Int. J. Remote Sens. **19**(4), 743–757 (1998)

C. Zhu, X. Yang, Study of remote sensing image texture analysis and classification using wavelet. Int. J. Remote Sens. **19**(16), 3197–3203 (1998)

Processing and Applications of Remotely Sensed Data

Siamak Khorram, Stacy A. C. Nelson, Cynthia F. van der Wiele, and Halil Cakir

Contents

---

*Dr. Halil Cakir did not contribute to this article as an employee of the US Environmental Protection Agency nor does this article reflect the views of this agency.

S. Khorram (✉)
Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA

Center for Geospatial Analytics, North Carolina State University, Raleigh, NC, USA
e-mail: khorram@Berkeley.edu

S.A.C. Nelson
Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, USA

Center for Geospatial Analytics, North Carolina State University, Raleigh, NC, USA
e-mail: Stacy_Nelson@ncsu.edu

C.F. van der Wiele
Region IV NEPA Program Office, US Environmental Protection Agency, Research Triangle Park, NC, USA
e-mail: cynthia.vanderwiele@alumni.duke.edu

H. Cakir
Air Quality Analysis Group/AQAD/OAQPS, US Environmental Protection Agency, Research Triangle Park, NC, USA
e-mail: Cakir.Halil@epa.gov

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_92

Abstract

Digital image processing, post-processing, and data integration techniques as applied to airborne and satellite remotely sensed data for the purpose of extracting useful Earth resources information will be discussed in this chapter. Image preprocessing and data reduction tools are described in the previous chapter. The concepts discussed in this chapter include:

- Image processing techniques such as unsupervised image classifications, supervised image classifications, neural network classifiers, simulated annealing classifiers, and fuzzy logic classification systems
- The most widely accepted indices and land use/land cover classification schemes
- Post-processing techniques such as filtering and change detection
- Accuracy assessment and validation of results
- Data integration and spatial modeling including examples of integration of remotely sensed data with other conventional survey and map form data for Earth observation purposes

Keywords

Digital image processing • Supervised classifiers • Unsupervised classifiers • Filtering • Accuracy assessment classification schemes • Geospatial modeling • Image validation • Image visualization • Post-processing • Satellite remote sensing geospatial modeling

Introduction

Various image processing and post-processing techniques and data integration with

other data types in a geographic information system (GIS) for spatial modeling purposes of remotely sensed data will be discussed in this chapter. Image processing, data reduction, and preparation of image form data for processing have been described in a previous chapter.

Image Processing

Digital image processing techniques are employed for the improvement of image visual display and analysis, presentation of data in an orderly and meaningful form, classification of data into defined categories of interest, and integration of image form data with other conventional surveys and maps. These techniques involve visual image data displays, vegetation indices, commonly used image classification systems, and widely accepted land use and land cover schemes.

Data Display, Visualization, and Reduction Methods

Raw data displays are most frequently provided as displays of a given band, as either a panchromatic, true color composite (TCC), or false color composite (FCC) image. Any individual band of a multispectral digital image can be displayed as grayscale (panchromatic) image, where the lowest-value pixels are displayed as black, the highest-value pixels are displayed as white, and pixels with intermediate values are displayed in corresponding shades of gray. True color composite displays the blue band in blue color, the green band in green color, and the red band in red color, while false color composite displays the combination of any three bands from a multispectral image other than the true color. Essentially, a TCC depicts its features in natural color. The standard FCC provides an infrared band displayed in red color, the red band displayed in green color, and the green band displayed in blue color. The differences between TCC and standard FCC data displays are illustrated in Fig. 1. In the standard FCC, all vegetation appears in red, whereas in TCC, vegetation appears in green (Fig. 2) (Hester 2008; Khorram et al. 2016).

The TCC of an area in coastal North Carolina and the standard FCC of the same area are shown in Figs. 1 and 2, respectively.

Using finer resolution data based on QuickBird satellite data, the TCC and standard FCC are depicted in Figs. 3 and 4, respectively.

Fig. 1 The true color composite of Open Grounds Farm area in coastal North Carolina as seen from Landsat TM in which all vegetation appears in various shades of green and the fallow fields are in various shades of brown. (Image courtesy of Khorram et al. 2016)

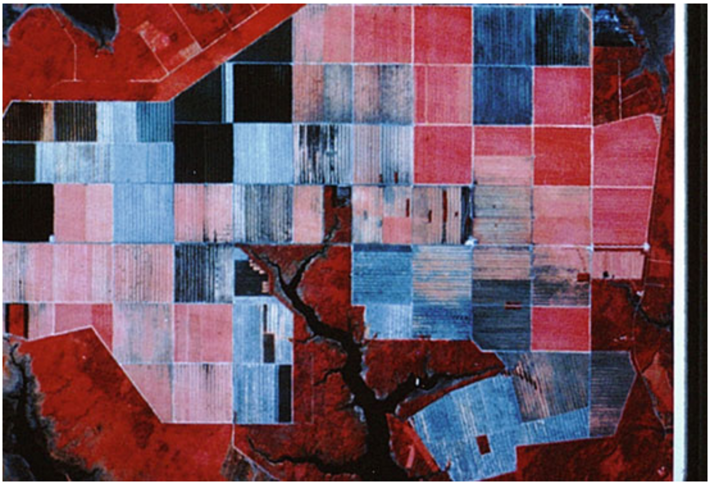


Fig. 2 The standard false color composite of Open Grounds Farm area in coastal North Carolina as seen from Landsat TM in which all vegetation appears in various shades of red and the fallow fields are in various shades of brown. (Image courtesy of Khorram et al. 2016)

Fig. 3 Two identical images illustrating false color composite (FCC) and true color composite (TCC) below (Graphic courtesy of QuickBird)

Band Combinations, Ratios, and Indices

Arithmetic combinations of certain bands via division, addition, subtraction, or multiplication can lead to enhanced information. Differencing and ratioing are primarily used for change detection and spectral enhancement studies. Band ratios typically include the following: infrared band over red band for vegetation distribution, green band over red band for mapping surface water bodies and wetland

Fig. 4 True color composite (TCC) as opposed to false color composite (FCC) as presented above (Graphic courtesy of QuickBird)

delineation, red band over infrared band for mapping turbid waters, and red band over blue band or red band over green band for mineral mapping, such as iron-rich or iron-poor rocks.

The most frequently used index for vegetation mapping is the **Normalized Difference Vegetation Index (NDVI),** which is defined as:

$$\text{NDVI} = \frac{(B2 - B1)}{(B2 + B1)}$$

where $B2$ represents the brightness values (i.e., the digital numbers or pixel values) from a near-infrared band of an image and $B1$ represents the corresponding values in the image's red band. NDVI images can be displayed in black and white, as shown in Fig. 5, or in color, as shown in Fig. 6, which illustrates the global distribution of vegetation in various shades of green. NDVI images are typically used for covering large geographic areas, thus reducing the cost of data processing for applications such as vegetation mapping (Khorram et al. 2016).

Figure 6 illustrates the extent of vegetation in various shades of green color based on NDVI.

Image Classification Systems

Categorizing the pixels of a digital image into particular land cover classes or themes is referred to as *image classification*. Most image processing techniques are based on hard logic, utilizing both spectral and temporal spatial patterns. Several other techniques have been recently explored for processing remotely sensed data utilizing including fuzzy logic (Cho and Kim 1995) neural network and decision trees (Dai and Khorram 1997a), simulated annealing, and hybrid methodologies. In this

Fig. 5 A black and white NDVI image of a study area in Oxnard, coastal California, created from Landsat data, adapted from Khorram et al. 2016

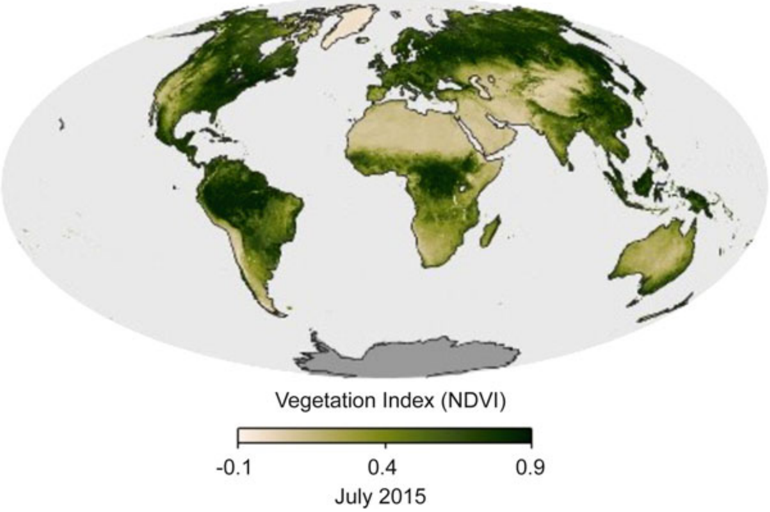


Fig. 6 Example of NDVI image depicting the global vegetation cover in various shades of *green* color (Graphic courtesy of NASA)

section, the focus will be on hard logic first and then briefly describe other classifiers. Image classification based on hard logic is divided into *supervised classification* and *unsupervised classification* as defined by Hester and others (Hester 2008).

Land cover, commonly referred to as land use/land cover classification scheme, must be utilized for identifying categories of interest. Three common classification schemes are based on the objectives of analysts. These are the US Geological Survey

(USGS) classification system; the Cowardin, Carter, Golet, and LaRoe system (Cowardin et al. 1979); and the NOAA Coastal Change Analysis Program (C-CAP) system (NOAA 2004) scheme (Anderson et al. 1976; USGS 2004) which is the predominant system used. It is designed to utilize remotely sensed data and is organized in four levels of progressively detailed information. Level I and Level II of these categories are most frequently used and shown in Table 1.

The other two most frequently used classification systems are primarily used for coastal and wetland analyses. These are the US Fish and Wildlife Classification System (Cowardin et al. 1979) and the National Oceanic and Atmospheric Administration (NOAA) Coastal Change Analysis Program (Klemas et al. 1993; also see Khorram et al. 1996).

Usually, there are overlaps and redundancies as information is contained in more than one band of multiband spectral data. The process of reducing this redundancy should be observed carefully so that one does not create an adverse impact on the accuracy of classified data through statistical and/or graphical analyses such as autocorrelation, bar graph spectral plots, feature spectral plots, and so forth.

Unsupervised Classification

Unsupervised classification methods generally use no or minimal analyst supervision to develop the resultant land use/land cover. This is a computerized process whereby each pixel is iteratively assigned to a class based on the similarity of the spectral properties of pixels in multiple bands. This is accomplished through the determination of class “spectral separability,” based on means and covariance matrices. The analyst assigns the categorical information to the classified data after classification. Two most common techniques for unsupervised classification are clustering algorithms and the Iterative Self-Organizing Data Analysis Technique (ISODATA).

Using the *clustering algorithms* approach, clustering is performed in two stages. During the first stage, a number of clusters are built up sequentially. Each cluster is composed of pixels having similar spectral values (within a range that is determined by the analyst), therefore occupying a common spectral space (Jain 1989; Celik 2009) with a well-defined mean vector for each class. In the second stage, a minimum-distance-to-means classification algorithm, described earlier, is applied to the entire data set for assigning each pixel to a cluster.

The *Iterative Self-Organizing Data Analysis Technique (ISODATA)* (Tou and Gonzalez 1977; also see Sabins 1987) is a very widely used multiple iterative procedure with minimal analyst supervision for processing remotely sensed digital data. ISODATA utilizes the mean and standard deviation in a number of bands in n-dimensional space. The parameters determined by the analyst initially to guide the algorithm include the maximum number of clusters, maximum percentage of pixels that can remain unchanged between iterations, maximum number of iterations, minimum percentage of pixels assigned to each cluster, maximum standard deviation, and minimum distance between clusters (Jensen 2005; also see Hester 2008). An example of an application of unsupervised classification for the City of San Francisco and surrounding areas is shown in Fig. 7.

Table 1 Land use and land cover classification system for the use with remote sensor data (Modified from Anderson et al. 1976)

Level I	Level II	
1. Urban or built-up land	11	Residential
	12	Commercial and services
	13	Industrial
	14	Transportation, communications, and utilities
	15	Industrial and commercial complexes
	16	Mixed urban or built-up land
	17	Other urban or built-up land
2. Agricultural land	21	Cropland and pasture
	22	Orchards, groves, vineyards, nurseries, and ornamental horticultural areas
	23	Confined feeding operations
	24	Other agricultural land
3. Rangeland	31	Herbaceous rangeland
	32	Shrub and brush rangeland
	33	Mixed rangeland
4. Forest land	41	Deciduous forest land
	42	Evergreen forest land
	43	Mixed forest land
5. Water	51	Streams and canals
	52	Lakes
	53	Reservoirs
	54	Bays and estuaries
6. Wetland	61	Forested wetland
	62	Non-forested wetland
7. Barren land	71	Dry salt flats
	72	Beaches
	73	Sandy areas other than beaches
	74	Bare exposed rock
	75	Strip mines, quarries, and gravel pits
	76	Transitional areas
	77	Mixed barren land
8. Tundra	81	Shrub and brush tundra
	82	Herbaceous tundra
	83	Bare ground tundra
	84	Wet tundra
	85	Mixed tundra
9. Perennial snow or ice	91	Perennial snowfields
	92	Glaciers

Supervised Classification Systems

Supervised classifications are a three-stage process performed by the analyst comprised of training, classification, and output.

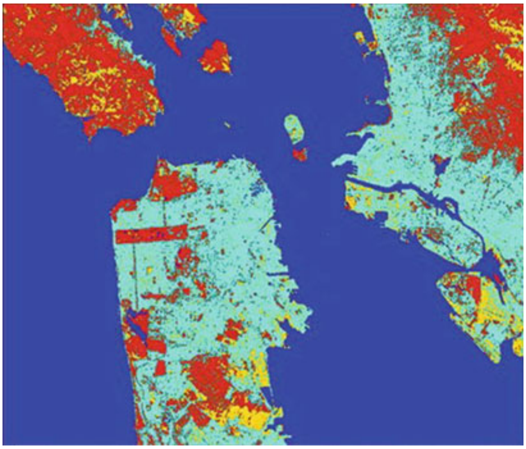


Fig. 7 Example of unsupervised classification depicting the City of San Francisco and surrounding area, California, USA (Graphic courtesy of Siamak Khorram)

During the *training stage*, training sites are selected by the analyst to represent areas with known cover types. In this stage, the analyst establishes the relationships between the cover types and the spectral data in multiple wavelength bands. Other data types such as aerial photography, ancillary spatial data, or field visits are typically used for establishing these relationships. The number of training sites should be at least three times the number of categories of interest. Training sites are usually selected to represent the spectral distribution of each cover type of interest and are randomly or systematically distributed throughout the study area. The training sites should each represent a completely homogenous window, which is a small portion of the full scene covering a given cover type.

The second stage of supervised classification is the *classification stage*: In order to differentiate spectral bands into accurate land use/land cover categories, a number of classification and pattern recognition algorithms have been developed for supervised classification. The most widely used classification algorithms include minimum distance to means, parallelepiped, and maximum likelihood and are briefly described.

The *minimum-distance-to-means classification algorithm* determines the mean spectral values for each category in each band, and then the spectral domains for multiple bands are computed. Each land cover category has a well-identified spectral space. A pixel of unknown identity (unassigned pixel) can be assigned to the category which has the minimum distance between the unknown pixel value and the mean vector in the category of interest. Although this method is simple and computationally efficient, it is insensitive to various degrees of variance in spectral response data (Lillesand et al. 2008), which eventually results in decreased classification accuracy.

Spectral variance is taken into account in the *parallelepiped classification algorithm* by establishing a range for each category in multiple bands to establish a decision region. The ranges are defined as the lowest and highest brightness values arranged in a rectangular or stepped rectangular scattergrams called parallelepipeds.

Fig. 8 Unclassified QuickBird satellite imagery of Black Creek Watershed, Cary, North Carolina, acquired on June 29, 2004 (*left*) and the associated land use/land cover thematic classification map (*right*) based on maximum likelihood classifier

This method is also simple, computationally efficient, and allowed for variance. But covariance, i.e., the tendency of spectral values to vary similarly in two or more bands, can be poorly taken into account at best. In this case, the interdependency of bands to each other is removed to an extent (Lillesand et al. 2008).

The *maximum likelihood classification algorithm* assumes the normal, Gaussian, distribution of training data statistics for each class in each band (Blaisdell 1993). Given this assumption, the distribution of spectral responses in each category can be described in mean vector and covariance matrix. The statistical probability of a given pixel belonging to a category of interest can then be determined. For a three-band data, the probability density functions. The unknown (unassigned) pixel can then be assigned the category of the highest probability value, which is maximum likelihood of belonging. The maximum likelihood classifier is based on equiprobability contours. The shape of these contour lines is due to their sensitivity to covariance (Lillesand et al. 2008). The maximum likelihood classification algorithm is the most commonly used method, but it is not as computationally efficient as the other methods described above. Figure 8 illustrates an application of the maximum likelihood classifier over the Black Creek Watershed, Cary, North Carolina [USA], and the associated land use/land cover map of the same area based on QuickBird satellite data.

Finally, a more complex version of the maximum likelihood classifier is the Bayesian classifier, in which a certain weight is assigned to probability estimates based on a prior knowledge of the anticipated likelihood of occurrence (Hester et al. 2008; also see Hord 1982).

The final stage is known as the *output stage*. Output products are used for presentation and visualization of the results. This is done by regrouping the results of classified data into a desired group of classes as determined by the analyst and presented in digital form, hard copy output product form (film, transparencies, paper, slides, etc.), or in graphic and tabular forms. The output product is accompanied by

Fig. 9 A comparison of true color composite (TCC) image on the left and a classified image of the same area based on maximum likelihood classifier on the right

the statistical parameters, accuracy assessment table, and other supporting information.

Figures 9 and 10 are examples of a classified image based on the supervised maximum likelihood classification system showing progressively more detailed information as compared to the TCC of the same areas.

Neural Network Classifiers

Artificial neural network (ANN) approaches have been widely used for image classification in remote sensing since the 1990s (Qiu and Jensen 2004; Dai and Khorram 1999; Yang and Chung 2002). The automated system consists of two ANN classification modules: (1) a Kohonen's Self-Organizing Mapping (KSOM) module based on *unsupervised* KSOM neural networks and (2) a multilayer perceptron (MLP) module based on *supervised* MLP neural networks that uses single or multilayer perceptrons to approximate the inherent input–output relationships that is the most commonly used network model for image classification in remote sensing (Kanellopoulos and Wilkinson 1997). MLP networks are typically trained with the supervised back propagation (BP) algorithm (Rumelhart et al. 1986) and consist of one input layer, one or more hidden layers, and one output layer. In the ANN system, each module is composed of several sub-modules: pattern conversion, network training, and the network generalization sub-module. A working flowchart of the ANN-based classification system and functions of each sub-module are summarized in Fig. 11 (Nogami et al. 1997; also see Xu and Wunsch 2005).

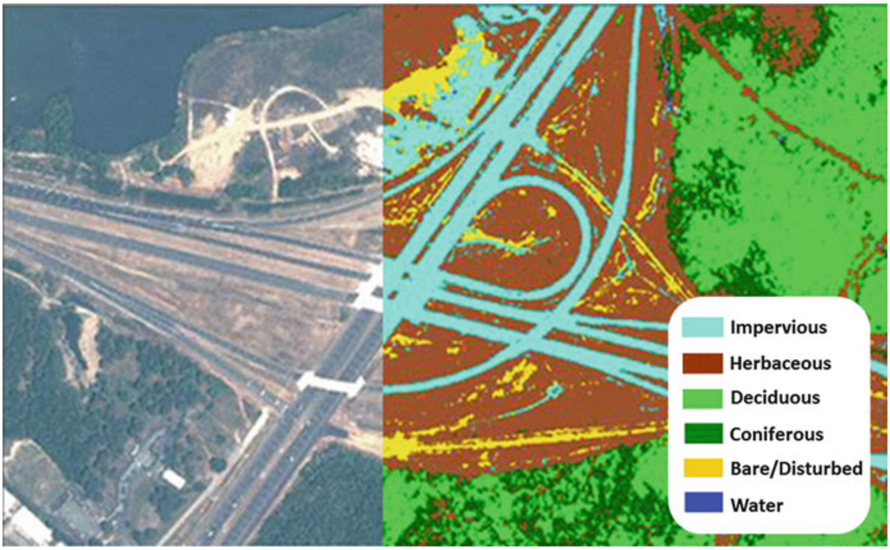


Fig. 10 An illustration of a more detailed classification based on maximum likelihood classifier as compared to a true color composite (TCC) of the same area. The image on the left represents a TCC display, and the image on the right a classified image of the same area based on maximum likelihood classifier

The pattern conversion sub-module performs the following functions: (1) sampling a certain number of training and testing patterns from a number of selected image subsets, (2) scaling the input pattern into the network operational interval, and (3) generating training or testing pattern files. In the pattern conversion sub-module of the supervised MLP network, the corresponding class label must be provided for each pattern. Network training sub-modules provide the graphical user interfaces to allow the user to interactively define the architecture and parameters needed and to perform the training once all the parameters are set. In the KSOM module, two training sub-modules are provided: the standard KSOM and the KSOM-SA training sub-module. In the MLP module, the BP training sub-module is used. During a network training trial using each of the training sub-modules, an error file is generated to record the training MSEs to assist in monitoring the training behavior and selecting the appropriate network and parameters. After training is completed, network generalization sub-modules are implemented to generalize the entire image using the trained network and to produce the classified map. The application of this methodology for land use/land cover classification is illustrated in Fig. 12a–c using an example in Coastal North Carolina [USA].

Simulated Annealing Classifiers

The commonly used *K*-means-type algorithms are known to produce good results only if the clusters are well separated in the feature space and are hyper-spherical in

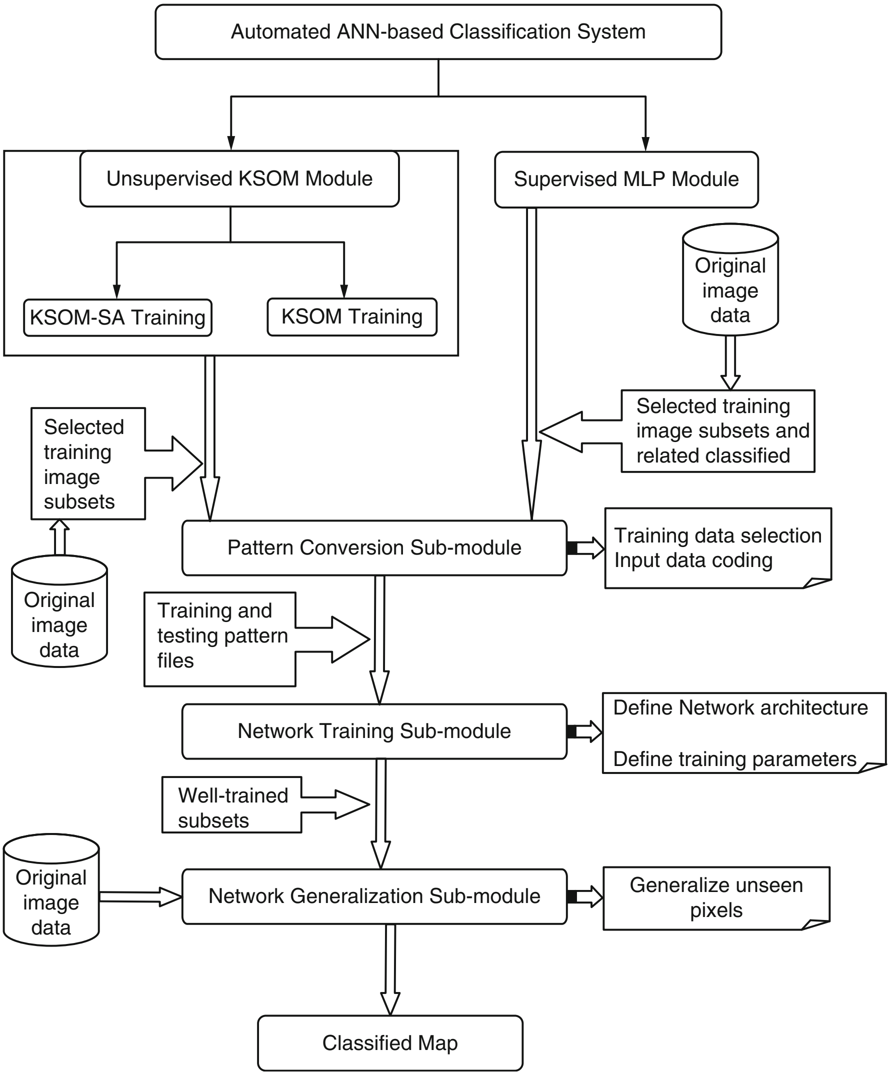
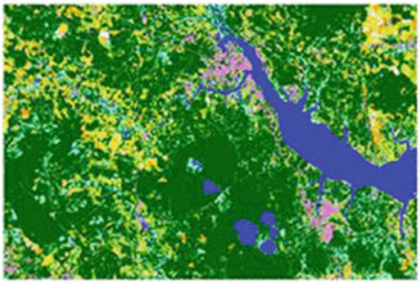


Fig. 11 Flowchart of an automated ANN-based classification system

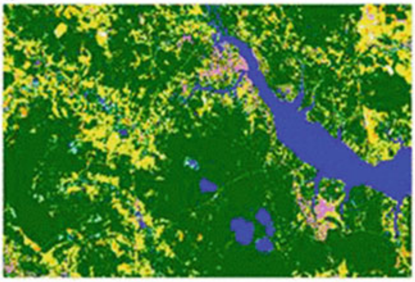
shape when *Euclidean* distance is used. However, in a complex problem that cannot be solved by a simple convex cost function, the local minima problem is inevitable.

Alternatively, simulated annealing (SA) was developed on the basis of an analogy between the physical annealing process of solids and the large combinatorial optimization problems (Das and Chakrabarti 2005. Also see De Vincente et al. 2003). SA was proven to have great potential to find or approximate the global or near-global optimal in a combinatorial optimization problem. The premise of SA is to incorporate some randomness in the assignments of cluster labels to pixels in the a



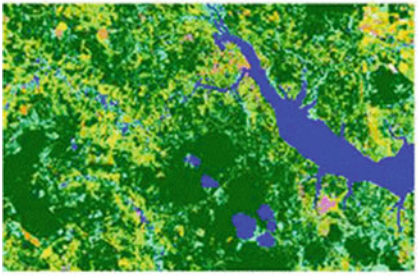
Classified map of the MLP network

b



Classified map of the KSOM network

c



Classified map of the KSOM-SA network

Fig. 12 (**a**–**c**) Classified images of Coastal North Carolina, USA, based on three versions of the ANN classification method (Graphic courtesy of USGS)

clustering procedure, thus reducing the limitation of local minima. As a result, using the SA-based approach to classification has the potential to improve the accuracy for land cover classification. Simulated annealing (SA) was developed on the basis of an analogy between the physical annealing process of solids and large combinatorial optimization problems (Kirkpatrick et al. 1983; see also Cerny 1985).

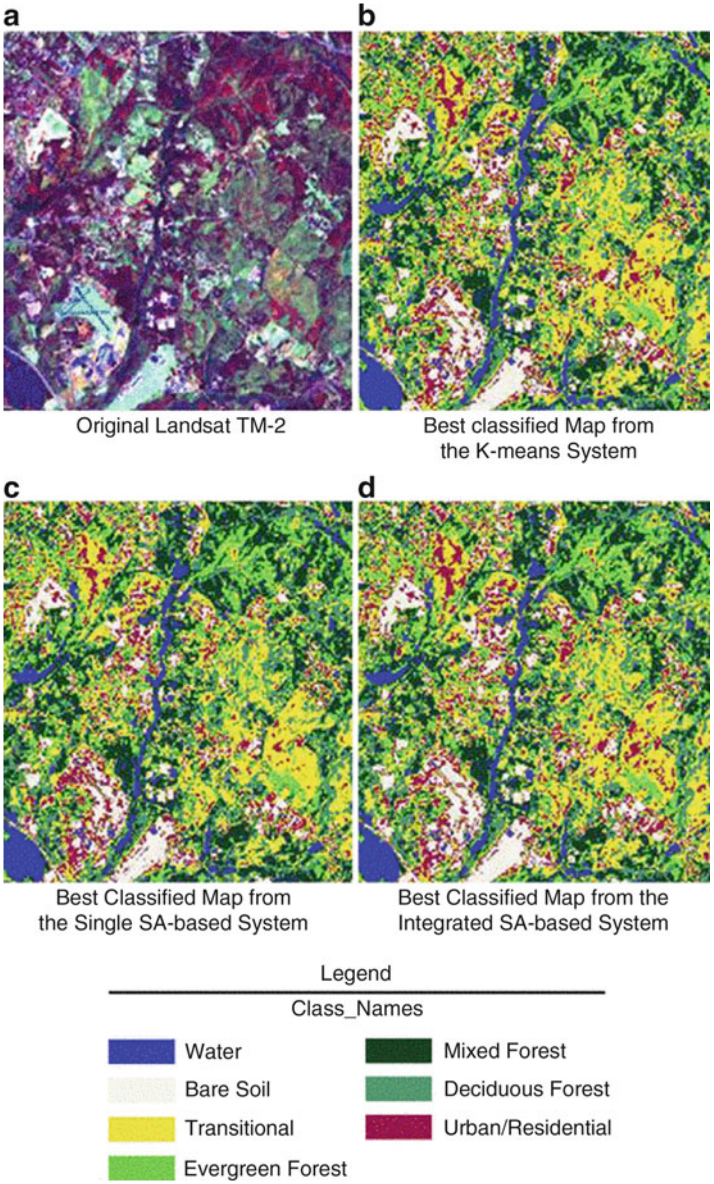
SA has the potential to find or approximate the global or near-global optimal in a combinatorial optimization problem. Although SA often requires greater computational time, the improved network can self-adapt to choose momentum parameters according to annealing temperature, thus enabling the network to escape from local minimum spots and converge stably. Mathematically, SA can be modeled by means of a Markov chain. The basic procedure involves a cooling schedule, in which a parameter called temperature *T* starts out sufficiently high and is gradually lowered in a given schedule to minimize the energy or cost function associated with a specific problem formulation. At each temperature *T*, a small, randomly generated perturbation is repetitively applied until the system reaches thermal equilibrium at that temperature. Then the algorithm moves to the next temperature in the given schedule. The rule of accepting a perturbation is based upon the *Metropolis criterion* (Shurr 1997). The original image and single simulated annealing and K-means clustering classified maps, the integrated K-means, and simulated annealing (I-SA) are shown in Fig. 13a–d.

Hyperspectral Data Classification

Hyperspectral remote sensing has been an evolving new technology in the world of remote sensing. In this process, hyperspectral sensor collects and processes spectral information across a broad range of the electromagnetic spectrum, at typically very narrow bandwidths. The narrower bandwidths of the hyperspectral sensor increase the ability to detect specific features on the Earth's surface, such as certain minerals or even subtle variances between similar structures, that may have strong reflectances within very narrow ranges in the electromagnetic spectrum. Additionally, these types of sensors are usually able to capture data using a large number of bands. For example, in comparison to the Landsat program sensors, the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) hyperspectral sensor records 224 different bandwidths that range between 400 and 2500 um.

Airborne LiDAR Data Processing

The processing of an airborne Light Detection and Ranging (LiDAR) data set is substantially different from multispectral or hyperspectral image processing primarily due to the nature of the data. (Airborne LiDAR data is also qualitatively different from spaceborne LiDAR data captured by sensors such as CALIOP, although some of the processing issues are similar; see Winker et al. 2009.) An airborne LiDAR data set consists of a dense (often greater than one pulse per square meter), three-dimensional, and geospatially referenced point cloud, where each point documents a single elevation estimate for a reflective object scanned by the system's laser. The resulting point cloud must be further filtered to separate points that consist of non-ground and true-ground returns. Once these returns have been separated, the resulting true-ground returns may be further processed to constitute a bare-Earth digital terrain model. This bare-Earth digital terrain model represents the surface of the Earth with all features above the bare ground removed. The non-ground measurements must further be classified according to the category of object they represent. Typically, most non-ground measurements in a LiDAR data set will correspond to either buildings or vegetation.



Original Landsat TM-2

Best classified Map from the K-means System

Best Classified Map from the Single SA-based System

Best Classified Map from the Integrated SA-based System

Fig. 13 (**a**–**d**) Land use/land cover using simulated annealing techniques (Graphic courtesy of NOAA)

Fuzzy Logic Classifiers

In order for remotely sensed land use/land cover classification data to be the most useful, the post-classification results should resemble accurate representation of the ground features being investigated. The incorporation of unsupervised and

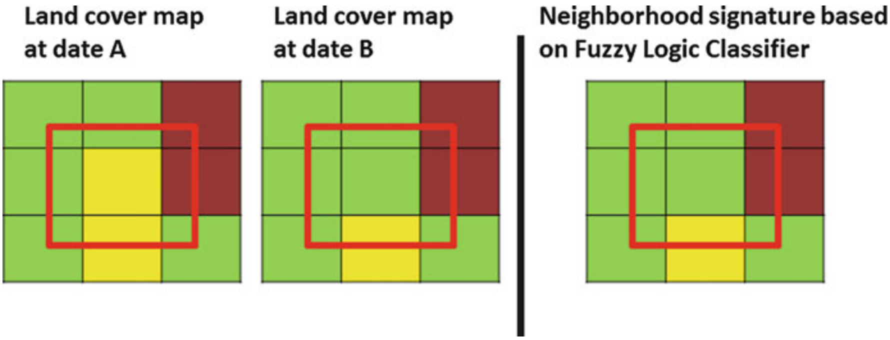


Fig. 14 Illustration of an application of fuzzy classification technique for land use/land cover mapping. In this example, the fuzzy logic classification algorithm is applied to the target cell (highlighted in *red*). At date A, the target cell is yellow; at date B the target cell is *green*. The resulting thematic output classifies the target cell as green based on class memberships of neighboring pixels in relation to the (spatial) distance from the target pixel between dates A and B.

supervised classification techniques, based on training sets and maximum likelihood or other spatial algorithms, develops classification groupings of objects belonging to exclusive categories. Often for general or moderate resolution land cover analysis, these methods are acceptable. However, in situations where highly accurate land cover data is required, it is important to note that the real world often does not conform to the "hard" rules imposed by the algorithms or boundaries that result from processes such as maximum likelihood, nearest neighbor, etc.

Fuzzy logic classifiers (Fig. 14) may be used to circumvent the "hard" classification boundaries by allowing for a transition or "fuzzy" region to be established between classes when a pixel has membership in two or more classification categories (Jensen 2005). The fuzzy region is established based on the vector direction of the spectral measurement space distance from the means of all established classes. This gradation of class values allows for the class transition region or fuzzy class to represent the value of member each pixel has in the established classes.

For example, for any pixel in a multispectral image, it is possible to measure the distance between the pixel's spectral values and the mean vectors (i.e., the sets of mean spectral values) for all classes in the classification scheme. These measured distances may then be translated to the pixel's likelihood of membership in each class.

Additionally, fuzzy logic may be incorporated into thematic classification efforts, and fuzzy classifiers can resolve other issues besides the problem of *mixed* pixels. Studies by Hester found that the capacity of fuzzy logic sets was able to overcome post-classification accuracy limitations in the analysis of an urban watershed utilizing high-resolution data (Hester 2008; Hester et al. 2010). The use of fuzzy classifiers addressed two major sources of change detection error: individual-date map misclassification and image "misregistration." The calculated overall "from–to accuracy" achieved in this study was over 78 % for all classes.

Each pixel is described as a “fuzzy neighborhood vector,” and those vectors are compared between maps based on neighborhood size, distance–decay function, and neighborhood vector comparisons. The result is a “similarity value” that is a measure of each pixel’s inter-map neighborhood similarity ranging from 0 to 1 (Hagen 2003).

Image Post-processing and Smoothing

The post-processing phase is often conducted based on the objectives of the project. At times, filtering is needed for smoothing the data and removing speckles in the classified image for better correspondence to conventional survey maps. Other techniques utilized in post-processing include accuracy assessment of the classified images for verification and validation of the results, change detection and monitoring based on two or more dates of classified images, and integration of the classified images with other conventional survey maps and data layers in a geographical information system (GIS) for geospatial modeling.

Filtering

After the data is classified, a number of algorithms can be applied to the classified image for better presentation and visualization. Kernels, which can be thought as moving windows, are applied to classified images to reduce noise and filter unwanted information. An example of a 5 × 5 median filter which has been applied to a classified map is shown in Fig. 15. Median values are calculated within a moving 5 pixel by 5 pixel window and assigned into the central pixel. Mean and median filters usually have a smoothing effect. Other filters that have been frequently used in image processing use functions that highlight differences between pixel values. They are usually used for sharpening the edges of objects in an image.

Accuracy Assessment

Accuracy assessment has been a key component and the focus of a significant number of remote sensing studies (Van Genderen and Lock 1977; Congalton et al. 1983; Goodchild et al. 1992; Khorram et al. 1992; Congalton and Green 1999; Paine and Kiser 2003). Without assessing the accuracy of a classified data, the reliability and repeatability of the output products are in question. Sophisticated statistical procedures in the analysis of error matrices are developed for the accuracy assessment of land cover classifications for a single date (Congalton et al. 1983; Aaronoff 1985; Khorram et al. 1999; Stehman 2001; Foody 2002; Lunetta 2003; Pal and Mather 2003; Morisette and Khorram 2003). However, the accuracy assessment of change detection procedures are not in operational stage yet and involve issues such as not yet widely accepted sampling techniques, image registration, boundary problems, and reference data. The error sources involved in the accuracy assessment

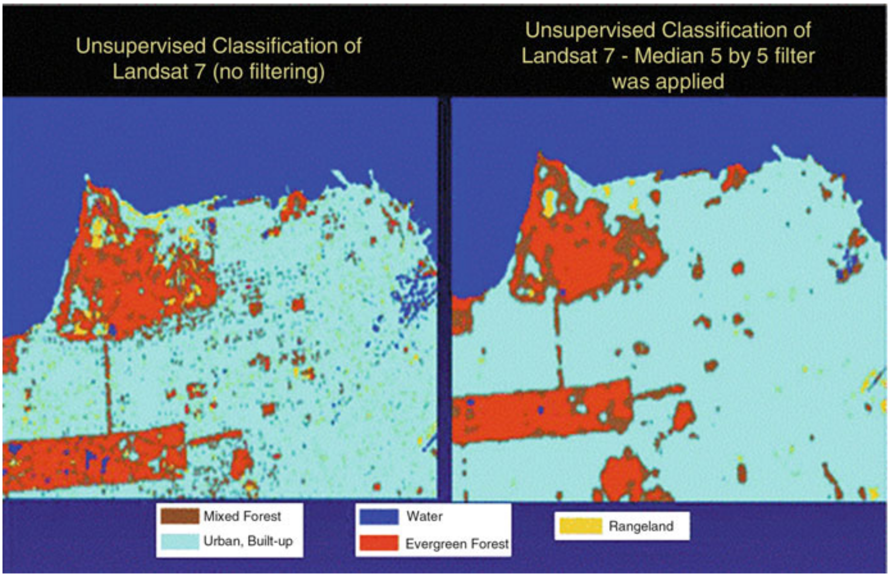


Fig. 15 A filtering technique has been applied to classified land use/land cover images on the *right* side as compared to a classified original image on the *left* side (Graphic courtesy of Siamak Khorram)

include registration differences between reference data and remotely sensed data, delineation errors in digitizing, data entry errors, errors in image classification and delineation, and errors involved in sampling, collection, and interpretation of reference data.

The most commonly used procedure for accuracy assessment is error matrix analysis. An error matrix can be constructed by the results based on reference data (e.g., data collected on the ground or from substantially higher spatial resolution image) on one side of a table and the results based on the classified image on the other side of the table. An adequate number of samples are identified on the classified image, and corresponding reference data are collected using various sampling strategies. The accuracy is determined in terms of percent correctly classified sample sites, as compared to their corresponding reference data, for each category of interest as well as the overall classification accuracy involving all categories. Traditionally, the total number of correct samples in a given category is divided by the total number of samples in that category based on reference data. This accuracy measure indicates omission errors and is often referred to as “producer’s accuracy” because the producer of image classification is interested in how well he has classified a certain category. If the total number of correct samples in a given category is divided by the total number of samples based on classified data, then this indicates the commission error. This measure is called the “user’s accuracy” or reliability because the user is interested in the probability that a classified sample represents the actual category on the ground (Story and Congalton 1986). Multivariate statistical procedures have also been used for accuracy assessment. The most commonly used is a discrete multivariate technique, called KAPPA, which is a measure of agreement or accuracy by KHAT statistics (Cohen 1960). KAPPA is computed from the error matrix table constructed from both reference and classified data sets:

$$K_{hat} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum \left(x_{i+} \times x_{+i}\right)}{N^2 - \sum_{i=1}^{r} \left(x_{i+} \times x_{+i}\right)}$$

where ***r*** is the number of rows in the matrix, X_{ii} is the number of observations in row i and column *i*, $_{Xi+}$ and X_{+i} are the marginal totals for row *i* and column *i*, respectfully, and N is the total number of observations.

Change Detection

Global environmental change has become a major national and international policy issue. Land cover change provides an important component to estimate and model changes in environmental and socioeconomic conditions resulting from regulatory or land use policy changes and is a potentially important indicator of the effects of local, national, and international policies on environmental quality and even human health. Remote sensing, combined with supporting data, provides the most feasible approach to land cover change detection. Issues specific to change analysis include the land cover classification scheme, hard logic versus fuzzy logic, large number of categories of change, the difficulty of field verification for past time periods, errors involved in classifications for both dates, remote sensor systems characteristics, digital image processing techniques, and environmental characteristics.

The spatial scale and information content of satellite-derived remote sensing data has inspired the development of automated change detection algorithms and methodologies, especially for evaluating and recording land use and land cover (LULC) change. These automated methods can be categorized as pre-classification and post-classification or even constitute more advanced procedures.

There are various methods for change detection. Some changes are easily detected through the visual analysis of color composites for the dates under consideration. Figure 16 illustrates change detection based on visual analysis of Advanced Very High-Resolution Radiometry (AVHRR) satellite data using an example of before and after oil wells that were set on fire in Kuwait. Figure 17 is another visual analysis illustration based on Landsat data from 1973 to 2000 to 2006. All urban and built up areas in this figure are shown in various gray colors and all vegetation in various green colors.

Pre-classification methods develop change maps from multi-temporal data (i.e., data captured over the same area on different dates) without first generating classified LULC maps from that data. The algorithms used in pre-classification procedures

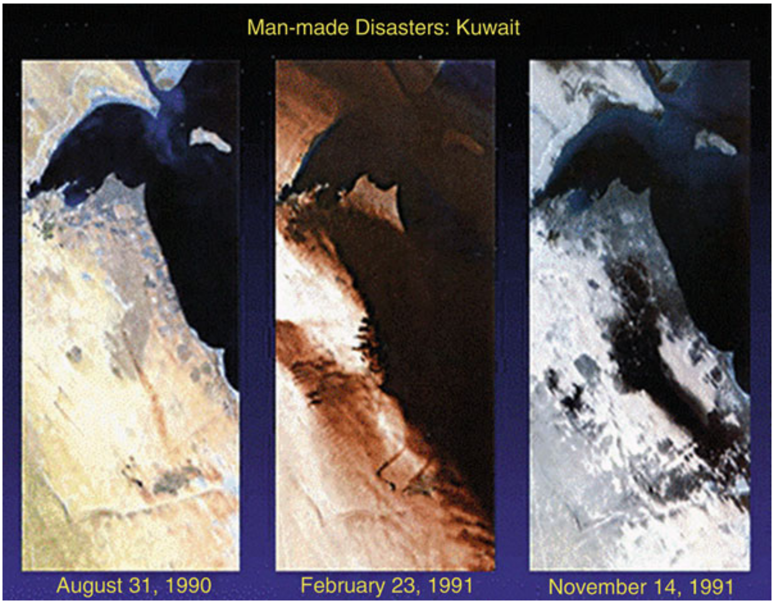


Fig. 16 Change detection as illustrated with images of before and after oil wells were set on fire in Kuwait (courtesy of Dai and Khorram 1997b)

may transform or simplify the original data before creating a change map, but they do not rely first on the generation of meaningful land use/land cover classifications of the individual image dates. One of the most important aspects of any pre-classification change detection algorithm is the specification of a change threshold. This parameter, either derived by the analyst using the initial algorithm output or by the analyst using only a priori knowledge, represents the interpretative mechanism by which the algorithm judges whether a change has occurred.

Another example of visual analysis of change detection based on the high-resolution QuickBird data is demonstrated in Fig. 18.

Visual analysis of temporary changes can also be conducted from satellite data. In Fig. 19, the 100-m resolution ESA Proba-V minisatellite images detail dramatic changes to Lake Poopó, Bolivia’s second largest lake.

Figure 20 illustrates the frozen ice swirl patterns off the coast of Greenland on October 7, 2012, which are depicted in green (NSIDC 2012).

The most common digital change detection techniques are image algebra change detection and post-classification comparison change detection.

Image Algebra Change Detection

The changes between two dates of remotely sensed data may be quantified by using a simple image algebra band “ratioing” or differencing technique for corresponding bands in two dates. The resulting change image produced has positive and negative values (that can be transformed to positive values by adding or constant). A

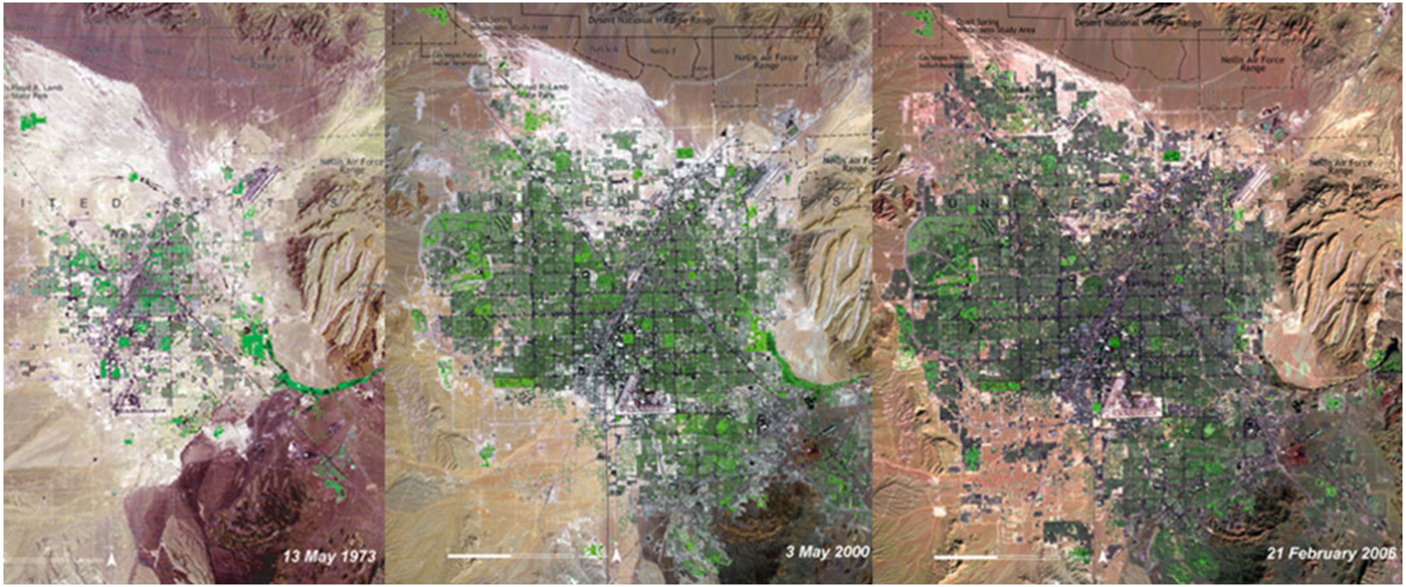


Fig. 17 Visual analysis change detection of Las Vegas, Nevada, area based on Landsat data (1973, 2000, 2006). Image adapted from http://earthobservatory.nasa.gov

Fig. 18 (**a**) and (**b**) Two QuickBird satellite images of a section of the Athens, Greece, stadium, showing the changes in the stadium: a) under construction (left) and b) completed (right) while preparing for 2004 Summer Olympics

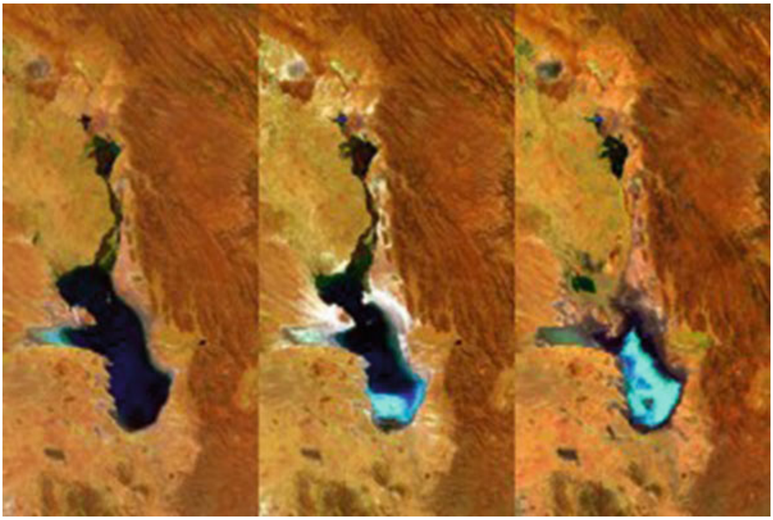


Fig. 19 (**a**), (**b**), and (**c**) Three 100-m resolution images recorded by the Proba-V minisatellite were acquired on April 27, 2014, July 20, 2015, and January 22, 2016, detailing the rapid evaporation of Lake Poopó in Bolivia. Once occupying a 3000-square km area, the 3-m deep lake was fully evaporated in December 2015. The lake is very sensitive to fluctuations in climate. (Image courtesy of ESA/Belspo, produced by VITO)

threshold for determining "change" and "no change" can be applied to the change image after examining the histogram (Elfishawy and Kesler 1991). Although this method does not provide "from–to" change information, it is simple and widely used.

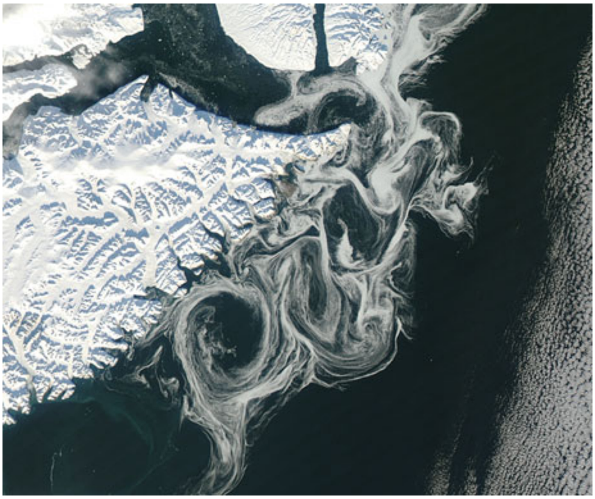


Fig. 20 Sea ice swirls off the coast of Greenland as acquired by MODIS Aqua. Image courtesy of Jeff Schmaltz, LANCE MODIS Rapid Response Team at NASA GSFC

Post-classification Comparison Change Detection

Post-classification comparison change detection methods are easy to use and require a complete classification of each individual date involved (Rutchey and Velcheck 1994; also see Jensen 2005). Unfortunately, any error present in the classification of individual dates will be incorporated in the change detection map. The number of classes increases geometrically. For example, for a six-category image for each date, 6 × 6 = 36 category of change is formed for the change map. The analyst can highlight very specific changes from a few desired categories to another few special interest categories and ignore the other land use/land cover categories in order to get an understanding of the change analysis. This is a commonly used technique and provides detailed “from–to” change information. Figure 21 shows the changes in land use and land cover that was produced using the post-classification method (Dobson et al. 1995).

In this figure, all changes are represented in various shades of gray. Figure 22 illustrates detailed post-classification change detection as compared to the TCC of the same area.

Data Integration and Spatial Modeling

The integration of remotely sensed data into available spatial analysis packages, such as ESRIs ArcGIS, and powerful statistical packages offers a tremendous opportunity for achieving historical and current information that is necessary for resource monitoring, mapping, and spatial modeling. The advantages provided by remote sensing include the ability to provide data collection and analysis over large and/or remote areas; a rapid, cost-effective method for monitoring, collecting, and analyzing data; and the ability to collect, model, and analyze current and archival

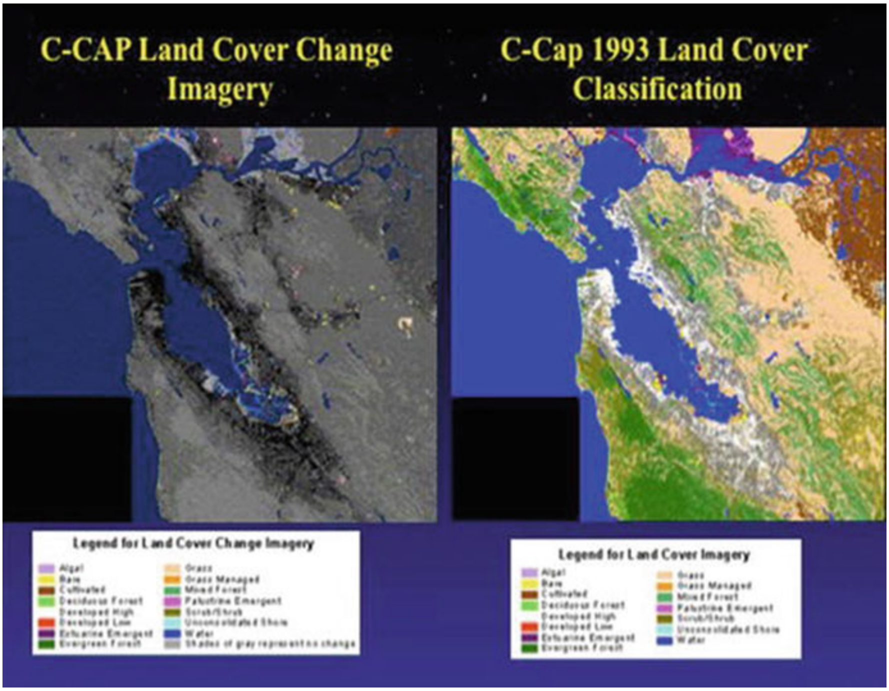


Fig. 21 Image of San Francisco Bay Area (courtesy of the coastal change analysis program of the US National Oceanic and Atmospheric Administration)

environmental data. Combining remotely sensed data with geographic information system (GIS) applications further provide the possibility of developing spatial analysis procedures that take advantage of temporal and spatial information on the structure of the landscape. Additionally, combining remotely sensed image data with geographic referenced data sets within a GIS allows users to take advantage of the extensive spatial database management system that lends itself easily to combining other geographic and tabular data of the respective area. Spatial and statistical models may be developed within a GIS or statistics program to determine likelihood of occurrence, probabilities, pattern and cluster analysis, and density and distribution functions that utilize classified image data as a base data source.

The true power of using remotely sensed data in a GIS is the fact that classified (processed) image data may serve as an input data source within the GIS. The GIS is able to use data in the form of layers to visualize patterns and trends that exist within the data, as shown in Fig. 23. The interactions and relationships that occur between these data layers can then be taken into account in a geospatial modeling context. Geospatial modeling results are then used by researchers and resource managers to solve problems and make appropriate decisions based on location-derived attributes.

Despite the potential analysis and modeling power provided by combination of remotely sensed image data with GIS and statistical packages, development of

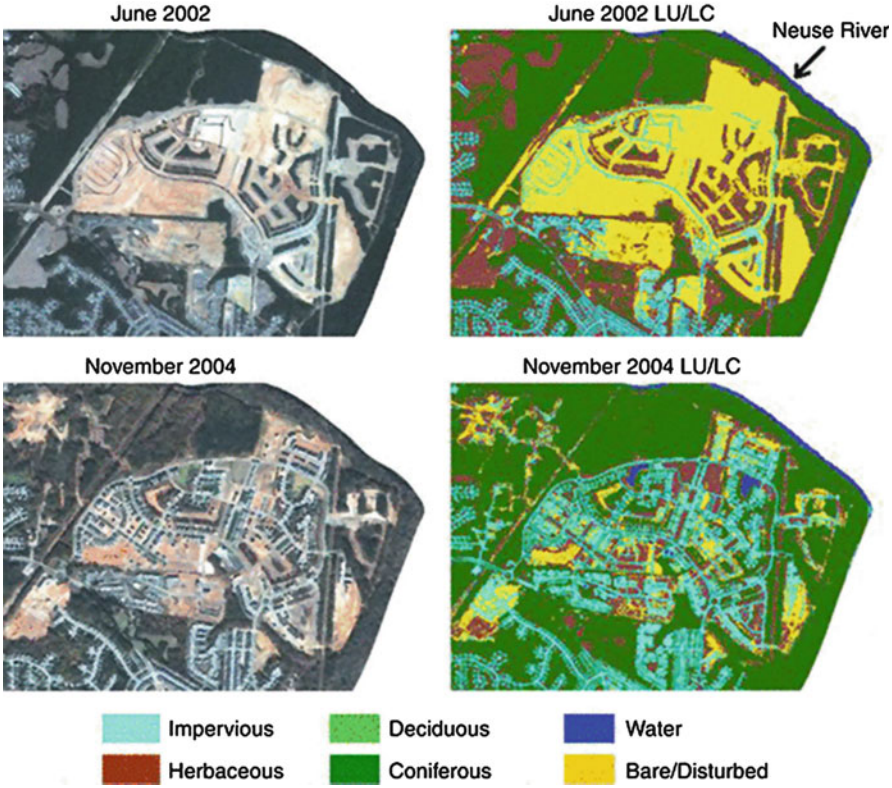


Fig. 22 Change detection using post-classification method (Graphic courtesy of Siamak Khorram)

integration protocols have only recently gained widespread attention. This may be due to differences in the data structures used to acquire and store data in the remote sensing and GIS environments. Digital imagery from spectral remote sensing detectors is stored as a raster data model. This data model is usually composed of a series of homogenous pixels organized in a row and column arrangement to represent the area of the Earth or targeted feature that the image data has been acquired over. The most popular data structures within a GIS are the vector file-based data models. Although many GIS allow for the use of raster data, problems may arise if the raster data is not available in a format that is readily usable for spatial analysis purposes. For example, unclassified (raw) raster data may require a considerable amount of processing outside of the GIS environment before it can be used for spatial modeling purposes.

Additionally, errors may arise from interpreting data accuracy of models developed from the combination of raster and vector data sources. This is largely due to the fact that accuracy assessments for remotely sensed data classifications are generated using the error matrix method. This method provides accuracy information on a global-image scale. Accuracy information developed within a GIS typically

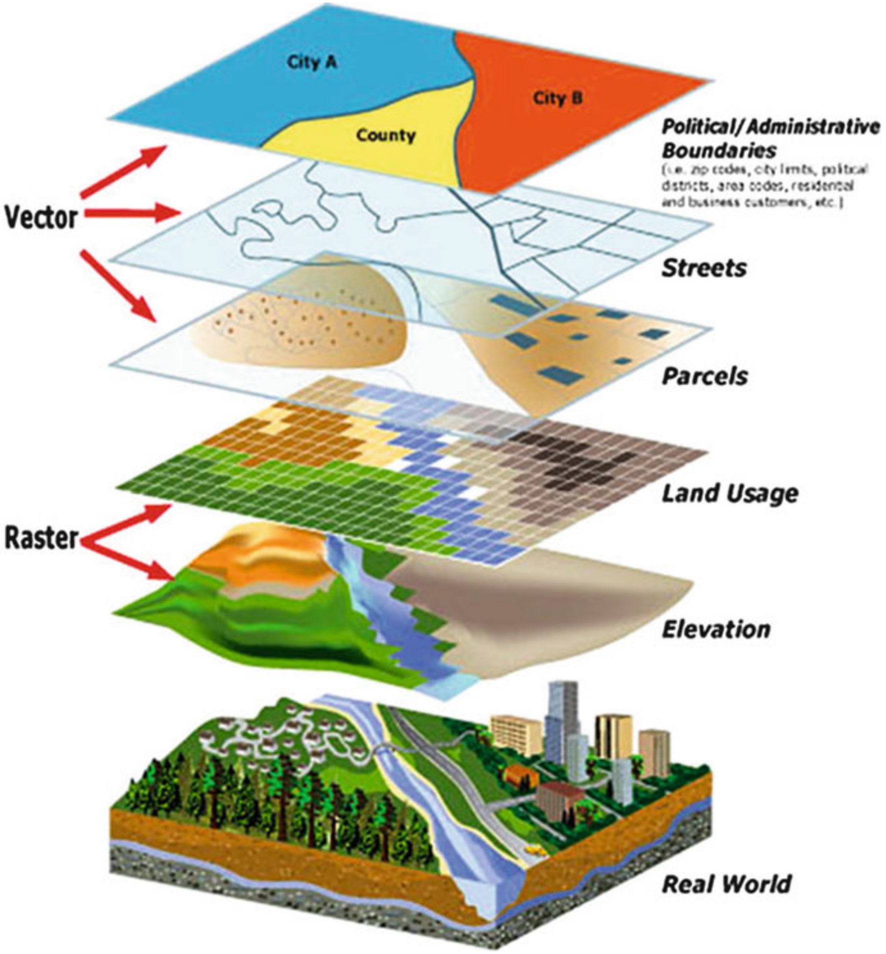


Fig. 23 A diagram displaying the integration of remote sensing and other data types in the form of geographic information systems (*GIS*) (Image courtesy of NOAA)

provides an analysis of spatial error on a local scale. Given this difference, care must be used to ensure accuracy assessment techniques are appropriate for all data types used in the analysis (Wang 1991; also see Khorram et al. 1999). Despite the limitations of using one package over another, be it an image or vector-based processing package, the integration, availability, and resolution advances of raster data have led to the development of new applications and new abilities to model data outputs. This trend is likely to continue as higher-quality data emerges, new uses for the data are discovered, new methods are developed to process them, new computing power is designed to analyze, and future missions and platforms are created to capture even higher-quality data.

Conclusion

The techniques described in this chapter are commonly used for processing airborne and satellite data for the use by a wide range of end users. Image processing techniques evolve as the higher-resolution data becomes more readily available from a variety of airborne and spaceborne platforms. With rapid technological advances in computer processing and cloud computing capabilities, removal of communication bottlenecks, improvements in screen technology, and the increase in high-volume data storage capacity, the future of remote sensing and geospatial technologies provide exciting opportunities for new applications in Earth observation and planetary exploration.

Cross-References

- ▶ Developments in Hyperspectral Sensing
- ▶ Electro-Optical and Hyperspectral Remote Sensing
- ▶ Geographic Information Systems and Geomatics
- ▶ Lidar Remote Sensing
- ▶ Operational Applications of Radar Images
- ▶ Remote Sensing Data Applications

References

S.G. Aaronoff, The minimum accuracy value as an index of classification accuracy. Photogramm. Eng. Remote Sens. **57**(5), 501–509 (1985)

J.R. Anderson, E. Hardy, J. Roach, R. Witmer, A land use and land cover classification system for use with remote sensing data, US Geological Survey Professional Paper 964, Washington, DC, 1976, p. 28ff

E.A. Blaisdell, *Harcourt Brace Javanovich* (Harcourt Brace Javanovich, New York, 1993), p. 653ff

T. Celik, Unsupervised change detection in satellite images using principal component analysis and k-means clustering. IEEE Geosci. Remote Sens. Lett. **6**(4), 772–776 (2009)

V. Cerny, Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. J. Optim. Theory Appl. **45**, 45–51 (1985) MathSciNet

S.B. Cho, J.H. Kim, Combining multiple neural networks by fuzzy integral for robust classification. IEEE Trans. Syst. Man Cybern. **25**(2), 380–384 (1995)

J.A. Cohen, A coefficient of agreement for nominal scales. Educ. Psychol. Meas. **20**, 37–46 (1960)

R.G. Congalton, K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices* (Lewis, Boca Raton, 1999). 137 p

R.G. Congalton, R.G. Oderwald, R.A. Mead, Assessing landsat classification accuracy using discrete multivariate statistical techniques. Photogramm. Eng. Remote Sens. **49**(12), 1671–1678 (1983)

L.M. Cowardin, V. Carter, F.C. Golet, E.T. LaRoe, *Classification of Wetlands and Deepwater Habitats of the United States* (U.S. Fish and Wildlife Service, Washington, DC, 1979), p. 103ff. FWS/OBS-79/31

X. Dai, S. Khorram, A new automated land cover change detection system for remotely-sensed imagery based on artificial neural networks, in *Proceedings of the IEEE/IGARSS 1997 International Geoscience and Remote Sensing Symposium*, Singapore, 1997a

X. Dai, S. Khorram, in *Proceedings of the IEEE/IGARSS 1997 International Geoscience and Remote Sensing Symposium*, Singapore, 1997b

X. Dai, S. Khorram, Data fusion using artificial neural networks: a case study on multitemporal change analysis. Comput. Environ. Urban Syst. **23**, 19–31 (1999)

A. Das, B.K. Chakrabarti, Quantum annealing and related optimization methods. Lect. Notes Phys. **679**, 239–257 (2005)

J. De Vincente, J. Lanchares, J. Hermida, Placement by thermodynamic simulated annealing. Phys. Lett. A **317**, 415–423 (2003)

J.R. Dobson, E.A. Bright, R.L. Ferguson, D.W. Field, L.L. Wood, K.D. Haddad, H. Iredale, J.R. Jensen, V. Klemas, R.J. Orth, J. P. Thomas, NOAA Coastal Change Analysis Program (C-CAP). Guidance for Regional Implementation. National Oceanic & Atmospheric Administration, Washington, DC, NMFS **123** (1995) p. 92ff

A.S. Elfishawy, S.B. Kesler, Adaptive algorithms for change detection in image sequence. Signal Process. **23**, 179–191 (1991)

G.M. Foody, Status of land cover classification accuracy assessment. Remote Sens. Environ. **80**, 185–201 (2002)

M.F. Goodchild, G.Q. Sun, S. Yang, Development and test of an error model for categorical data. Int. J. Geogr. Inf. Syst. **6**(2), 87–104 (1992)

A. Hagen, Fuzzy set approach to assessing similarity of categorical maps. Int. J. Geogr. Inf. Sci. **17**, 235–249 (2003)

D.B. Hester, Dissertation, North Carolina State University, 2008a

D.B. Hester, H.I. Cakir, S.A.C. Nelson, S. Khorram, Per-pixel classification of high spatial resolution satellite imagery for urban land cover mapping. Photogramm. Eng. Remote Sens. **74**, 463–471 (2008a)

D.B. Hester, S.A.C. Nelson, H.I. Cakir, S. Khorram, H. Cheshire, High resolution land cover change detection based on fuzzy uncertainty analysis and change reasoning. Int. J Remote Sens. **31**, 455–475 (2010)

R.M. Hord, *Digital Image Processing of Remotely-Sensed Data* (Academic, New York, 1982), p. 256

A.K. Jain, *Fundamentals of Digital Image Processing* (Prentice Hall, Englewood Cliffs, 1989), pp. 418–421 MATH

J.R. Jensen, *Introductory Digital Image Processing*, 3rd edn. (Pearson Prentice Hall, Upper Saddle River, 2005), p. 316

I. Kanellopoulos, G.G. Wilkinson, Strategies and best practice for neural network image classification. Int. J. Remote Sens. **18**, 711–725 (1997)

S. Khorram, C.F. van der Wiele, F.H. Koch, S.A.C. Nelson, M.D. Potts, *Principles of Applied Remote Sensing* (Springer, New York, 2016), p. 307. ISBN 978-3-319-22559-3

S. Khorram, F. Koch, C. van der Wiele, S.A.C. Nelson, *Remote Sensing, Book* (Springer, New York, 2012). doi:10.1007/9781-4614-3103-9. ISBN 978-1- 4614-3102-2

S. Khorram, H.M. Cheshire, K. Sidrellis, Z. Nagy. *Mapping and GIS Development of Land Use/Land Cover Categories for the Albemarle-Pamlico Drainage Basin* (NC Department of Environmental, Health, and Natural Resources, Raleigh, NC, USA Dept. No. 91–08, 1992), p. 55ff

S. Khorram, H. Cheshire, X. Dai, J. Morisette, Land cover inventory and change detection of coastal North Carolina using landsat thematic mapper data. ASPRS/ACSM Annu. Conv. Expos. **1**, 245–250 (1996) Remote Sensing and Photogrammetry

S. Khorram, G.S. Biging, N.R. Chrisman, D.R. Colby, R.G. Congalton, J.E. Dobson, R.L. Ferguson, M.F. Goodchild, J.R. Jensen, T.H. Mace, *Accuracy Assessment of Remote Sensing-Derived Change Detection*. American Society of Photogrammetry and Remote Sensing, Monograph (American Society of Photogrammetry and Remote Sensing, Bethesda, MD, 1999)

S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, Optimization by simulated annealing. Science **220**, 671–688 (1983) MathSciNet MATH

V.V. Klemas, J.E. Dobson, R.L. Ferguson, K.D. Haddad, A coastal land cover classification system for the NOAA coastWatch change analysis program. J. Coast. Res. **9**(3), 862–872 (1993)

T. Lillesand, R. Kiefer, J. Chipman, *Remote Sensing and Image Interpretation*, 6th edn. (Wiley, New York, 2008), p. 763

R.L. Lunetta, J.G. Lyons (eds.), *Geospatial Data Accuracy Assessment*. Report No. EPA/600/R-03/064 (US Environmental Protection Agency, Las Vegas, 2003), p 335

J.T. Morisette, S. Khorram, Exact Confidence Interval for Proportions, *Photogrammetric Engineering and Remote Sensing*, 66(7):875–880 (2003)

NOAA, Coastal Change Analysis Program (C-CAP), (NOAA Coastal Services Center, Charleston, 2004), http://www.csc.noaa-gov/crs/lca/ccap_program.html. Accessed 22 Dec 2015

NSIDC (National Snow & Ice Data Center), State of the Cryosphere: is the cryosphere sending signals about climate change? *Sea Ice* (2012), https://nsidc.org/cryosphere/sotc/sea_ice.html. Accessed 21 Jan 2016

Y. Nogami, Y. Jyo, M. Yoshioka, S. Omatu, Remote sensing data analysis by Kohonen feature map and competitive learning. IEEE SMC'97 **1**, 524–529 (1997)

D.P. Paine, J.D. Kiser, Chapter 23: mapping accuracy assessment, in *Aerial Photography and Image Interpretation*, 2nd edn. (Wiley, New York, 2003), pp. 465–480

M. Pal, P.M. Mather, An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sens. Environ. **86**, 554–565 (2003)

Polar Science Center: Arctic Sea Ice Volume Anomaly, Version 2, (University of Washington 2012). http://psc.apl.uw.edu/research/projects/arctic-sea-ice-volume-anomaly/. Accessed 21 Jan 2016

F. Qiu, J.R. Jensen, Opening of black box of neural networks for remote sensing image classification. Int. J. Remote Sens. **9**, 1749–1768 (2004)

D.E. Rumelhart, G.E. Hinton, R.J. Williams, *Parallel Distributed Processing* (MIT Press, Cambridge, MA, 1986)

K. Rutchey, L. Velcheck, Development of an everglades vegetation map using a SPOT image and global positioning system. Photogramm. Eng. Remote Sens. **60**(6), 767–775 (1994)

M.J. Sabins, Convergence and consistency of fuzzy C-means/ISODATA algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **9**, 661–668 (1987)

P.C. Shurr, Acceptance of the acceptance criteria for the simulated annealing algorithm. Math. Oper. Res **22**(2), 266ff (1997)

S.V. Stehman, Statistical rigor and practical utility in thematic map accuracy assessment. Photogramm. Eng. Remote Sens. **67**, 727–734 (2001)

M. Story, R.G. Congalton, Accuracy assessment: a user's perspective. Photogramm. Eng. Remote Sens. **52**(3), 397–399 (1986)

J.T. Tou, R.C. Gonzalez, *Pattern Recognition Principles* (Addison-Wesley, Readings, 1977), p. 377

J.L. Van Genderen, B.F. Lock, Testing land use map accuracy. Photogramm. Eng. Remote Sens. **43** (9), 1135–1137 (1977)

USGS, USGS National Land Cover Data (EROS Data Center, Sioux Falls, 2004), http://landcover.usgs.gov/prodescription.html. Accessed 20 Dec 2015

F. Wang, Integrating GIS and remote sensing image analysis systems by unifying knowledge representation scheme. IEEE Trans. Geosci. Remote Sens. **29**, 656–664 (1991)

D.M. Winker M.A. Vaughan, A.H. Omar, Y. Hu, K.A. Powell, Z. Liu, W.H. Hunt, and S.A. Young, Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms. J. Atmos. Oceanic Technol **26**, 2310–2323 (2009)

R. Xu, D. Wunsch, Survey of clustering algorithms. IEEE Trans. Neural Netw. **16**, 32f (2005)

C. Yang, P. Chung, Knowledge-based automatic change detection positioning system for complex heterogeneous environments. J. Intell. Robotic Syst **33**, 85–98 (2002). MATH

Remote Sensing Data Applications

Haruhisa Shimoda

Contents

H. Shimoda (✉)
Research and Information Center, Tokai University, Shibuya-ku, Tokyo, Japan
e-mail: smd@keyaki.cc.u-tokai.ac.jp

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_47

Abstract

Application areas of remote sensing are very wide. They can be divided into two areas. One is applications in the Earth environmental monitoring and process studies of the Earth system, and another is operational applications. The former can be divided into atmosphere, ocean, land, cryosphere, and their interactions. In this chapter, temperature, water vapor, aerosols and clouds, atmospheric constituents, greenhouse gases, sea surface temperature, sea surface salinity, sea surface wind, ocean color, sea surface height, topography, land cover, soil moisture, carbon cycle, sea ice, snow, and glaciers are described. The latter has wide variety. This chapter cannot cover all the operational application areas. Among them, NWP and weather forecasting, fisheries, disasters such as biomass burnings, floods, ship navigations, and agriculture are described. In addition to these application areas, some basic processings for applications are also described. These processings include radiative transfer and inversion problem, geometric and radiometric corrections, and classification algorithms.

Keywords

Radiative transfer • Temperature • Water vapor • Aerosol • Cloud • Atmospheric constituents • Greenhouse gases • Sea surface temperature • Sea surface salinity • Sea surface wind • Ocean color • Sea surface height • Topography • Geometric correction • Radiometric correction • Land cover • Soil moisture • Carbon cycle • Sea ice • Snow • Weather forecasting

Introduction

Application areas of remote sensing are very wide. The application areas at the start of remote sensing in sixties were mostly operational applications, i.e., land cover/ use, mineral explorations, agriculture, forest industry, fisheries, and disasters. However, recent development of global change has created new application fields, i.e., Earth environmental monitoring. In order to understand the full degree of worldwide environmental change key global geophysical parameters should be measured for a long time. These parameters include elements such as atmospheric and environmental constituents like greenhouse gases and ozone, large scale forest decrease and desertification. It is almost impossible to monitor these global phenomena using conventional in situ measurements. It is critical to use remote sensing for these purposes.

In this chapter, measurement and monitoring techniques for each sphere, i.e., atmosphere, ocean, land, and cryosphere are described. After these descriptions, some of the operational fields, like weather forecast, fisheries, disaster, etc., are described. In addition to these application areas, some basic processings for applications are also described. These processings include radiative transfer and inversion problem, geometric and radiometric corrections, and classification algorithms.

Atmospheric Applications

Radiative Transfer and Inversion Problem

In remote sensing, sensors on satellites receive electromagnetic waves from the Earth. However, these electromagnetic waves are reflected or emitted from the ground or from the atmosphere and absorbed or scattered by the atmosphere and finally arrive at the sensor. So, the received electromagnetic waves by the sensor can be described using radiative transfer codes. Radiative transfer equation when light of spectral radiance I_λ propagates ds in a medium can be written as follows:

$$\frac{dI_\lambda}{K_\lambda \rho ds} = -I_\lambda + S_\lambda$$

Here,

I_λ: Spectral radiance of wavelength λ
K_λ: Mass attenuation coefficient
ρ : Density of medium in which wave propagates
S_λ: Source function of radiation

and

$$K_\lambda = \chi_\lambda + \sigma_\lambda$$

χ_λ: Mass absorption coefficient
σ_λ: Mass scattering coefficient

$$S_\lambda \equiv j_\lambda / K_\lambda$$

where

j_λ: Mass emission coefficient

Unfortunately, this equation cannot be solved analytically. So, many kinds of approximation codes are developed and used. Most popular codes are MODTRAN (Berk et al. 1998), 6S (Kotchenova et al. 2006), RSTAR (Nakajima et al. 2004), and

LBLRTM (Clough et al. 2005). LBLRTM is used for thermal infrared region, while other codes are mainly used in visible and near-infrared region. ARTS (Buehler et al. 2005) is sometimes used for the microwave region.

In the visible and near-infrared region, main atmospheric effects are scattering by atmospheric molecules and aerosols. The former can be modeled by Rayleigh scattering (Rayleigh 1871) and the latter can be modeled by Mie scattering (Mie 1908a). In order to calculate atmospheric attenuation of light by these scatterings, characteristics of atmospheric molecules and aerosols are necessary. These characteristics are usually given from MODTRAN. On the other hand, molecular absorption characteristics are necessary to calculate in the infrared region. These characteristics are given by databases. Most popular databases are HITRAN (Rothman et al. 2009) and GEISA (Jacquinet-Husson et al. 2009).

In order to retrieve geophysical parameters from data obtained by satellite born sensors, it is necessary to invert the radiative transfer equation. Usually, this process is an ill-posed problem and many kinds of inversion algorithms are proposed. The most popular algorithms are MAP (maximum a posteriori probability) (Rodgers 2000) and artificial neural networks. Here, a brief description of MAP algorithm is described.

$$\hat{\boldsymbol{x}} = \boldsymbol{x_a} + \left(\boldsymbol{K}^T \boldsymbol{S}_\varepsilon^{-1} \boldsymbol{K} + \boldsymbol{S}_a^{-1}\right)^{-1} \boldsymbol{K}^T \boldsymbol{S}_\varepsilon^{-1} (\boldsymbol{y} - \boldsymbol{K} x_a)$$

Here,

$\hat{\boldsymbol{x}}$: Retrieved geophysical parameter (vector)
$\boldsymbol{x}^a$: A priori parameter (vector)
$\boldsymbol{y}$ **:** Observed data (vector)
S_ε: Variance-covariance matrix of observation noise
S_a: Variance-covariance matrix of the parameter
K: Jacobian matrix

$$K = \partial F(x)/\partial x$$

where

$F(x)$: Observed radiance spectra

Temperature and Water Vapor

Atmospheric temperature and water vapor are the most important geophysical parameters for weather forecasting. There are two ways to measure these variables from space. One is to use optical sensors and another is to use microwave sensors. For the optical sensor, high spectral resolution IR optical sensors are usually used. Two kinds of sensors are now used. One is a grating spectrometer and another is a Fourier transform IR spectrometer. AIRS on EOS Aqua is an example of the former instrument, while IASI on METOP is an example of the latter instrument. Both instruments can measure temperature with 0.1 K accuracy and water vapor in 10 % accuracy with 1 km vertical resolution. Another kind of optical sensor which can measure column density of water vapor over land uses near-infrared water vapor absorption lines. The disadvantage of optical sensors is that it cannot measure parameters under clouds. Figures 1 and 2 show global atmospheric temperature and water vapor measured by AIRS on Aqua.

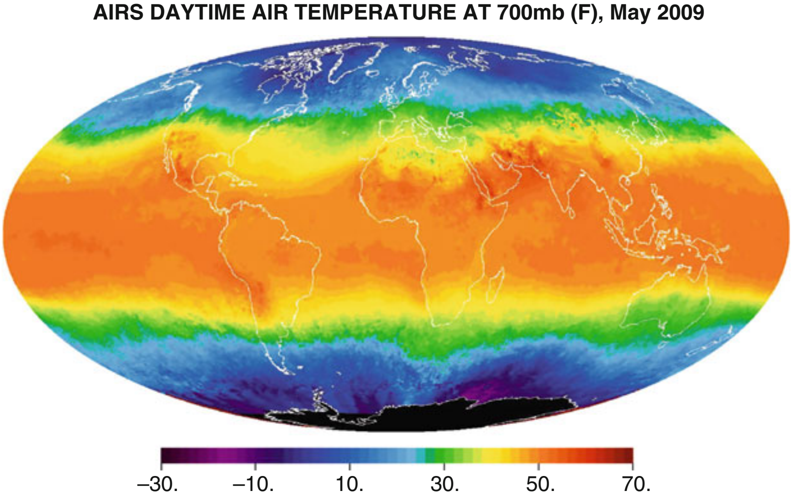


Fig. 1 Global atmospheric temperature at 700 hPa measured by AIRS on Aqua (2009)

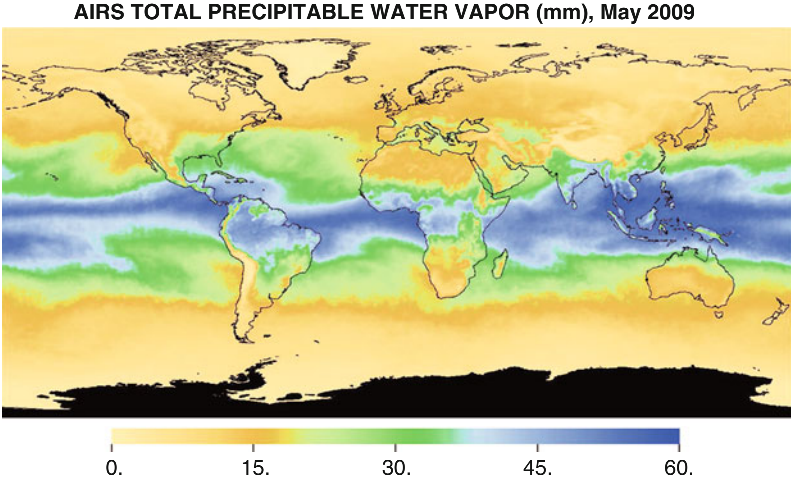


Fig. 2 Global atmospheric water vapor (total precipitable water) measured by AIRS on Aqua (2009)

Vertical profiles of temperature and water vapor also can be retrieved by microwave sensors like AMSU on METOP. Temperature is measured mainly using 50 GHz band, while water vapor is measured using 183 GHz band. The accuracy of microwave sounders is not so good as optical sounders. The vertical resolution depends on the number of channels in each absorption band. The advantage of microwave sounders is that it can measure over clouds. Also, microwave instruments can measure column water vapor using 23 and 31 GHz bands.

Another way of retrieving temperature and water vapor vertical distribution is to use GPS occultation technologies. GPS signals received at the satellites through atmosphere are refracted by atmosphere. The extent of this refraction depends on temperature and water vapor concentrations of the atmosphere. Hence, temperature and water vapor concentrations can be inferred from the extent of the refraction (Melbourne et al. 1994, Kursinski et al. 1997). COSMIC system on FORMOSAT is now composed of six satellites carrying GPS occultation instruments, and several other satellites also carry GPS occultation instruments. From these instruments several thousand measurements are done in one day now.

Aerosols and Clouds

Aerosols and clouds are one of the most important parameters to measure from space. In the last IPCC report, aerosols and clouds are still the most uncertain radiative forcings in the climate models. There are two ways to measure aerosols and clouds from space. One is to use passive optical sensors, and other is to use active sensors, i.e., lidar and radar. In the passive system, aerosols are usually measured using visible wavelengths. Over the ocean, it is rather easy. In most of methods, near-infrared or short wave infrared wavelength is used to acquire aerosol free reflection from the ocean, and this information is used to infer visible wavelength Mie scattering properties. In most cases, retrieved geophysical parameters are optical thickness and Ångstrom exponent. The former is the total quantity of aerosols, and the latter corresponds to aerosol particle radius. Figure 3 shows a typical example of this type of aerosol retrieval obtained from visible and near-infrared channels of GLI on ADEOS2. Other important parameters are kinds of aerosols, refractive indices, and height of aerosols.

Over land, the retrievals are rather difficult. Ocean is very homogeneous and dark, but land is very inhomogeneous and bright. Two kinds of aerosol retrievals over land are tried using ultraviolet wavelength and polarization. Ultraviolet absorbing aerosols such as dust and soot can be retrieved using ultraviolet wavelengths. Figure 4 shows aerosol distribution retrieved from TOMS using ultraviolet, and Fig. 5 shows aerosol retrievals over land using POLDER on ADEOS using polarizations.

Active sensors, i.e., lidar and radar can retrieve vertical distribution of aerosols and clouds. However, it is very difficult to retrieve horizontal distributions using these sensors. For aerosol retrieval, Mie scattering lidar is used, while for cloud retrieval, cloud profiling radar (CPR) is used. CPR is a very high frequency radar using around 94 GHz frequency. Figure 6 shows an example of aerosol vertical

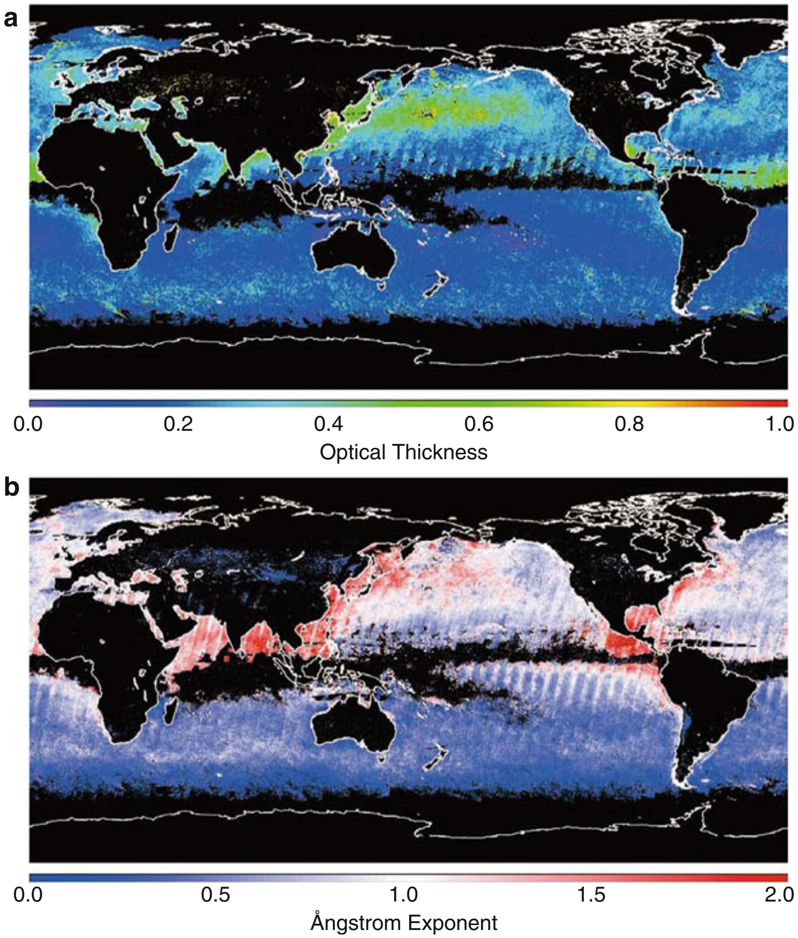


Fig. 3 Global aerosol distribution over ocean retrieved from GLI on ADEOS2. (**a**) Aerosol optical thickness. (**b**) Aerosol Angstrom exponent (Nakajima et al. 2009)

distribution retrieved from Mie lidar CALIOP on CALIPSO, and Fig. 7 shows an example of cloud vertical distribution retrieved from CPR on CloudSat.

Atmospheric Constituents

Many kinds of atmospheric constituents can be measured from space. Figure 8 shows the atmospheric absorption spectrum of major constituents. From this figure, wide range of electromagnetic wavelength can be used for the measurements. Usually, ultraviolet, visible and near-infrared, thermal infrared, microwave, millimeter, and submillimeter regions are used for measuring atmospheric constituents.

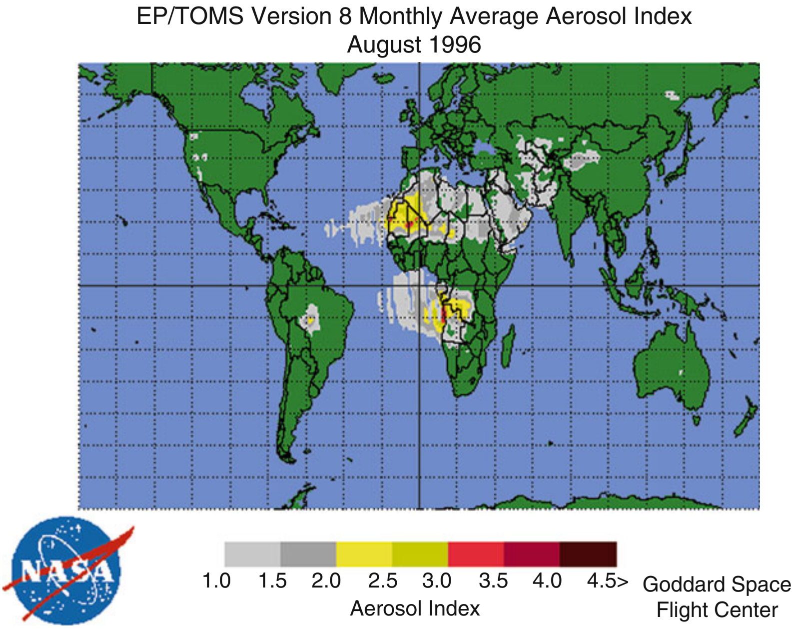


Fig. 4 Aerosols over land from TOMS (EP/TOMS Version 8 Monthly Average Aerosol Index 1996) (Courtesy of NASA)

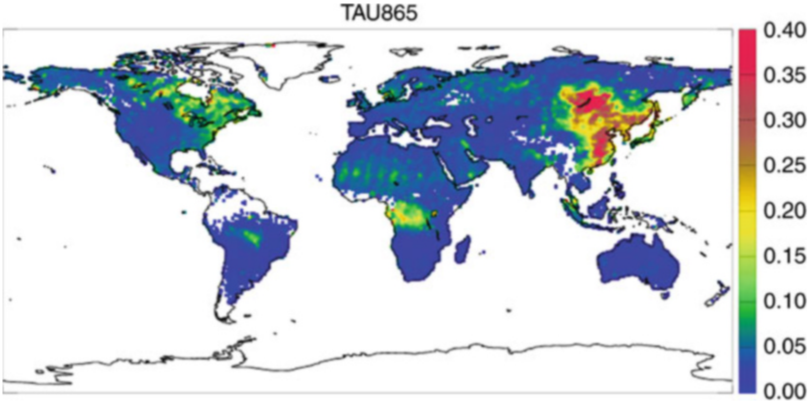


Fig. 5 Aerosols over land and ocean using POLDER (AEROSOLS RESULTS OVER LAND 1997) (Courtesy of CNES)

There are two ways of measurements. One is to observe the Earth in the vertical direction (nadir observation) and another is to observe in the limb direction (limb observation) as shown in Fig. 9. The latter can be divided into two methods depending on the light source. One is to measure the emissions from the atmosphere and another is to observe absorption by the atmosphere from light sources, e.g., sun,

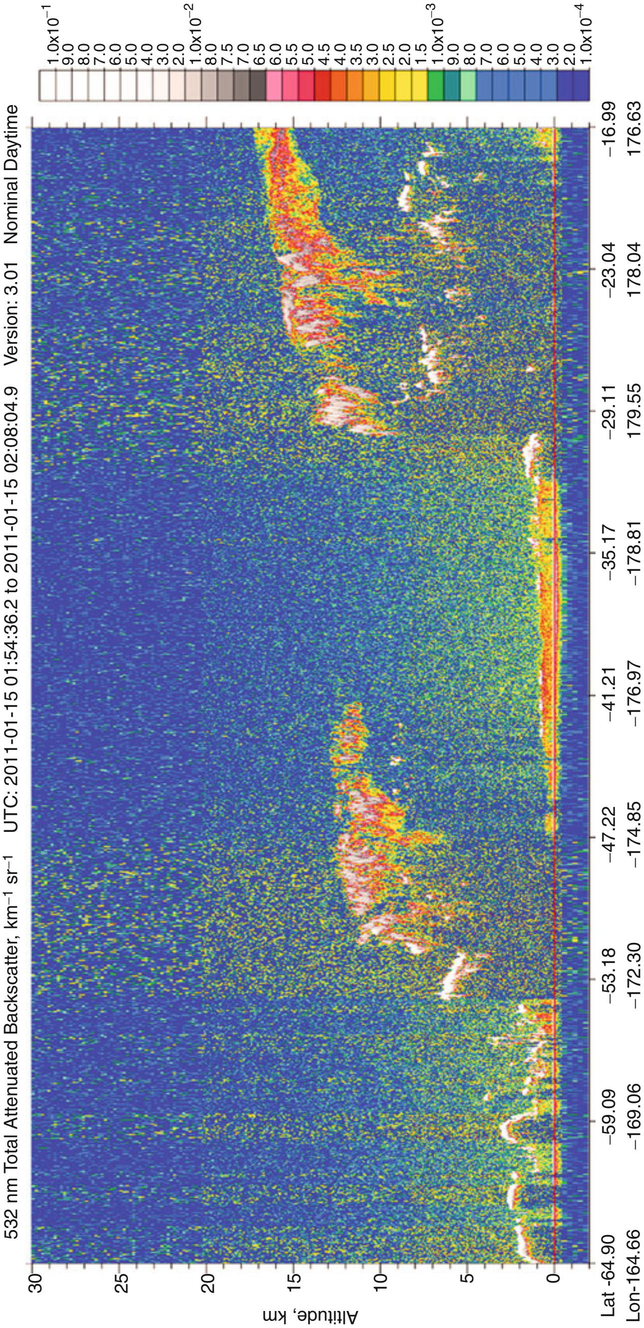


Fig. 6 Aerosol vertical distribution from CALIOP (LIDAR LEVEL1 BROWSE IMAGES 2011) (Courtesy of NASA)

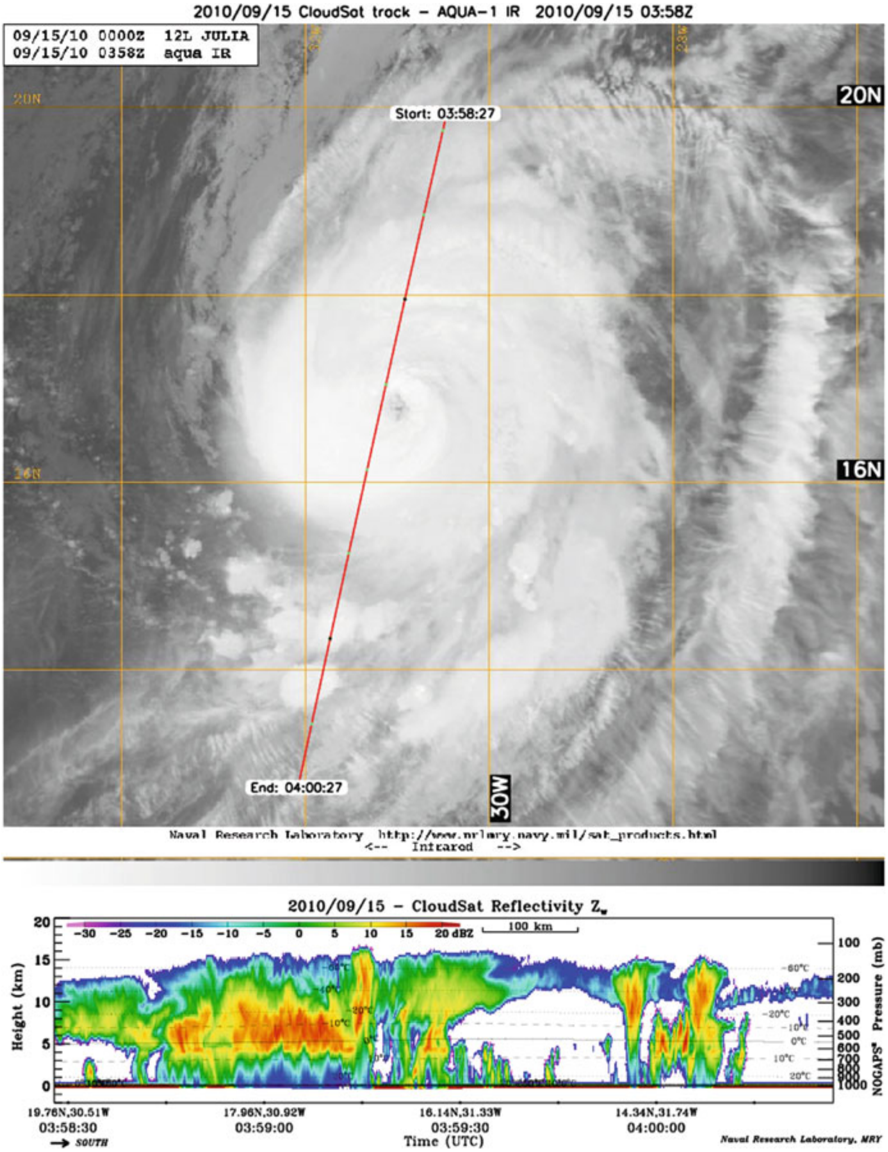


Fig. 7 Cloud vertical distribution from CloudSat (Cloud vertical distribution from CloudSat 2010) (Courtesy of Colorado State University)

moon, and stars. The latter method is called an occultation measurement. Each observation method has advantages and disadvantages. Nadir observation is good for observing horizontal distributions, while its vertical resolution is not so good. Limb observation has very high vertical resolution and also very sensitive, but its horizontal resolution is bad, and it is very difficult to observe middle and lower

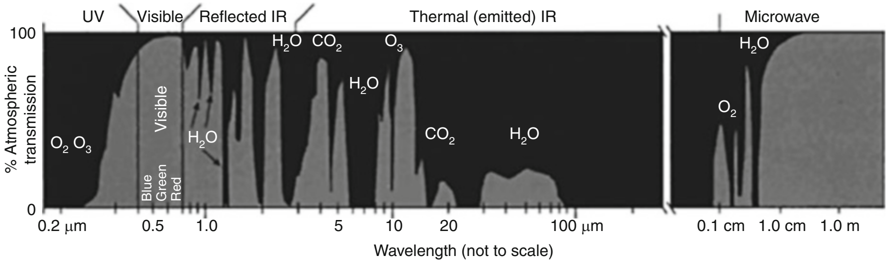


Fig. 8 Atmospheric absorption spectra by major atmospheric constituents (Earth Observatory, Absorption Bands and Atmospheric Windows 1999) (Courtesy of NASA)

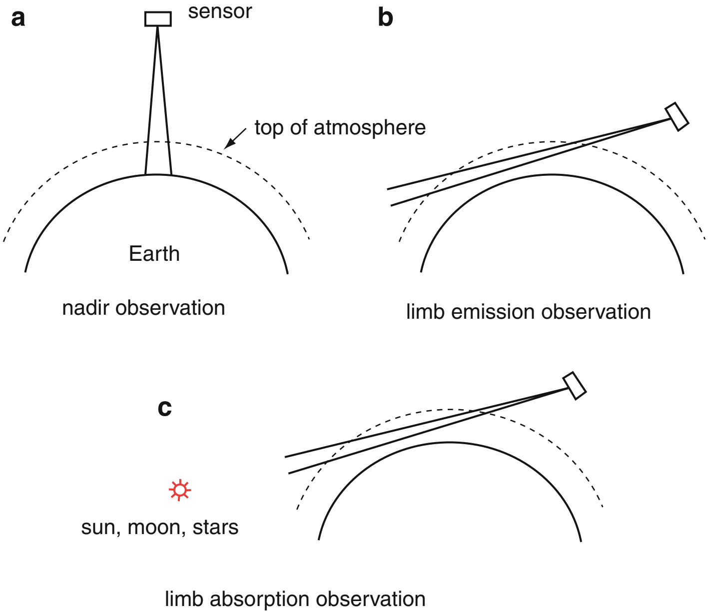


Fig. 9 Observation method of atmosphere

troposphere. Solar occultation measurements are very stable and have high signal to noise ratio, but the observation areas are very limited when used from sun synchronous orbit satellites. Limb emission measurements can measure in day and night and can cover very wide areas, but its sensitivity is limited by the instrument sensitivity.

In the UV region, the main target is ozone, sulfur dioxide, and NO_2. Figure 10 shows the total ozone measured by EP/TOMS. The Antarctic ozone hole can be clearly seen in this image. Figure 11 shows the total NO_2 measured by SCIAMACHY on ENVISAT. In the visible and near-infrared region, there are very few steep absorption lines. There are some lines of ozone and detailed absorption

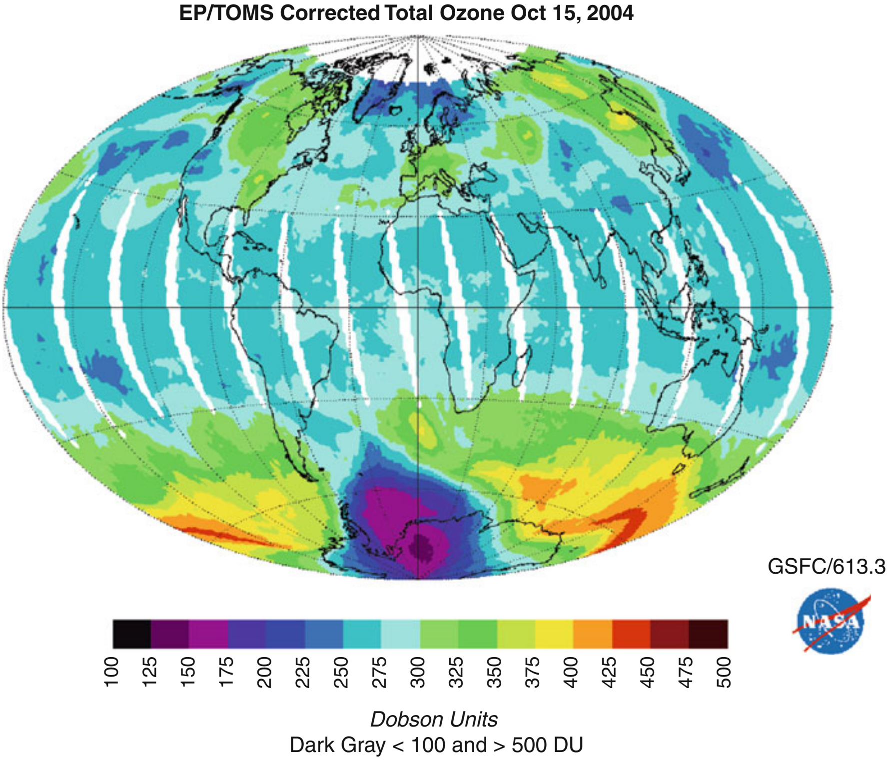


Fig. 10 Total ozone measured by TOMS on EP (Earth Probe TOMS Data & Images 1997) (Courtesy of NASA)

lines of oxygen. Oxygen lines are used to retrieve pressures. Figure 12 shows the vertical distribution of ozone over Antarctica obtained from solar occultation sensor ILAS on ADEOS.

In the infrared region, there are many absorption lines from many molecules. For the nadir measurement, major molecules which can be measured are H_2O, CO, N_2O, O_3, and CH_4. Figure 13 shows the distribution of CO measured by SCIAMACHY on ENVISAT.

Limb emission measurement can measure many kinds of atmospheric molecules. For instance, MIPAS on ENVISAT can measure following molecules: C_2H_2, C_2H_6, CH_4, ClO, $ClONO_2$, CO, F11, F12, F22, H_2O, H_2CO, HCN, HCOOH, HNO_3, HNO_4, HOCl, N_2O_5, N_2O, NO_2, NO, and O_3. Figure 14 shows global distributions of HCN and C_2H_6 showing biomass burning observed by MIPAS on ENVISAT.

Global HCN (left) and C_2H_6 (right) distributions at 200 hPa (10.5–12.6 km) have been measured by MIPAS in September 2003. White areas are data gaps due to cloud contamination (or insensitive values in case of C_2H_6 south of 60°S). Red solid lines show the monthly averaged tropopause intersection from the NCEP reanalysis.

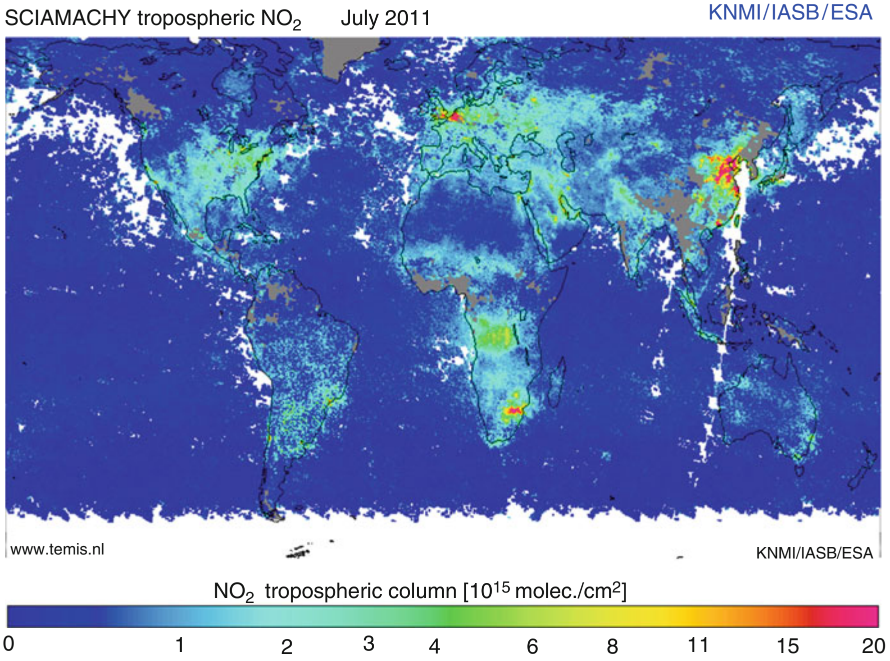


Fig. 11 Total NO_2 measured by SCIAMACHY on ENVISAT (Total NO2 measured by SCIAMACHY on ENVISAT 2011) (Courtesy of ESA)

Strongly enhanced values of HCN between South America, Africa, and Australia reflect the southern hemispheric biomass burning plume. Enhanced C_2H_6, an indicator for both biomass burning and industrial/urban pollution, is also visible in this region, but additionally west of Peru and in the northern tropics and subtropics. Trajectory calculations (not shown) hint towards pollution sources in Northern South America. For further details see: http://www.atmos-chem-phys.net/9/9619/2009/acp-9-9619-2009.html

Greenhouse Gases

Greenhouse gases, especially distributions of CO_2 and CH_4 and their fluxes are very important parameters to understand the carbon cycle. Atmospheric greenhouse gases, especially carbon dioxide is very difficult to measure. In order to retrieve useful CO_2 concentrations, required accuracy is at least 1 % and 1 ppm accuracy is desirable. On the contrary, CO_2 measurement is very sensitive to aerosols, clouds, and surface pressures. This is the reason that there were no spaceborne sensors to measure CO_2 until recently. There are several CO_2 absorption lines in short wave infrared and thermal infrared. In the short wave infrared region, absorption lines are in 1.6 μm and 2.0 μm region. In the thermal infrared region, CO_2 absorption lines are

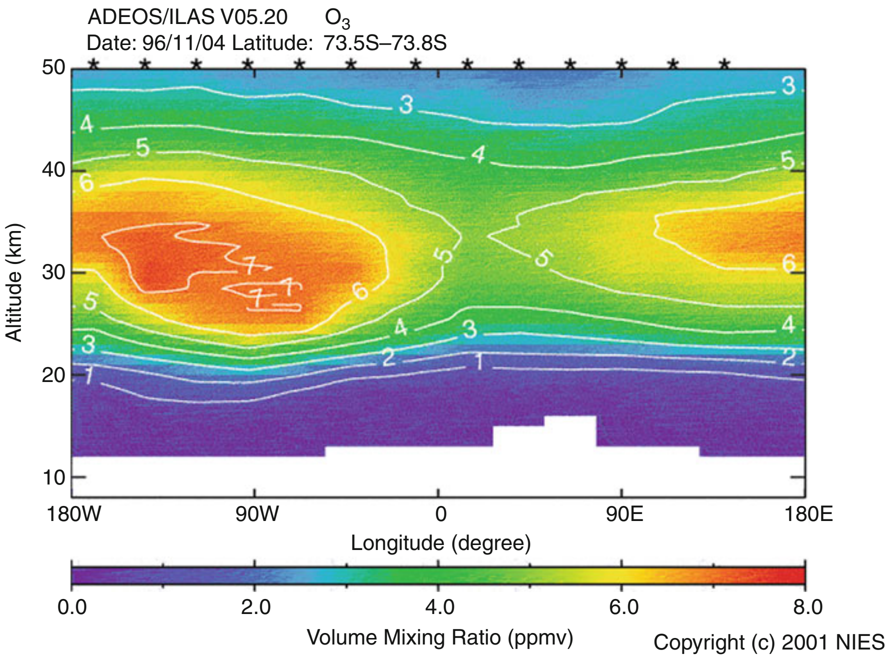


Fig. 12 Vertical distribution of ozone over Antarctica obtained from solar occultation sensor ILAS on ADEOS (ADEOS EarthView 1998) (Courtesy of NIES)

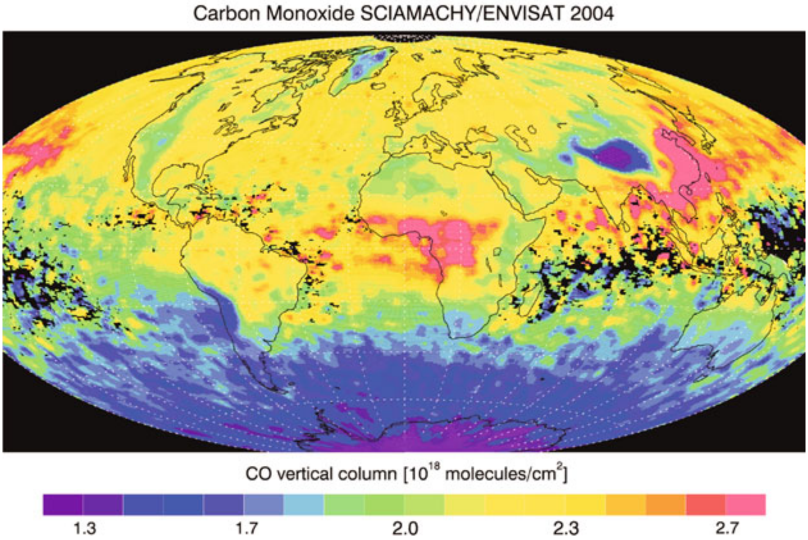


Fig. 13 Global CO distribution measured by SCIAMACHY on ENVISAT (Carbon Monoxide SCIAMACHY/ENVISAT 2004) (Courtesy of University Bremen)

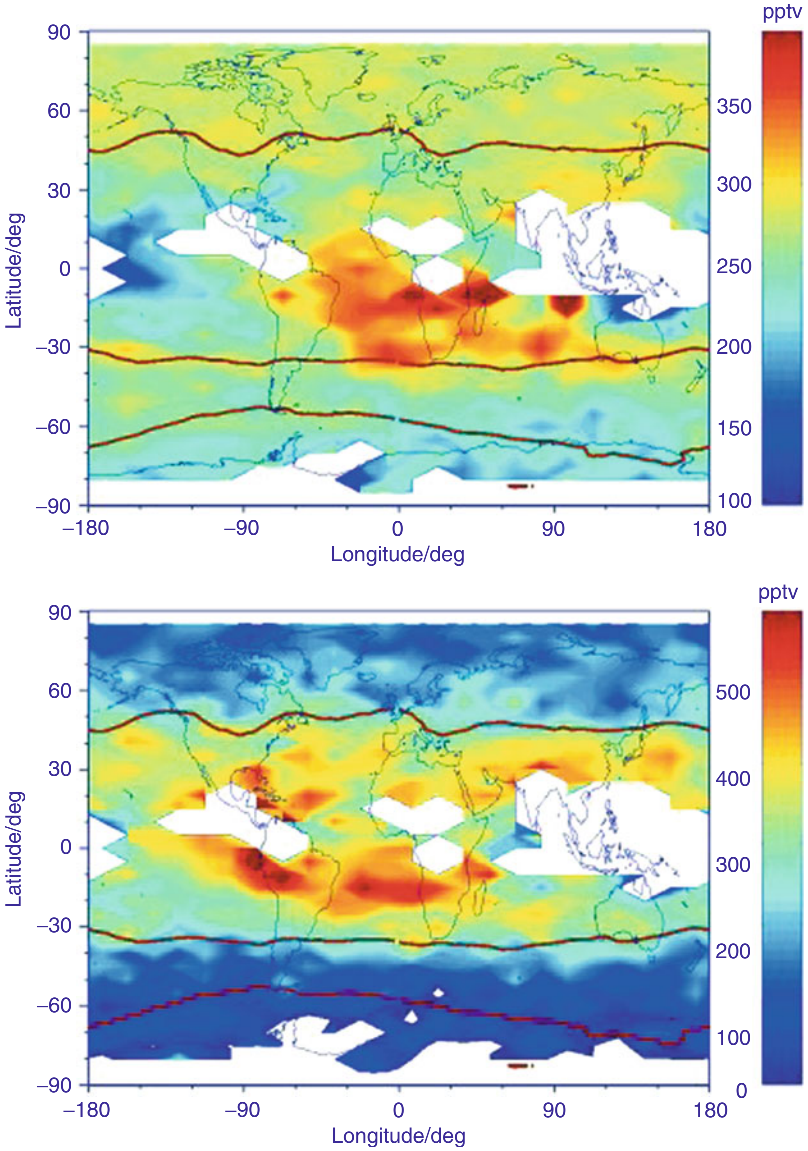


Fig. 14 Global distributions of HCN and C_2H_6 showing biomass burning (Global distributions of HCN and C2H6 showing biomass burning 2003) (Courtesy of IMK)

in 11 µm region and in 15 µm region. Figure 15 shows CO_2 and CH_4 absorption lines in short wave and thermal infrared region. As for the vertical sensitivity, short wave infrared region has almost flat sensitivity in the troposphere, while thermal infrared region has high sensitivity in mid-troposphere but very low sensitivity in lower

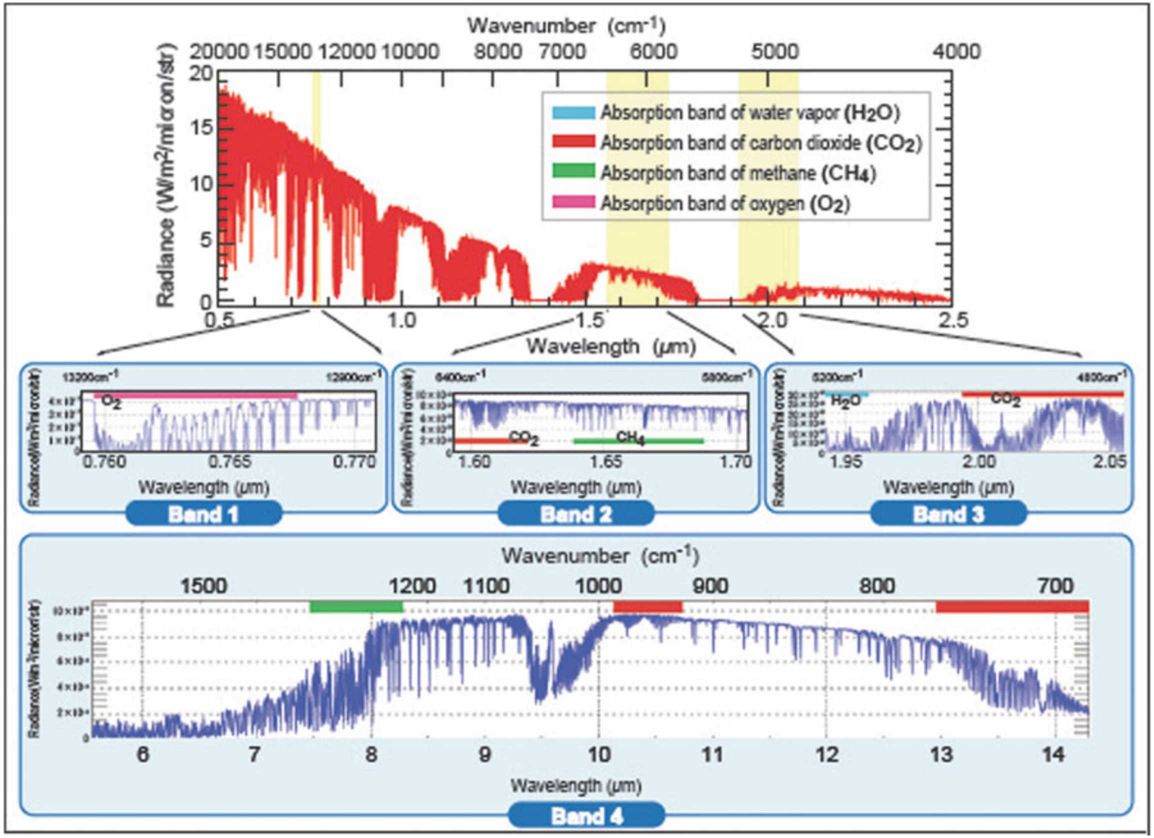


Fig. 15 CO_2 and CH_4 absorption lines in short wave and thermal infrared region (GOSAT Project 2009) (Courtesy of NIES)

troposphere. From short wave infrared spectra, only total column density can be retrieved, while in the thermal infrared region, vertical profile from around 3000 m and above can be retrieved. Figures 16 and 17 show CO_2 and CH_4 distributions retrieved from short wave infrared bands of TANSO-FTS on GOSAT. TANSO-FTS is a Fourier transform spectrometer with short wave infrared bands and thermal

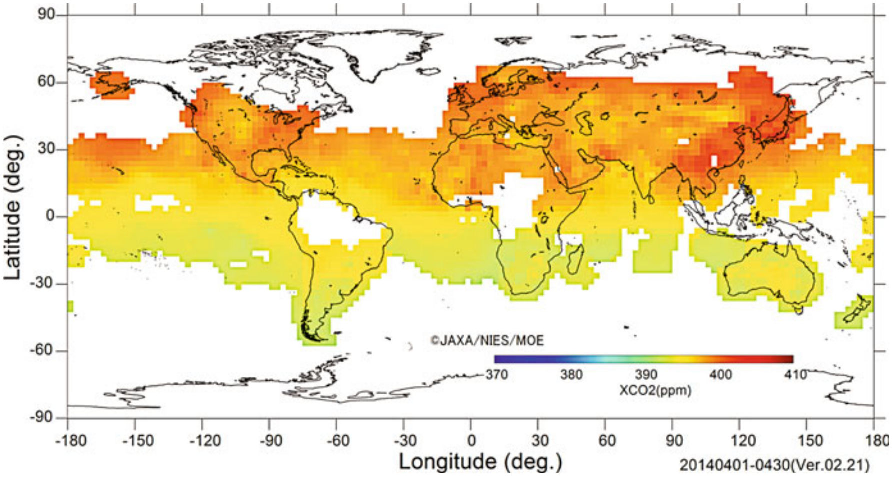


Fig. 16 CO_2 distribution from TANSO on GOSAT (XCO_2 distribution level 3 2014) (Courtesy of NIES)

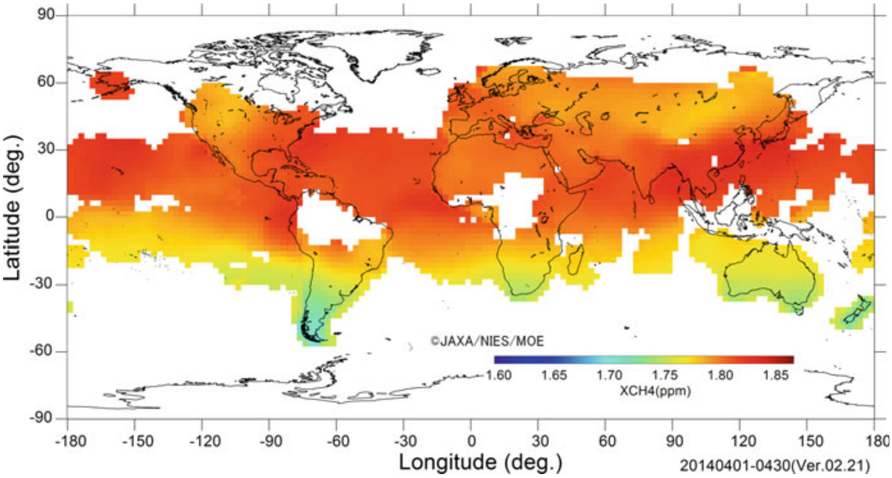


Fig. 17 CH_4 distribution from TANSO on GOSAT (XCH_4 distribution level 3 2014) (Courtesy of NIES)

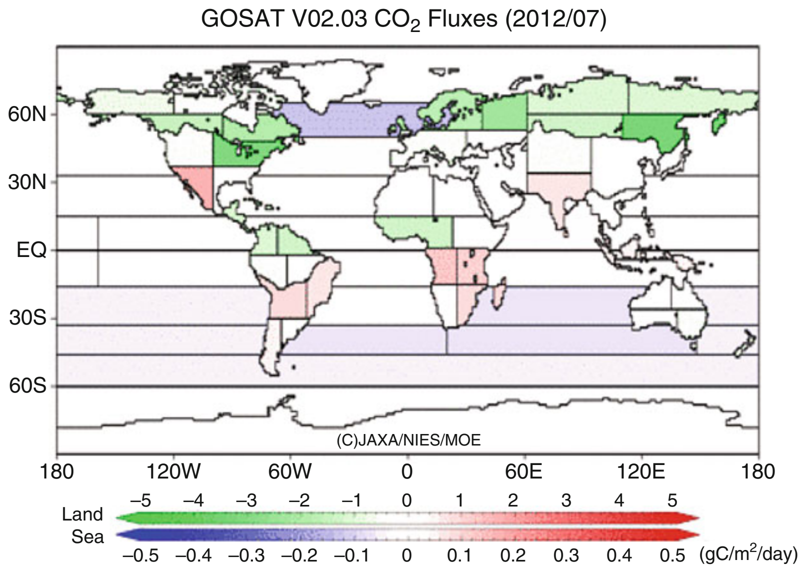


Fig. 18 CO_2 flux obtained from inverse model using TANSO-FTS CO_2 concentration data (CO_2 flux 2012) (Courtesy of NIES)

infrared band. Figure 18 shows the CO_2 flux obtained from the inverse model using TANSO-FTS CO_2 concentration data.

Precipitation

Precipitation not only largely affects the life of humankind but also dominates global energy and water cycle. Precipitation can be measured by microwave radiometers as well as by microwave radars. Microwave radiometers have a long history of observation, but the accuracy on land is not sufficient. TRMM was launched on 1997, and it carried both microwave radiometer (MSI) and precipitation radar (PR). Precipitation radar can measure three-dimensional structure of precipitation, and its accuracy over land is also quite good. In both cases, precipitation retrievals depend on many assumptions. With 17 years of continuous measurements, global precipitation retrievals both over land and ocean have rather good accuracy now. In 2014, GPM core satellite which is a follow on of TRMM was launched. It carries microwave radiometer (GMI) and dual frequency precipitation radar (DPR). PR was a Ku-band radar, while DPR is Ku and Ka band radar. With this capability, precipitation as well as droplet size distribution can be retrieved, and also Ka band radar is sensitive to solid precipitation. Figure 19 shows an example of three-dimensional precipitation retrieved from DPR.

There are many microwave imagers and sounders in the world. Using these microwave radiometers based upon precipitation radar statistics and models, real-

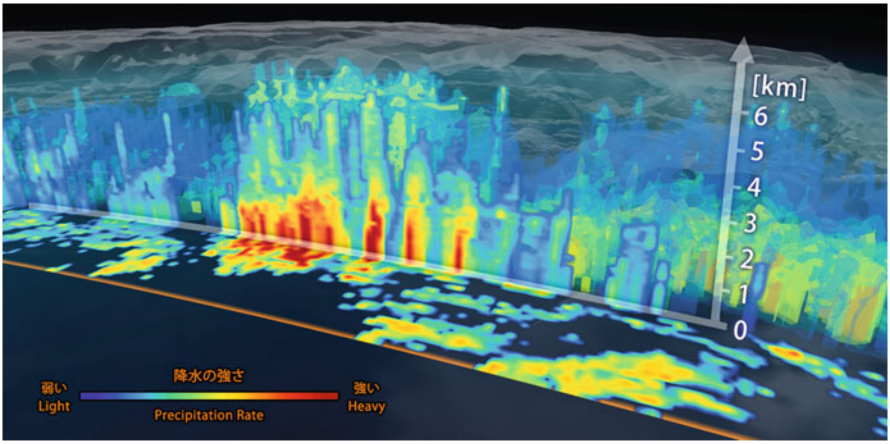


Fig. 19 Three-dimensional precipitation retrieved from DPR (3D view 2014) (Courtesy of JAXA/NASA)

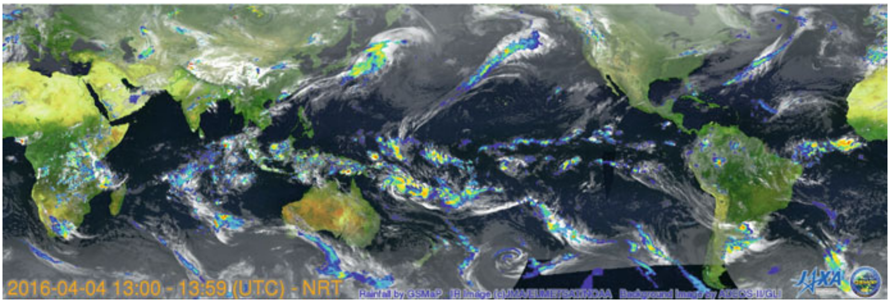


Fig. 20 An example of real-time global precipitation map (GSMaP) (JAXA GLOBAL RAINFALL WATCH 2016) (Courtesy of JAXA)

time global precipitation map can be generated. Figure 20 shows an example of global precipitation map generated by JAXA.

Oceanic Applications

Sea Surface Temperature

Sea surface temperature (SST) is very important parameter to understand the surface ocean circulations, as well as to monitor the global warming trend. It is also very useful for efficient fisheries. SST is the first geophysical parameter retrieved from satelliteborne sensors. SST can be measured using thermal infrared spectra, but these spectra are also sensitive to atmospheric water vapor. In order to eliminate the water vapor absorption, two or three thermal infrared bands are used. The most popular algorithm to retrieve SST was developed by NOAA for AVHRR on NOAA satellites. It is an empirical algorithm and called MCSST (multichannel SST algorithm) (McClain et al. 1985). The two band MCSST algorithm is shown as follows:

$$T_S = a_1 T_4 + a_2 (T_4 - T_5) + a_3 (T_4 - T_5)(\sec\theta_z - 1) + a_4$$

Here,

T_S: Sea surface temperature
T_4: Brightness temperature of 11 μm band
T_5: Brightness temperature of 12 μm band
θ_z: Satellite zenith angle at the ground
a_1–a_4: Coefficients which are determined experimentally

This algorithm works well, and its accuracy is around 0.5 K. However, it fails when the sea surface is very calm. One of the problems of SST measured by thermal IR spectra is that it measures sea skin temperature, i.e., less than 1 mm depth temperature, while in situ measurement usually measures 1 m depth temperature. Therefore, the SST retrieved from satellite IR sensors is inherently different from in situ measurements. However, usually, the differences are within the accuracy.

Another way of measuring SST from space is to use microwave region. IR sensors cannot be used over clouded areas, but microwave sensors can measure SST under clouds. In the microwave region, lower frequency is more sensitive to SST, and C band is the most appropriate frequency. Several corrections are necessary to retrieve SST from microwave radiometers. Refer to the AMSR-E case (AMSR/AMSR-E SST algorithm 2002). The problem is that the spatial resolution of low frequency microwave is very low. For the AMSR-E case, the spatial resolution of SST is around 50 km compared to 1 km resolution of MODIS which use thermal IR. Figure 21a shows 8 day composite of global SST retrieved from MODIS on Aqua, and Fig. 21b shows 7 day composite of global SST retrieved from AMSR-2.

SST is also very useful for monitoring oscillations like El Nino. Figure 22 shows very strong El Nino occurred in 2015.

Sea Surface Salinity

Sea surface salinity (SSS) is one of the key parameters which controls the deep sea water circulation (Lagerloef and Schmitt 2006; Lagerloef 2002; Office and US CLIVAR 2007). However, only low frequency microwave, i.e., L-band microwave, is sensitive to SSS. Also it depends on sea surface winds and column water vapor. Aquarius satellite can simultaneously measure SSS, sea surface winds, and water vapor. Figure 23 shows global sea surface salinity measured by Aquarius.

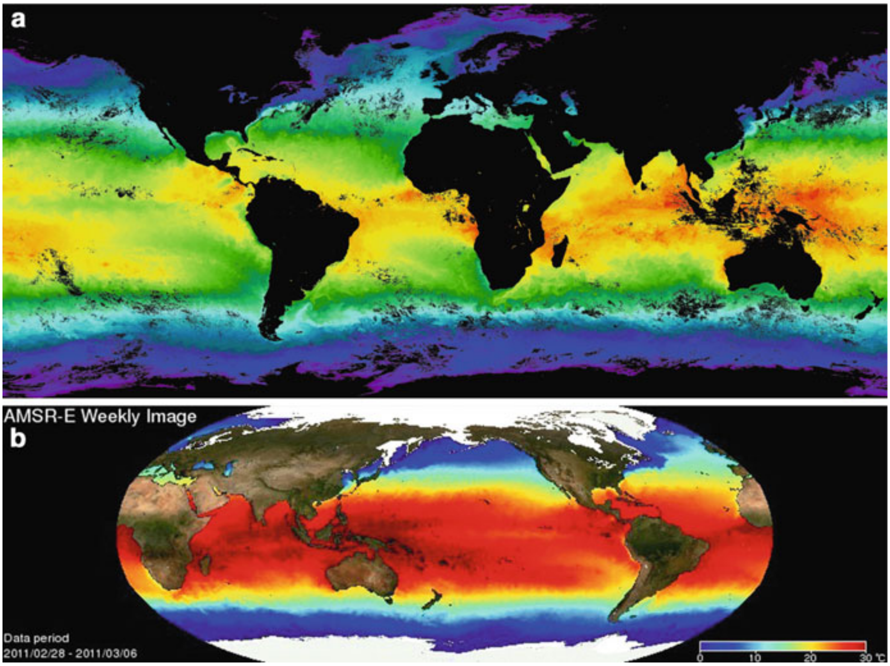


Fig. 21 Sea surface temperature retrieved from thermal infrared (**a**) and C-band microwave radiometer (**b**). (**a**) SST retrieved from MODIS on Aqua (SEA SURFACE TEMPERATURE 2015) (Courtesy of NASA). (**b**) SST retrieved from AMSR-2 (AMSR-2 Weekly Image 2014) (Courtesy of JAXA)

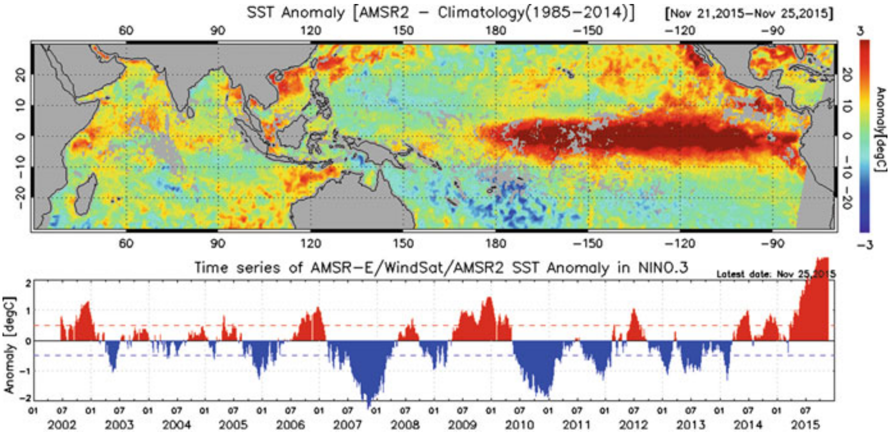


Fig. 22 (a) SST anomaly from climatological values. (b) Temporal variation of averaged SST anomaly of the El Niño monitoring region-3 (NINO.3) from June 2002 to November 2015 (El Niño phenomenon being close to the strongest on record 2015) (Courtesy of JAXA)

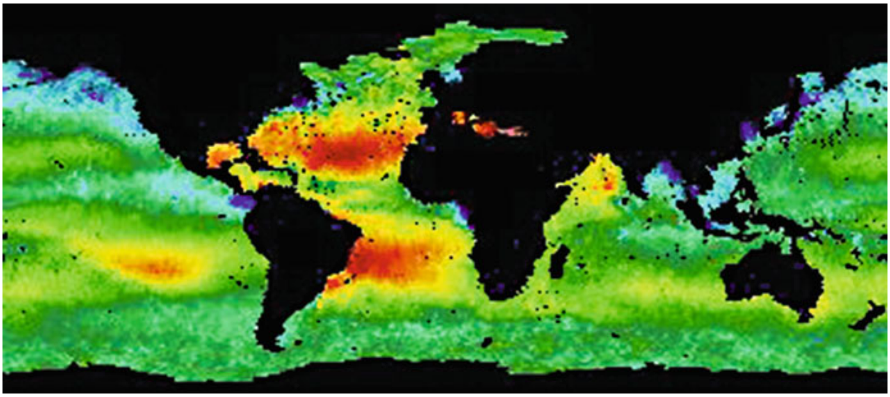


Fig. 23 Global sea surface salinity measured by Aquarius (Aquarius L3 Image Browser 2015) (Courtesy of NASA)

Sea Surface Wind

Sea surface winds are very important parameters which decide surface ocean circulations. It is also very important to find and track cyclones. Sea surface winds can be retrieved from microwave scatterometers, microwave altimeters, and microwave radiometers. For a microwave scatterometer, it measures Bragg scatterings from capillary waves on the sea surface. This scattering depends on surface wind speeds, directions, and incidence angles. The relationship between these parameters and back scattering coefficient can be described by geophysical model function, which is determined by empirical data (Jones et al. 1981, Naderi et al. 1991). Figure 24 shows a sample of geophysical model function. As can be seen from this figure, there are four ambiguities for the surface wind directions. In order to eliminate this ambiguity, back scatterings are measured from three directions. Still, there remain 180° ambiguities and these ambiguities are removed by postprocessings. Figure 25 shows an example of sea surface wind vectors measured by RapidScat. Until now, two frequencies are used for microwave scatterometer, i.e., Ku band and C band. Ku-band scatterometer is more accurate and can measure slow wind speed, while C-band scatterometer can measure higher wind speed and more insensitive to the rainfall.

Sea surface winds measurement using microwave altimeter depends on other mechanisms. It measures quasi-specular scatterings from the sea surface, and these scatterings are inversely proportional to the wind speed. Microwave altimeter cannot measure wind directions. Microwave radiometer also can measure sea surface wind speed. Most of microwave radiometers can measure only wind speed and not wind directions. Wind speed modifies the emissivity of the sea surface, and these modifications depend on frequencies, and wind speed also modifies the polarization. Depending upon these modifications, sea surface wind speed can be retrieved from microwave radiometer data. However, wind direction also slightly modifies brightness, temperature, and polarizations. With the Stokes vector measurement (i.e., four

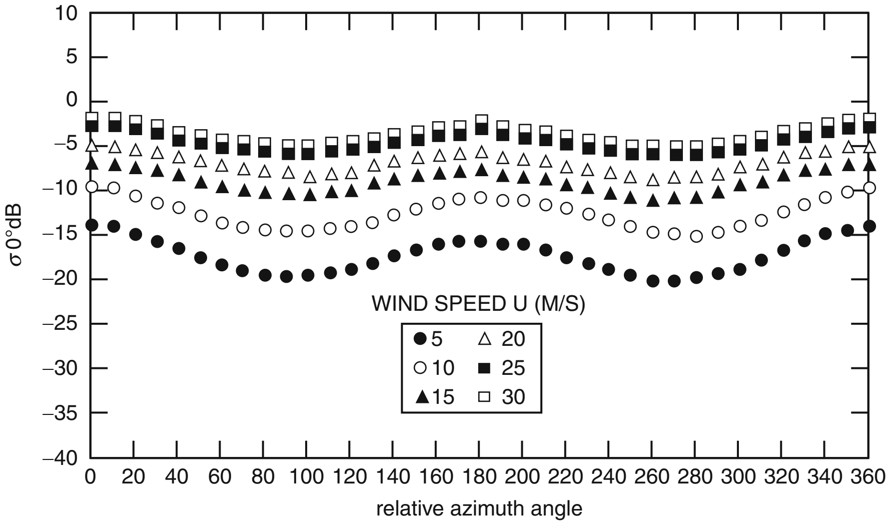


Fig. 24 A sample of geophysical model function (Japan Association of Remote Sensing 2001)

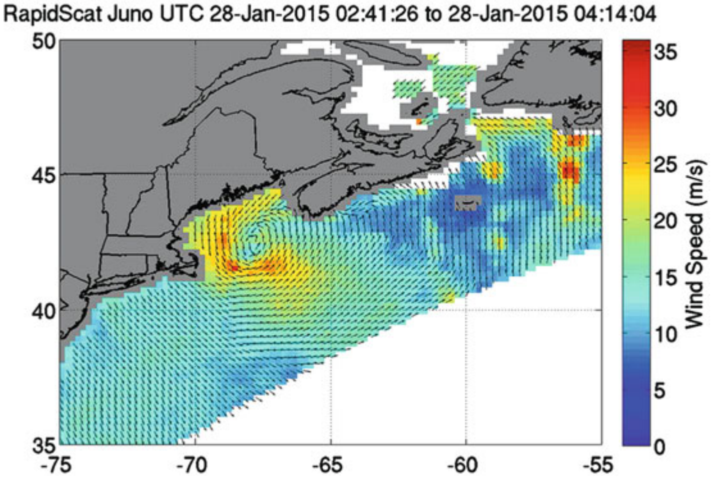


Fig. 25 A sample of sea surface wind vector measured by SeaWinds on RapidScat (NASA RapidScat Proving Valuable for Tropical Cyclones 2015) (Courtesy of NASA/JPL)

polarization measurement), wind direction also can be retrieved (Wentz et al. 2002, Jelenak et al. 2004). Windsat has this capability and measures sea surface wind vector. However, wind vector measurement using microwave radiometer can be done over rather small range of wind speed (cannot measure less than 6 m/s), and also it is largely affected by rainfall. Figure 26 shows an example of sea surface wind vector retrieved from Windsat.

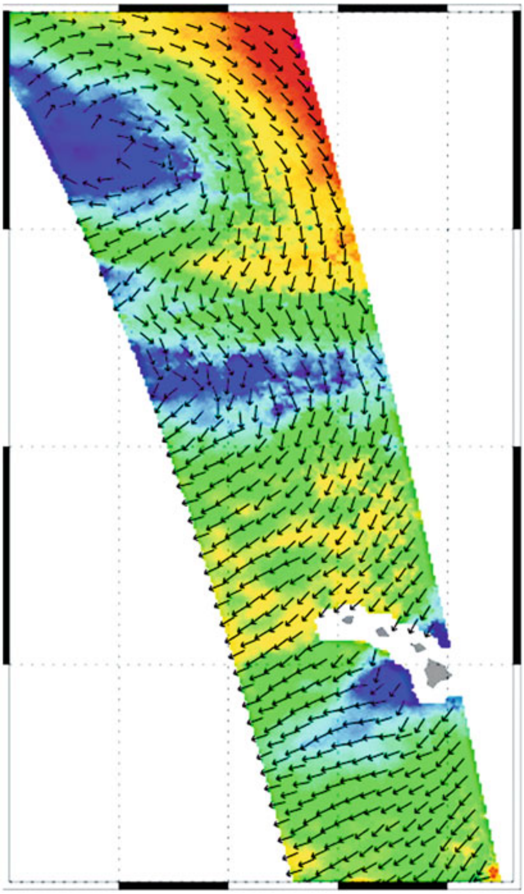


Fig. 26 A sample of sea surface wind vector measured by Windsat (Windsat images 2003) (Courtesy of NRL)

Sea Surface Height

Sea surface height is important to understand the ocean dynamics of sea current, geoid, etc. Sea surface height (SSH) is usually measured by microwave altimeter. The principle of satellite altimetry is shown in Fig. 27. The ground height or the sea surface height is measured from the reference ellipsoid. If the altitude of the satellite, Hs, is given as the height from the reference ellipsoid, the sea surface height $HSSH$ is calculated as follows:

$$HSSH = Hs - Ha$$

Here,

Ha: Measured distance between satellite and the sea surface

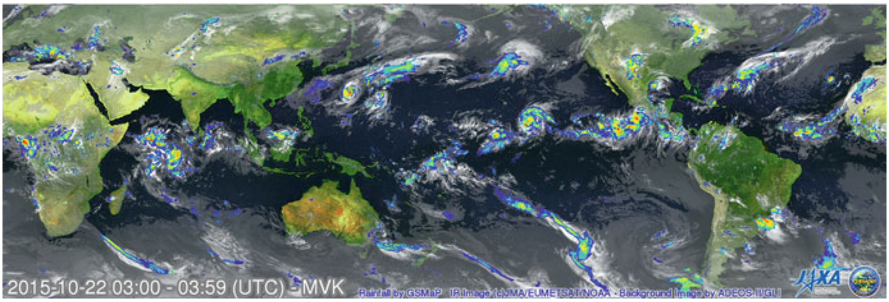
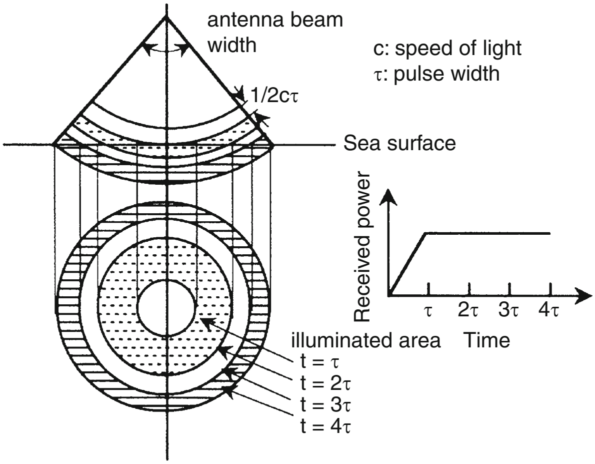


Fig. 27 The principle of satellite altimetry (Japan Association of Remote Sensing 1985)

Fig. 28 Change of illuminated area and received power (Japan Association of Remote Sensing 2001)



Ha is measured on the basis of the travel time of the transmitted microwave pulses. From the time ($t = 0$), when the first edge of pulse arrives at the surface, to the time ($t = t$), when the end edge of a pulse with a width of arrives at the surface, the received power increases linearly as shown in Fig. 28. The received pulses are composed of echoes from various parts of the sea surface. Therefore the travel time from a satellite to the sea surface can be calculated by averaging the received pulses. Pulse compression techniques will be also applied in order to obtain a high frequency pulse for improvement of the resolution.

The accuracy of altimeter itself is around 1 cm, but it measures only distance between the satellite and the sea surface, hence it is most important to know the satellite position. In order to measure the satellite position, several means have been developed. Now, the most used instrument is GPS. Another way which can measure satellite position in high accuracy is to use ground-based lasers and corner mirrors onboard the satellite. Both means have similar accuracy, i.e., 2–3 cm. Figure 29 shows an example of sea surface height measured by Jason-1. Anyway, microwave

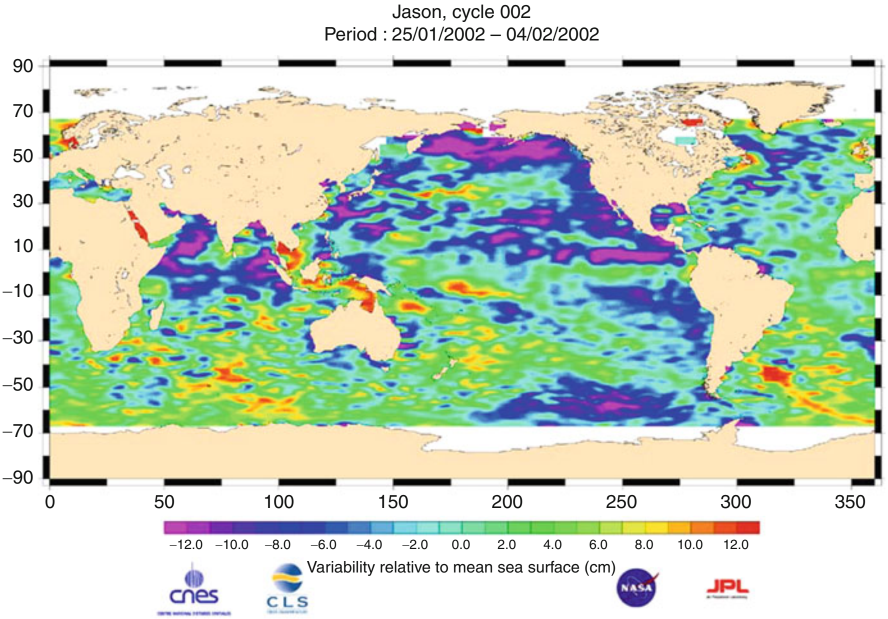


Fig. 29 Global sea surface height measured by JASON1 (Jason 2002) (Courtesy of NASA and CNES)

altimeter data are affected by the water vapor. So, microwave radiometer which can measure total water vapor is usually installed with microwave altimeter for water vapor correction.

Another way of measuring sea surface height is to use lidar. Icesat measured sea surface height using a lidar. Lidar has much more smaller IFOV compared to microwave altimeter, but as ocean is very dark, signal to noise ratio of lidar is not so high.

Ocean Color

Ocean color is one of the key components to characterize the oceanic living matters. Usually, main retrieved geophysical parameters from ocean color sensors are chlorophyll-a concentrations. Chlorophyll-a concentrations are assumed to be proportional to phytoplankton concentrations. Phytoplankton is the bottom of oceanic food chain and also plays an important role in carbon cycles in ocean. Until now, most of chlorophyll-a retrieval algorithms use empirical algorithms after atmospheric correction. The reflected light from open ocean is mostly composed of scattered light from the atmosphere. Only 10 % of received light at the satellite comes from the ocean surface. So, the atmospheric corrections are most important in chlorophyll retrievals. Conventional algorithms use near-infrared channels for aerosol estimation. As ocean can be thought to be dark in these channels, near-infrared radiances are thought to be reflected from atmospheric aerosols (Gordon and Clark 1981; Gordon and Wang 1994; Siegel et al. 2000).

Here, one of such algorithms is presented according to Gordon (Gordon and Voss 2004). The radiance received by a sensor at the top of the atmosphere (TOA) at a wavelength λ_i,$L_t(\lambda_i)$ can be written as follows:

$$L_t(\lambda_i) = L_{\text{path}}(\lambda_i) + T(\lambda_i)L_g(\lambda_i) + t(\lambda_i)L_{wc}(\lambda_i) + t(\lambda_i)L_w(\lambda_i)$$

Here,

$L_{\text{path}}(\lambda_i)$: The radiance generated along the optical path by scattering in the atmosphere and by specular reflection of atmospherically scattered light (skylight) from the sea surface
L_g (i): The contribution arising from specular reflection of direct sunlight from the sea surface (sun glitter)
L_{wc} (i): The contribution arising from sunlight and skylight reflecting from individual white caps on the sea surface
$L_w(i)$: The desired water leaving radiance
T: The direct transmittance of the atmosphere
t: The diffuse transmittance of the atmosphere

The above equation can be converted to the reflectance as follows:

$$\rho_t(\lambda_i) = \rho_{\text{path}}(\lambda_i) + T(\lambda_i)\rho_g(\lambda_i) + t(\lambda_i)\rho_{wc}(\lambda_i) + t(\lambda_i)\rho_w(\lambda_i)$$

In the first approximation, the sun glitter term and the white cap term can be ignored. So, the most important part is to estimate $\rho_{\text{path}}(\lambda_i)$. It can be written as follows:

$$\rho_{\text{path}} = \rho_r(\lambda) + \rho_a(\lambda) + \rho_{ra}(\lambda)$$

Here, ρ_r is the reflectance resulting from multiple scattering by air molecules (Rayleigh scattering) in the absence of aerosols, ρ_a is the reflectance resulting from multiple scattering by aerosols in the absence of the air, and ρ_{ra} is the interaction term between molecular and aerosol scattering (Antoine and Morel 1998, Deschamps et al. 1983). In the single scattering approximation (called CZCS algorithm), ρ_{ra} can be neglected. As ρ_r can be calculated rather accurately, the problem is how to estimate ρ_a. The next approximation is that ocean totally absorbs near-infrared light, i.e., there is no reflected light from the ocean in near-infrared channels. Thus, the ρ_a in near-infrared channels can be computed by subtracting ρ_r from ρ_t. Using two near-infrared channels, i.e., λ_s (shorter wavelength) and λ_l (longer wavelength), the ratio can be calculated as follows:

$$\varepsilon(\lambda_s, \lambda_l) = \frac{\rho_a(\lambda_s)}{\rho_a(\lambda_l)}$$

For other i^{th} channels, $\varepsilon(\lambda_i, \lambda_l)$ should be known. For this purpose, radiative transfer calculations are done for many kinds of aerosols. The proper aerosol model is selected from the $\varepsilon(\lambda_s, \lambda_l)$ value. For real MODIS ocean color retrievals, more sophisticated algorithms including multiple scatterings are used, but the principle is the same.

These algorithms have rather higher accuracy in open ocean (called case 1 water), but not so good accuracy in turbid coastal waters (called case 2 water). Analytical algorithms are now developed for case 2 waters (Chomkoa et al. 2003, Ruddick et al. 2005, Bo-Cai Gao et al. 2007).

Chlorophyll-a data are used to estimate the primary production in the ocean. The most popular algorithm to estimate ocean primary production was proposed by Behrenfeld and Falkowski (Berenfeld and Falkowski 1997). This algorithm can be described as follows:

$$IPP = 0.66125 \times P_{opt}^{B} \times \frac{E_0}{E_0 + 4.1} \times DL \times Z_{eu} \times Chl.a$$

Here,

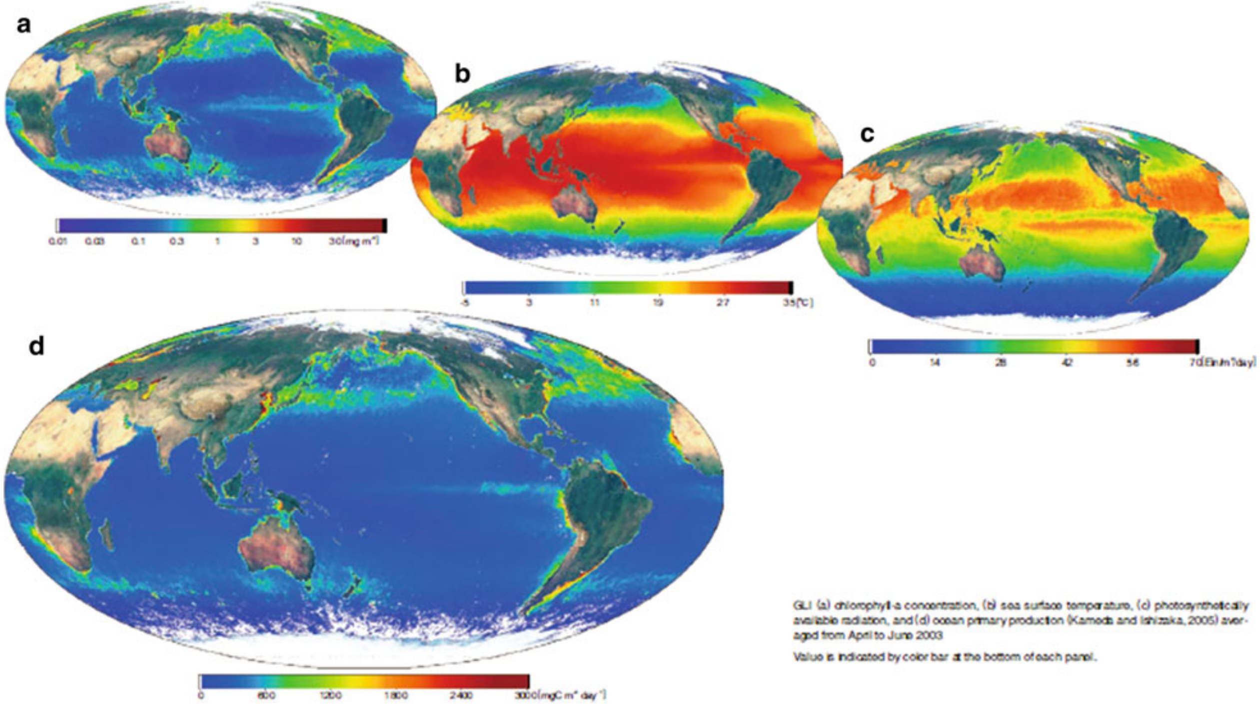
IPP: Integrated primary production of 1 day (mgCm^{-2}day^{-1})
P_{opt}^{B}: Maximum carbon fixation quantity within euphotic zone per unit chlorophyll a (mgC/mgChl•hour)
E_0: Photosynthetic active radiation (PAR)(mol quanta m^{-2})
DL: Day length (hour)
Zeu: Depth of euphotic zone (m)
$Chl.a$: Chlorophyll-a concentration at the sea surface

In this algorithm, P_{opt}^{B} is expressed as seventh order polynomials of sea surface temperature. From satellite data, SST and PAR can be retrieved. Figure 30 shows SST, PAR, Chlorophyll-a, and primary production estimated from GLI on ADEOS2. Many algorithms which modify this algorithm are proposed (Behrenfeld et al. 2005; Carr et al. 2006; Asanuma 2006; Ishizaka et al. 2007; Westberry et al. 2008).

Land Applications

Topography

Topography is one of the most important parameters of land and used for most of maps. Before the satellite era, topography or altitudes of land were mainly measured by stereo aerial photographs. However, it is rather difficult to measure global topography by aerial photographs. Satelliteborne sensors have provided the means to generate global topographic data. There are several kinds of sensors which can be used to measure land topography. One is the stereo images which are very similar with aerial photographs. The main part of retrieving DSM (digital surface model) from stereo pairs is image matching algorithms. Many kinds of image matching



GLI (a) chlorophyll-a concentration, (b) sea surface temperature, (c) photosynthetically available radiation, and (d) ocean primary production (Kameda and Ishizaka, 2005) averaged from April to June 2003

Value is indicated by color bar at the bottom of each panel.

Fig. 30 SST, PAR, Chlorophyll-a, and primary production estimated from GLI on ADEOS2 (Primary Productivity of Phytoplankton 2004) (Courtesy of JAXA) algorithms are proposed, but the least squares matching algorithm is the most powerful (Gruen 1985). There still remain problems, especially, occlusion problems and mismatching caused by very homogeneous targets. Second is the synthetic aperture radar (SAR) (Zebker and Goldstein 1986; Goldstein et al. 1988; Madsen and Zebker 1998). The third is to use lidar from satellites.

The stereo imaging from satellites has started from SPOT in 1986. However, SPOT made stereo pairs from different orbit. So the stereo pair images were taken on different days. It is rather difficult to collect the stereo pair images with the same conditions. After SPOT, several satellites were installed with multiple sensors, which made possible to get stereo images almost simultaneously. These multiple sensors were on JERS-1, ASTER on Terra, and PRISM on ALOS. Only from ASTER, global digital surface model (DSM) was generated and distributed (ERSDAC 2002). It has 30 m resolution and covers almost all over the world. Figure 31 shows an example of ASTER DEM (GDEM). Higher resolution DSMs can be generated from high-resolution sensors like PRISM and many commercial high-resolution sensors, but the DSM generations are limited and cover very few areas.

The principle of DSM generation from stereo pairs can be described as follows. Figure 32 shows the configuration of stereo pair images taken from parallel directions. Here, object coordinate space is expressed by (x, y, z) and image coordinates are expressed by (u, v). The origin of image coordinates is the crossing point of image projection plane and the optical axis, and u and v axes are parallel with x and y axes, respectively. The distance between two cameras L and R are c for both cameras, and the projection center of each camera OL and OR is $(-b/2, 0, 0)$ and $(b/2, 0, 0)$ both on the x-axis. The optical axes of camera L and R are parallel with z-axis.

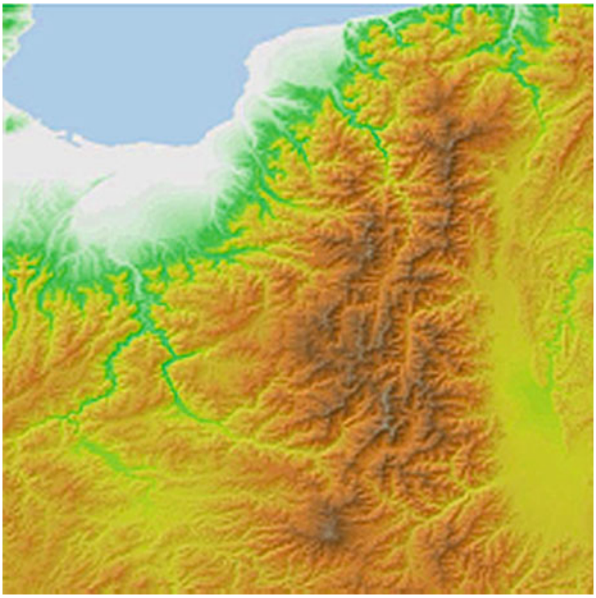


Fig. 31 An example of ASTER DEM (Perry and Kruse 2011) (Courtesy of ERSDAC)

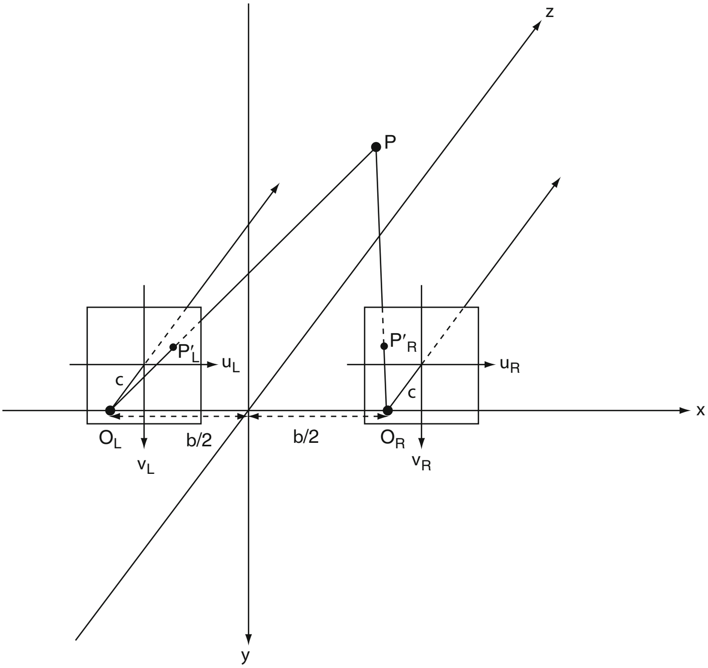


Fig. 32 Principle of DSM generation from stereo pairs (Takagi and Shimoda 2004a)

The next equation stands for the image P'_L (u_L, v_L) of target P (x, y, z) in image L.

$$\frac{u_L}{c} = \frac{x + b/2}{z} \tag{1a}$$

$$\frac{v_L}{c} = \frac{y}{z} \tag{1b}$$

Similarly, the next equation stands for the image P′R (u_R, v_R) of target P(x,y,z) in image R.

$$\frac{u_R}{c} = \frac{x - b/2}{z} \tag{2a}$$

$$\frac{v_R}{c} = \frac{y}{z} \tag{2b}$$

From Eqs. (1a), (1b), (2a), and (2b), the next equation can be deduced.

$$x = \frac{u_L z}{c} - \frac{b}{2} = \frac{u_R z}{c} + \frac{b}{2} \tag{3a}$$

$$y = \frac{v_L z}{c} = \frac{v_R z}{c} \tag{3b}$$

From the above equations, z can be calculated as follows:

x and y can be obtained by substituting z to Eqs. 3a and 3b.

$$z = c \frac{b}{u_L - u_R} \tag{4}$$

Another way to measure topography is to use interferometric SAR. The most important project to generate global DSM using interferometric SAR was Shuttle Radar Topography Mission (SRTM) (NASA/JPL 2016). SRTM was conducted in 2000 in STS99 mission. Two antennas with distance of 60 m were used, and C and X band SAR were used. It covered ±60° latitudinal areas and the original resolution is 30 m. Global 90 m resolution data are distributed to the public. Figure 33 shows the SRTM DSM of Africa and Middle East.

Higher resolution DSM is sometimes required for specific applications. Now, there is a 5 m grid global DSM dataset obtained from ALOS PRISM. Figure 34 shows comparison between 5 m, 30 m, and 90 m DSM.

Lidar also can be used to generate land topography. The lidar on Icesat was used to retrieve land topography (GLAS/ICESat L1 and L2 Global Altimetry Data 2014), but its main mission was to measure the height of ice sheet and the changes of ice

Fig. 33 SRTM DSM of Africa and Middle East (NASA/JPL 2004) (Courtesy of NASA)

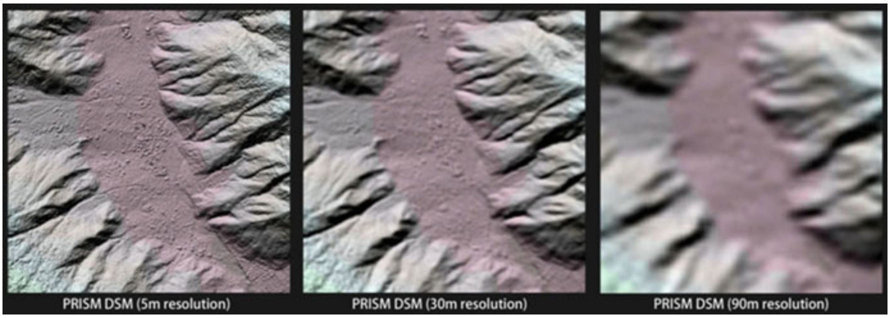


Fig. 34 Comparison of DSM between 5 m, 30 m, and 90 m grid (Tadono et al. 2014) (Courtesy of JAXA)

sheets. There are land topography products, but as the lidar can measure only the nadir of the satellite, the coverage is rather sparse.

Geometric Corrections and Map Projection

Images obtained from satellite include many kinds of geometric distortions. In the remote sensing applications, especially in the land applications, it is very important to correct these distortions. These distortions include inner distortions (lens distortion, distortions within the sensor, etc.) and outer distortions (Earth curvature, spacecraft position, and attitude error, etc.). There are two kinds of correction algorithms. One is to correct these distortions systematically based on known parameters. Another algorithm is to correct the image distortions and map to existing maps using ground control points (GCP). There are also some combinations of the above algorithms. Until some time ago, the parameters obtained from the satellite were not accurate enough to achieve accurate corrections. Especially, the accuracy of satellite position and attitude were not enough. However, recent satellite has GPS receivers and star trackers, providing almost sufficient accurate position and attitudes. GCPs are corresponding points in images and in corresponding maps. It is a rather tedious operation to select accurate GCPs. However, in the GCP correction algorithm, this process is of course very important.

In the case of systematic corrections, map projection is finally required. There are many kinds of map projections, but usually several projections are used, i.e., universal transverse Mercator (UTM), Mercator, polar stereo, and latitude longitude grid. UTM is usually used for large-scale maps. Mercator is mainly used for oceanic applications, and polar stereo is mainly used for meteorological applications. Latitude longitude grid is also frequently used because it can be used as starting data for any kinds of projections. Before map projections, orthographic projection is also sometimes important when the target area is not flat. When there are mountains in the image, these mountains are projected at a slant. Orthographic projection eliminates these distortions but need DEM of the target areas.

Radiometric Corrections

Radiometric corrections are important when geophysical parameters should be retrieved from satellite data. Radiometric corrections start from converting sensor output values to radiance. This procedure is based upon calibrations. Three kinds of calibrations are usually conducted. The first one is based on the ground calibrations before launch. Ground-based calibrations for optical sensors are done using standard light sources like integrating spheres with standard light sources. However, sensitivities of sensors usually change after launch. After launch, two kinds of calibrations are conducted. One is to use onboard calibration sources. For visible and near-infrared sensors, onboard calibration sources are diffused sunlight, light sources like light bulbs or LEDs, Moon, and dark space. For thermal infrared sensors and microwave radiometers, onboard calibration sources are deep space and onboard black body. The third calibration method is vicarious calibrations. In vicarious calibrations, homogeneous ground target and atmospheric measurements are used. The accuracies of calibrations are as follows. For ground-based visible and near-infrared regions, accuracies are 2–3 %. The highest accuracy in these regions on board is to use the Moon, and its accuracy is also around 2–3 %. The accuracy of vicarious calibrations is around 5–6 %. For thermal infrared, the accuracies depend on the emissivity of the onboard black body, and the highest accuracy is around 0.1 K. For active microwave sensors, different approaches are used. They use ground-based corner reflectors or active radar calibrators, which receive the radar pulse and then send back to the satellite sensor.

Second-step radiometric correction is atmospheric correction. In the visible and near-infrared region, corrections are done to atmospheric scattering by atmospheric molecules and aerosols. Scattering by atmospheric molecules is Rayleigh scattering (Jackson 1962; Craig and Thirunamachandran 1989), and it is rather easy, because the compositions of atmospheric molecules are fairly steady. Aerosols cause Mie scattering (Mie 1908b), but its correction is very difficult. It is because aerosol concentration and other properties (radius, species, etc.) change largely depending on time and place. Many kinds of aerosol correction methods are proposed (Chavez 1988; Gordon and Clark 1981; Kaufman 1989; Vermote et al. 1997), but still their accuracies are not so high. Figure 35 shows MODIS images before and after aerosol correction.

Land Cover and Land Use

Land Cover Categories

Land cover and land use are the basic information of land. These data were first used for urban planning, but now they are the fundamental data of land to understand the Earth environment. Land cover and land use are different concepts. Land cover is just what is there. On the contrary, land use is a functional concept. Sometimes they are the same, e.g., forest, agricultural land and lake. However, for instance, commercial areas and industrial areas are land use categories. From satelliteborne

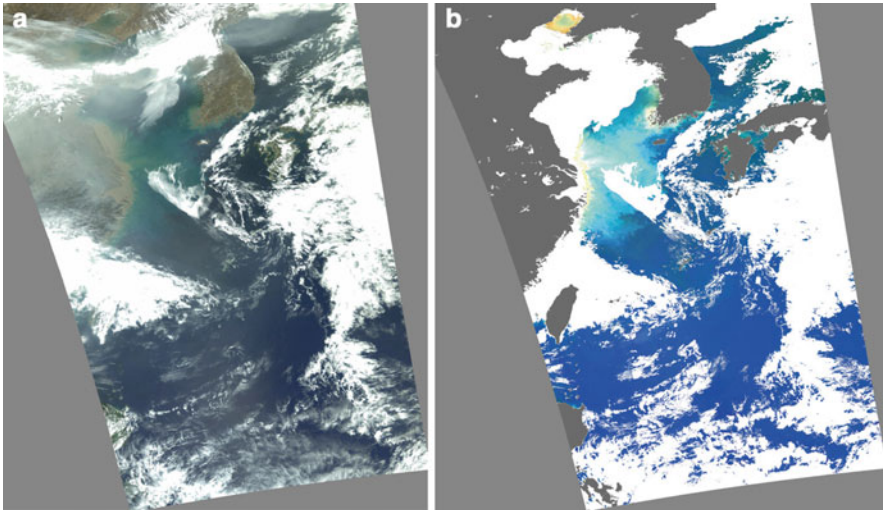


Fig. 35 MODIS images before (**a**) and after (**b**) the aerosol correction near west Japan and Korea (JAXA 2015b)

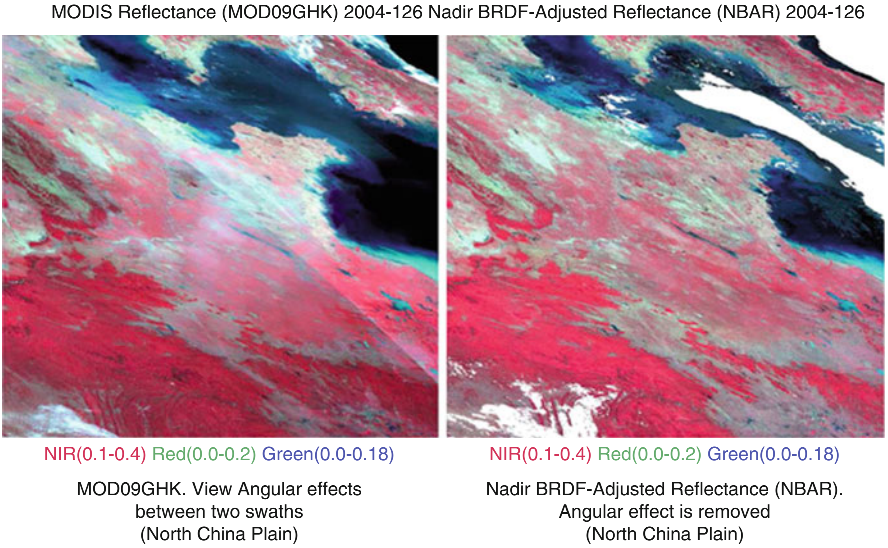
sensors, only land cover can be observed, and land use may be estimated using land cover data. Land cover maps are used, e.g., for urban plannings, vegetation cover estimates, or carbon cycle estimations.

In order to generate land cover maps, usually image classifications are used. Before starting the image classifications, geometric and radiometric corrections are necessary. Geometric corrections are usually necessary, but for radiometric corrections, it is dependent upon the applications. For instance, when the target area is not so large, land cover classifications can be done without radiometric corrections. However, when the target areas are large, e.g., global land cover, radiometric corrections are inevitable. There are effects of sun angles, aerosols, etc., which obscure the accurate classifications. MODIS is distributing land surface reflectance product for these purposes (Vermote and Vermeulen 1999). Also, there is a product called Nadir BRDF adjusted reflectance product (NBAR) which corrects the bidirectional reflectance distribution function (BRDF) effect (Schaaf 2010). Figure 36 shows the difference between MODIS surface reflectance product and NBAR product.

Land cover categories should be defined before classifications. There are several standards on the land cover categories. Tables 1, 2, and 3 show samples of these standard land cover categories for global land cover mapping. These standard classification categories are sometimes useful, but for local land cover mapping, much more specified categories are usually necessary.

Classification Features

Many kinds of features are used for land cover classifications. The most appropriate features depend on the resolutions of images. For high-resolution sensors, i.e., less



MOD09GHK. View Angular effects between two swaths (North China Plain)

Nadir BRDF-Adjusted Reflectance (NBAR). Angular effect is removed (North China Plain)

Fig. 36 The difference between MODIS surface reflectance product and NBAR product. *Left*: MODIS reflectance; *Right*: NBAR (Schaaf et al. 2007)

Table 1 Land cover categories of IGBP (FAO 2000)

Class	Class name	Description
11.	Cultivated and managed terrestrial areas	Tree crops Shrub crops Herbaceous crops Graminoid crops Nongraminoid crops Managed lands
12.	Natural and seminatural terrestrial vegetation	Forest Woodland Thicket Shrubland Grasslands Sparse vegetation Lichens/mosses
23.	Cultivated aquatic or regularly flooded areas	Aquatic or regularly flooded graminoid crops Aquatic or regularly flooded nongraminoid crops
24.	Natural and semi-natural aquatic or regularly flooded vegetation	Forest Woodland Closed shrubs Open shrubs Grasslands Sparse vegetation Lichens/mosses
15.	Artificial surfaces and associated areas	Built-up areas Nonbuilt-up areas
16.	Bare areas	Consolidated areas Unconsolidated areas

than 3 m IFOV, spatial features are used in addition to spectral features. Image segmentation (sometimes called an object-oriented classifications or object-based classifications) is sometimes very useful for these kinds of images (Neubert 2001; Blaschke and Hay 2001; Hofmann 2001). For medium-scale images, i.e., 30–80 m IFOV, spectral features are used in most cases.

Table 2 Land cover of Food and Agricultural Organization (FAO)

Major land cover type with their structural domains	
A11. Cultivated and managed terrestrial areas	Tree crops Shrub crops Herbaceous crops Graminoid crops Nongraminoid crops Managed lands
A12. Natural and semi-natural terrestrial vegetation	Forest Woodland Thicket Shrubland Grasslands Sparse vegetation Lichens/mosses
A23. Cultivated aquatic or regularly flooded areas	Aquatic or regularly flooded graminoid crops Aquatic or regularly flooded nongraminoid crops
A24. Natural and semi-natural aquatic or regularly flooded vegetation	Forest Woodland Closed shrubs Open shrubs Grasslands Sparse vegetation Lichens/mosses
B15. Artificial surfaces and associated areas	Built-up areas Nonbuilt-up areas
B16. Bare areas	Consolidated areas Unconsolidated areas
B27. Artificial surfaces and associated areas	Artificial waterbodies Artificial snow Artificial ice
B28. Natural waterbodies, snow, and ice	Natural waterbodies Snow Ice

For global or regional land cover classifications, other features are used frequently. One of the most popular features is the time series of normalized difference vegetation indices (NDVI). NDVI is expressed by the following equation:

$$NDVI = \frac{NIR - R}{NIR + R}$$

Here,

NDVI: NDVI value
NIR: Value of near-infrared band
R: Value of near red band

Table 3 Land cover categories of USGS (Anderson et al. 1976)

Level I	Level II
1 Urban or built-up land	11 Residential
Agricultural land Rangeland Forest land Water Wetland Barren land Tundra Perennial snow or ice	12 Commercial and services 13 Industrial 14 Transportation, communications, and utilities 15 Industrial and commercial complexes 16 Mixed urban or built-up land 17 Other urban or built-up land 21 Cropland and pasture 22 Orchards, groves, vineyards, nurseries, and ornamental horticultural areas 23 Confined feeding operations 24 Other agricultural land 31 Herbaceous rangeland 32 Shrub and brush rangeland 33 Mixed rangeland 41 Deciduous forest land 42 Evergreen forest land 43 Mixed forest land 51 Streams and canals 52 Lakes 53 Reservoirs 54 Bays and estuaries 61 Forested wetland 62 Nonforested wetland 71 Dry salt flats 72 Beaches 73 Sandy areas other than beaches 74 Bare exposed rock 75 Strip mines, quarries, and gravel pits 76 Transitional areas 77 Mixed barren land 81 Shrub and brush tundra 82 Herbaceous tundra 83 Bare ground tundra 84 Wet tundra 85 Mixed tundra 91 Perennial snowfields 92 Glaciers

As NDVI is a kind of ratio between spectral bands, it has some ability to eliminate the radiometric distortions.

Classifiers

Maximum Likelihood Classifier

Many kinds of classifiers are used for land cover classifications. Here, maximum likelihood classifier (MLC), neural net, and support vector machine (SVM), which give the highest classification accuracy, will be briefly described. The most popular classifier is stochastic classifier. MLC is the most popular within stochastic classifiers. MLC can be described as follows:

The classifier which gives the minimum loss is to classify vector **x** to category C_i which gives the minimum $\sum \lambda_{ij} P\left(C_i \middle| \boldsymbol{x}\right)$, when classification target data vector is denoted by **x**, categories are $C = \{C_1, C_2, \cdot\cdot\cdot, C_\mathrm{n},\}$, and the loss is λ_{ij} when category C_i is misclassified to category C_j. This is called Bayesian decision rule. In the case of image classifications, the losses by misclassifications can be thought to be constant, so the target of the classification is to acquire $P\left(C_i \middle| \boldsymbol{x}\right)$. From Bayes theorem, $P\left(C_i \middle| \boldsymbol{x}\right)$ can be expressed as follows:

$$P\left(C_i \middle| \boldsymbol{x}\right) = \frac{p\left(\boldsymbol{x} \middle| C_i\right) P\left(C_i\right)}{\sum p\left(\boldsymbol{x} \middle| C_i\right)}$$

Here, probability density function $P(\mathbf{x}| C_i)$ is called a likelihood, and $P(C_i)$ is called an a priori probability of category C_i. As the denominator of the above equation is the same to each category, $p(\mathbf{x} |C_i)P(C_i)$ should be set up. As $P(C_i)$ is usually difficult to estimate, the vector **x** is usually classified to the category with the maximum $P(\mathbf{x}| C_i)$. This classifier is called a maximum likelihood method. When $P(\mathbf{x}| C_i)$ follows a normal distribution, $P(\mathbf{x}| C_i)$ is expressed as follows:

$$p(\boldsymbol{x}|C_i) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{V}_i|^{1/2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \overline{\boldsymbol{x}}_i \ )^t \boldsymbol{V}_i^{-1} (\boldsymbol{x} - \overline{\boldsymbol{x}}_i \ ) \right\}$$

Here,

$\boldsymbol{V}_i$: Variance-covariance matrix of category C_i
$\mathbf{x}_i$: Mean vector of **x** of category C_i

For the calculation simplicity, the logarithm of the above equation with inverted sign is used, and the **x** is classified to the category with maximum of the following equation:

$$\log|\boldsymbol{V}_i| + (\boldsymbol{x} - \overline{\boldsymbol{x}}_i \ )^t \boldsymbol{V}_i^{-1} (\boldsymbol{x} - \overline{\boldsymbol{x}}_i \ )$$

The second term of the above equation is called a Mahalanobis' distance. Many other kinds of classifiers are used in the remote sensing applications. They are, for instance, decision tree classifiers, neural net, and support vector machine (SVM).

Neural Net

Artificial neural network is one of the strongest classifier for remote sensing data. The most popular architecture of neural network for classification is the multilayer perceptron as shown in Fig. 37. Multilayer perceptron is composed of input layer,

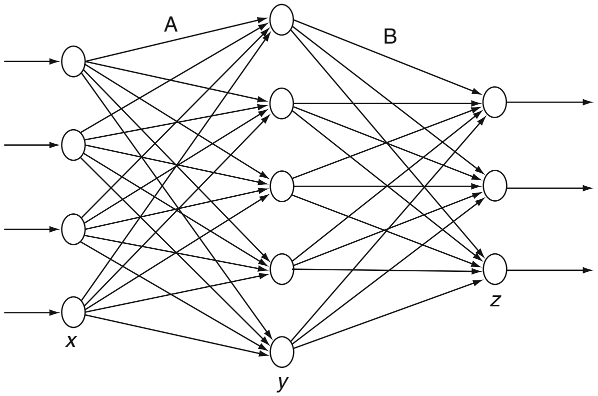


Fig. 37 A block diagram of multilayer perceptron (Takagi and Shimoda 2004b)

output layer, and hidden layers. The number of hidden layers may change according to the problem. In Fig. 37, only one hidden layer is shown.

The output signal $\boldsymbol{z} = (z_1, \cdots, z_k)^t$ to the input signal $\boldsymbol{x} = (x_1, \cdots, x_n)^t$ is expressed as follows:

$$\left.\begin{aligned} \xi_j &= \sum_{i=1}^{I} a_{ij} x_i + a_{0j} \\ y_j &= f_{hidden}\left(\xi_j\right) \\ \eta_k &= \sum_{j=1}^{J} b_{jk} y_j + b_{0k} \\ z_k &= f_{out}(\eta_k) \end{aligned}\right\} \qquad (5.3.12)$$

Here,

a_{ij}: Weight from ith input to jth hidden layer unit
b_{jk}: Weight from jth hidden layer unit to kth output unit
a_{0j} b_{0k}: Biases of jth unit of hidden layer and kth unit of output layer, respectively
f_{hidden} f_{out}: Input–output function of hidden layer unit and output layer unit, respectively

Usually, logistic functions are used for f_{hidden}, and f_{out} is defined according to each problem.

For the learning process of this kind of neural network, error back propagation learning (Rumelhart 1986a, b) is used. This algorithm is shown as follows:

Suppose the combinations of input data and training data as $\left\{\boldsymbol{x}_p, \ \boldsymbol{u}_p \middle| \ p = 1, \cdots, P\right\}$. In this algorithm, the following evaluation criterion is minimized.

$$\varepsilon_{\text{emp}}^2 = \sum_{p=1}^{P} \left\|\boldsymbol{u}_p - \boldsymbol{z}_p\right\|^2 = \sum_{p=1}^{P} \varepsilon_{\text{emp}}^2(p)$$

Using the method of steepest descent, the following iterative equation is deduced.

$$a_{ij}^{l+1} = a_{ij}^{l} + 2\alpha \sum_{p=1}^{P} \gamma_{pj} \nu_{pj} x_{pi}$$

$$b_{jk}^{l+1} = b_{jk}^{l} + 2\alpha \sum_{p=1}^{P} \delta_{pk} y_{pj}$$

Here,

α: Learning rate (>0)

$$\nu_{pj} = y_{pj}\left(1 - y_{pj}\right)$$

$$\gamma_{pj} = \sum_{k=1}^{K} \delta_{pk} b_{jk}$$

$$\delta_{pk} = u_{pk} - z_{pk}$$

Neural network classifier usually gives higher accuracy than MLC. It can deal with nonlinear problem as well as nonnormal distribution problems.

Support Vector Machine

Support vector machine (SVM) (Vapnik 1995) started from the pattern classifier of linearly separable two classes. It generates a hyperplane with the largest margin (the minimum distance from training samples to the hyperplane). The learning process uses Lagrange's method of undetermined multipliers and formulated by convex quadratic programming. However, most problems are not linearly separable and land cover classification needs many class classifications. In order to deal with nonlinearly separable problems, the original space is nonlinearly projected to higher order space. In the real calculations, this projection is not actually calculated. Instead, this calculation is replaced by kernel function calculations. This process is called a kernel trick (Schölkopf et al. 1996).

Examples of Land Cover Classifications

Several approaches have been made to generate global land cover maps. Three kinds of approaches are briefly introduced here. The first one is MODIS land cover product generated by NASA using MODIS data. Many kinds of features are used in this classification, i.e., land/water mask, Nadir BRDF adjusted reflectance (NBARs), spatial texture derived from 250 m red band, directional reflectance information, EVI, snow cover, land surface temperature, and DTM. Classifier is the decision tree classifier (Strahler 1999). An example of this product is shown in Fig. 38.

The second example is generated by ESA using MERIS data. In this approach, also NBAR-like products are used for the classification. The classifier is mainly

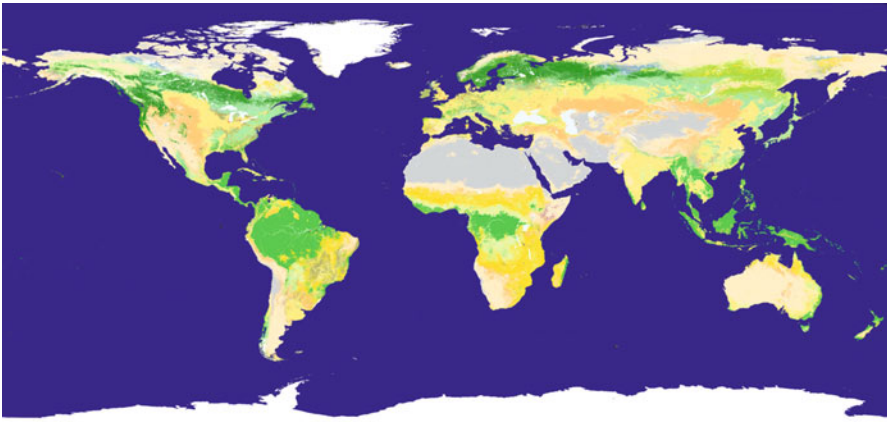


Fig. 38 Global land cover map generated by NASA (Global land cover map 2002) (Courtesy of NASA)

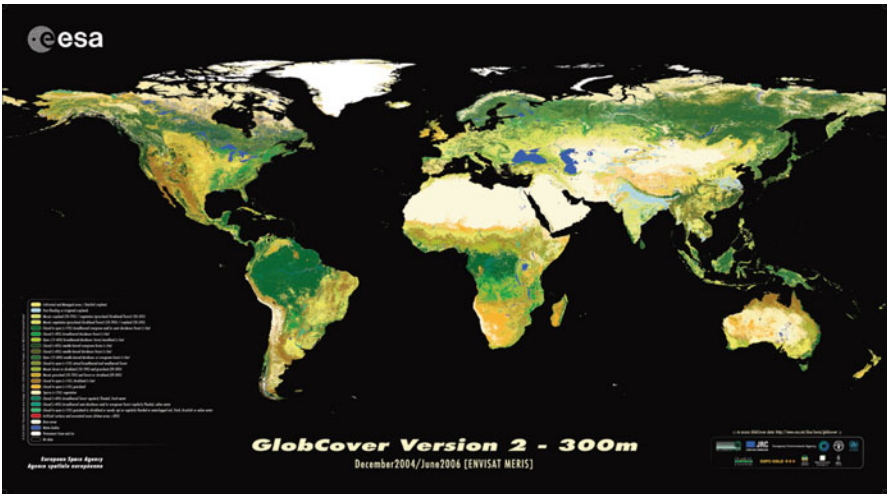
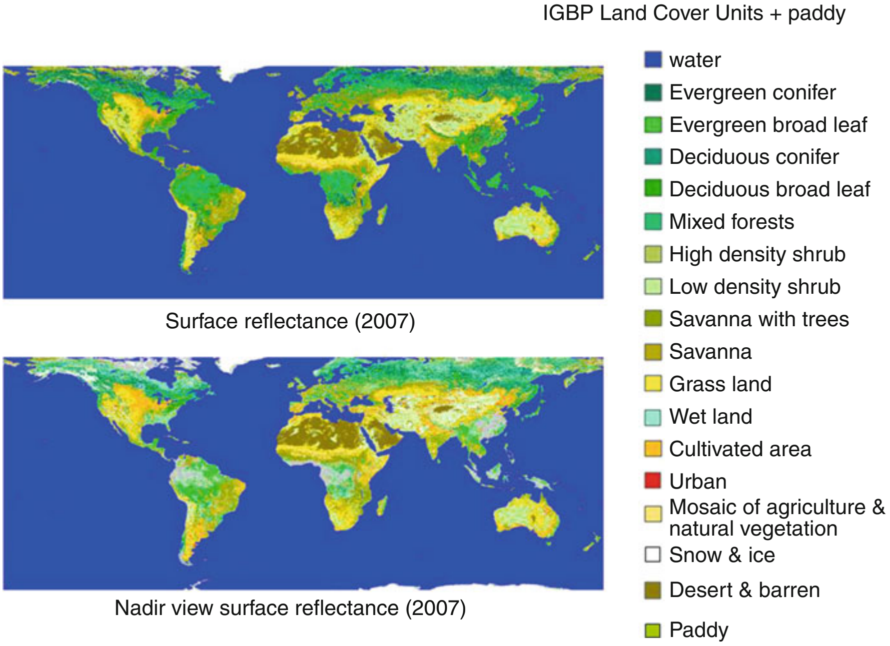


Fig. 39 Global land cover map using MERIS data generated by ESA (Global land cover map using MERIS data 2008) (Courtesy of ESA)

unsupervised clustering (ISODATA). These products are called GlobCover (GLOBCOVER, Products Description and Validation Report 2008). An example of this product is shown in Fig. 39.

The third example is made by the author's lab. It uses the same MODIS product with NASA product, i.e., NBAR, but the features used are very different from other two examples. Usually, there remain some clouds or snow in mosaicked images. In order to avoid these noises, we have developed a time domain co-occurrence matrix.



Surface reflectance (2007)

Nadir view surface reflectance (2007)

Fig. 40 Global land cover map using time domain co-occurrence matrix (Fukue et al. 2010)

This matrix is similar to the usual co-occurrence matrix, but the distance is not the space distance, but time difference is used as distance (Fukue et al. 2010). An example of this product is shown in Fig. 40. In this figure, upper image shows the result using MODIS surface reflectance and the bottom image shows the result of NBAR.

Geological Applications

Geological applications of remote sensing are one of the fastest application. Four kinds of applications are used in this field. The first application was to use satellite imagery in logistics. Usually, mineral exploration target areas are very large, and it is common that there are no large-scale base maps. In these circumstances, satellite images can be used as base maps. The second application has been the geological structures. Some of distinct geological structures, like anticlinal structures, circular structures etc., can be easily interpreted from the satellite images. Petroleum oils can only exist under anticlinal structure, and noble metals can be found along circular structures. Figure 41 shows an example of anticlinal structures observed by OPS on JERS-1. Other structures which can be interpreted from satellite images are lineaments. Some of these linear lines in the images are considered to express the underground fault structures. Figure 42 shows a sample of lineament extraction from a SAR image.

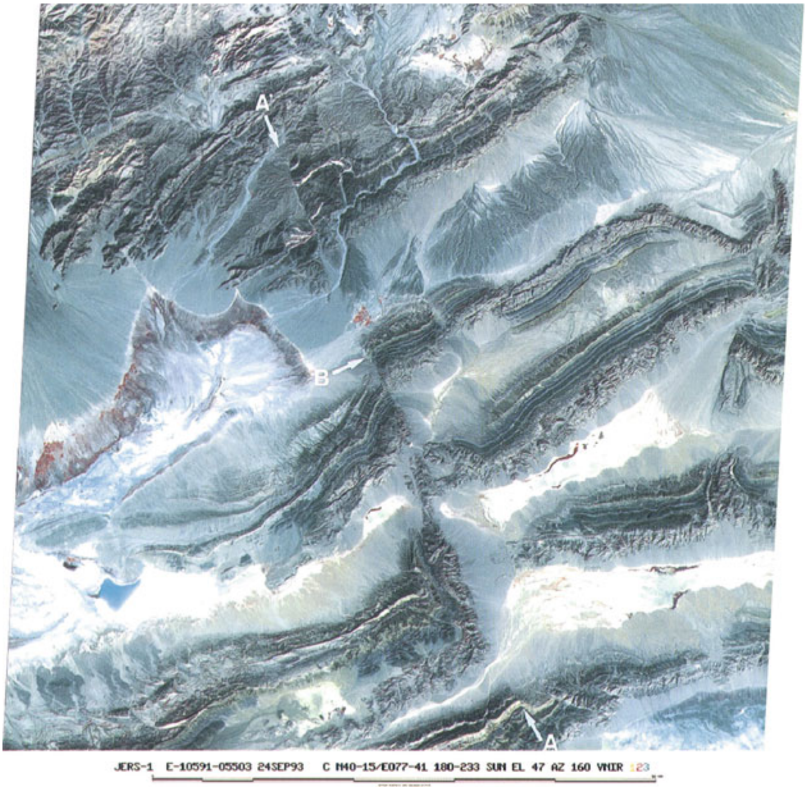


Fig. 41 An example of anticlinal structure observed by OPS on JERS-1

The third application is to directly detect rock types. For bare rock areas, rock types can be classified using their spectral signatures. Especially, metamorphic rocks have discriminative spectral signatures in short wave infrared region. Spectral features of mineral rocks and corresponding ASTER bands are shown in Fig. 43. Figure 44 shows the rock types extracted results from ASTER data. Now, most of petroleum fields over land are found. So, the effort to find a new petroleum field is concentrated over jungle and the ocean. For ocean explorations, remote sensing data are sometimes used to find out oil spills, which may be caused by ocean underground petroleum fields. SAR data are also used for geological applications. It is sometimes effective to find lineaments.

Soil Moisture

Soil moistures are not only important parameters to estimate thermal fluxes over land and to estimate evapotranspiration but also affect crop yields. Soil moistures

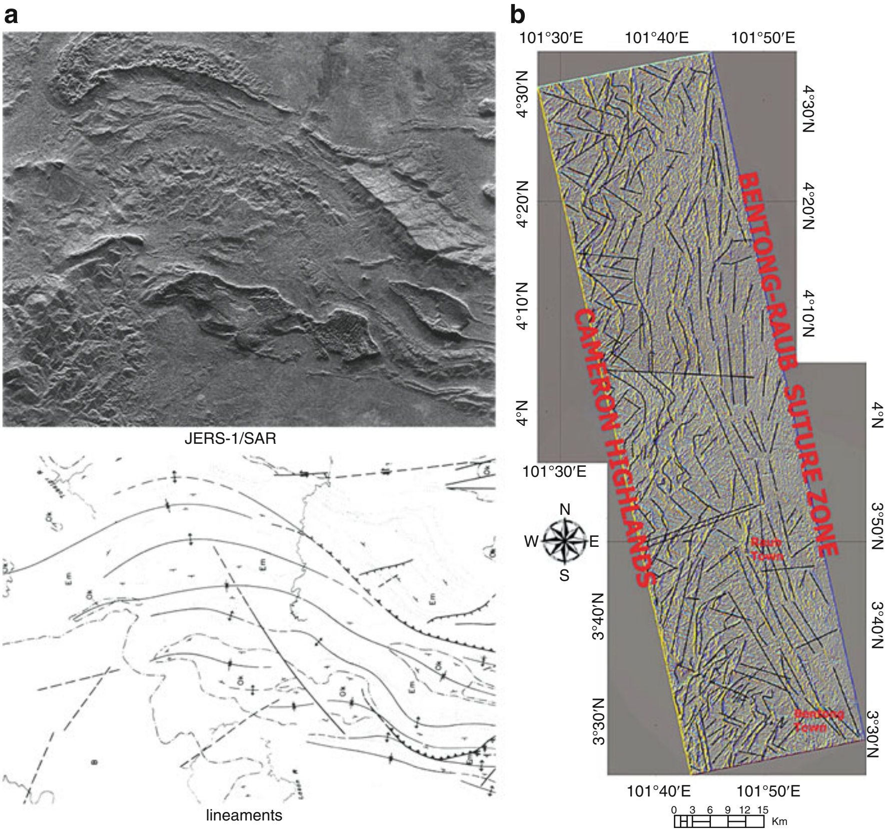


Fig. 42 Lineament extraction. (**a**) PALSAR image. (**b**) Extracted lineaments from the above image (Pour and Hashim 2015)

are usually retrieved from microwave radiometers. The sensitivity of microwave radiometer increases with the decrease of frequencies. Until recently, the lowest microwave radiometer frequency was C band. AMSR, AMSR-E, and AMSR2 were the only radiometers which have C band. The problem of C-band radiometer is that the spatial resolution is very low. With 2 m aperture antenna, the spatial resolution is around 50 km. This is a very wide area over land, and its validation is very difficult. Figure 45 shows changes of soil moisture of Africa between February and August obtained from AMSR-E. Very recently, much lower frequency radiometer was launched. It is SMOS and has L-band radiometer. Figure 46 shows an example of soil moisture over Australia obtained from SMOS.

Another approach is to use SAR for soil moisture retrievals. SAR has higher spatial resolution compared to microwave radiometers. Several attempts have been made, but

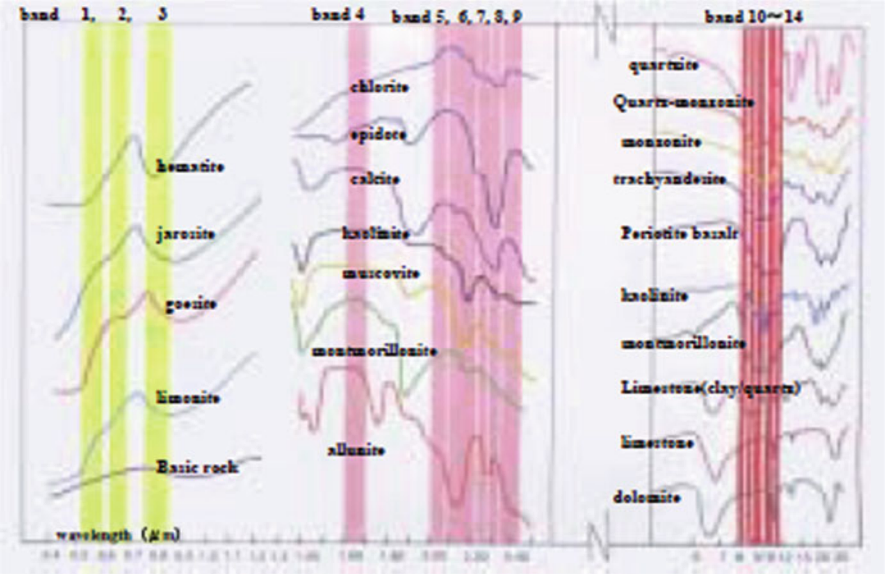


Fig. 43 Spectral signatures of mineral rocks and corresponding ASTER bands (ERSDAC 2003) (Courtesy of ERSDAC)

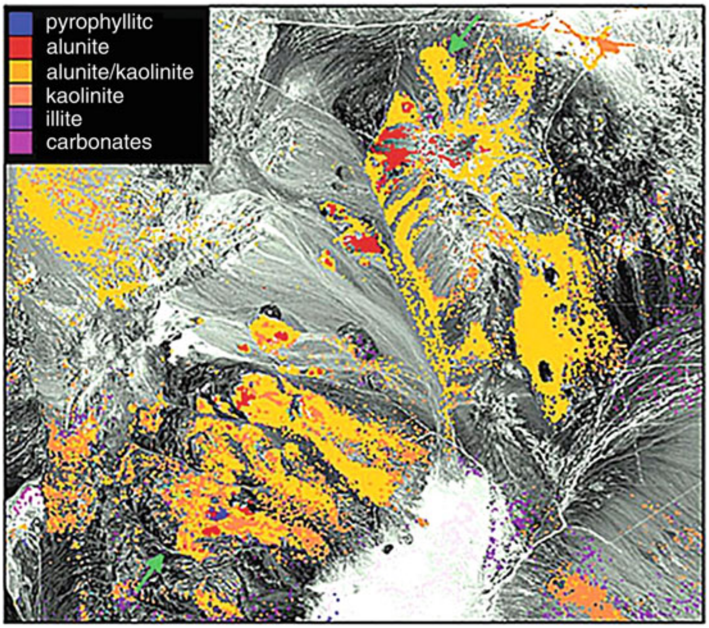


Fig. 44 Rock types extracted from ASTER data. *Left*: calcite; *Right*: mica (Perry and Kruse 2011) (Courtesy of ERSDAC)

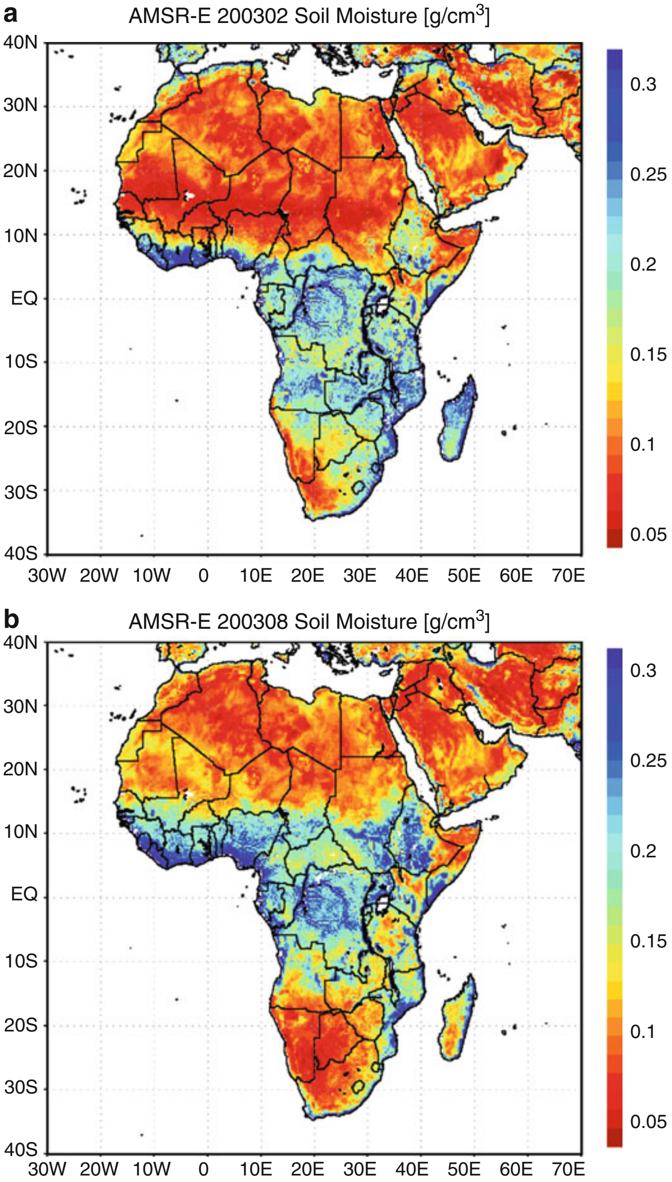


Fig. 45 Soil moisture of Africa obtained from AMSR-E. (**a**) February, 2003 (AMSR-E 200302 Soil Moisture 2004). (**b**) August, 2003 (AMSR-E 200308 Soil Moisture 2004) (Courtesy of JAXA)

the soil moisture retrievals are also difficult because of the sensitivity change associated with incidence angle change. Accuracies of the retrieved soil moistures from C-band radiometers and SAR are not sufficiently good. SMOS and Aquarius may provide higher accuracy soil moistures after sufficient validation activities.

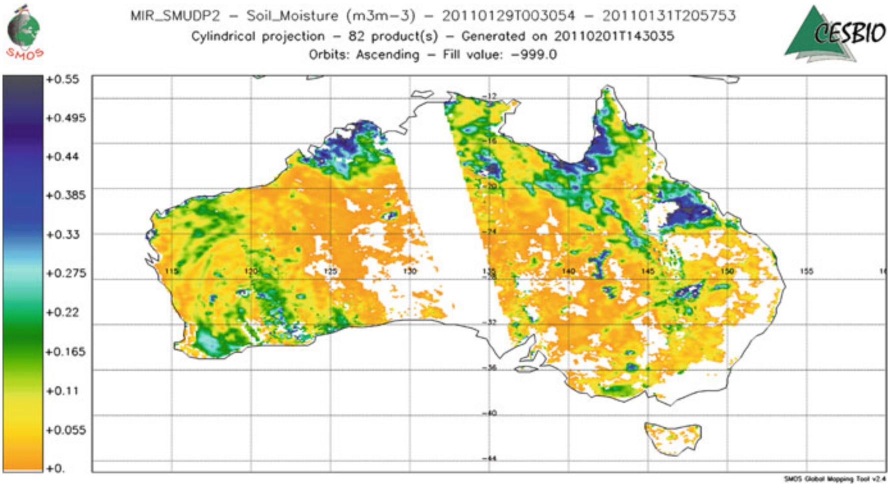


Fig. 46 Soil moisture over Australia obtained from SMOS during January 29–31, 2011 (Australia and Yasi 2011) (Courtesy of CNES)

Carbon Cycle

In order to accurately project the global warming, understandings of carbon cycle are very important. Land vegetation absorbs carbon dioxide, but the quantity of these absorptions is not clearly understood. The carbon flux of vegetation should be clarified, but it is rather difficult from satellite data. The first step is to estimate terrestrial primary productions of vegetation from satellite data. The gross primary production (GPP), which is the fixed amount of carbon by photosynthesis can be expressed as follows (Monteith 1972; Running et al. 2000; Nemani et al. 1982; Running et al. 2004):

$$GPP = \varepsilon \times fAPAR \times PAR$$
$$\varepsilon = \varepsilon_{\max} \times T_f \times VPD_f$$

Here,

ε: Light use efficiency parameter (gCMJ^{-1})
PAR: Photosynthetically active radiation
$fAPAR$: Fraction of absorbed PAR
$\varepsilon_{\max}$: Potential under optimal conditions (no environmental stress)
T_f: Reductions in photosynthesis under low temperature condition
VPD_f: Reductions in photosynthesis under suboptimal surface air vapor pressure deficit

PAR can be derived from satellite data, and *fAPAR* has correlation with satellite-derived LAI (leaf area index, calculated from NDVI or EVI) or NDVI or EVI. EVI is expressed as follows:

$$EVI = G \times \frac{NIR - \mathrm{Red}}{NIR + C_1 \cdot \mathrm{Red} - C_2 \cdot \mathrm{Blue} + L}$$

Here, G, C_1, C_2, and L are empirically defined coefficients.

From *GPP*, *NPP* (net primary production) is derived as follows:

$$NPP \ = \ GPP - R$$

Here, R is aboveground respirations and can be determined from LAI and temperature. Figure 47 shows an example of global NPP derived from MODIS data.

Another approach to derive GPP is to use fluorescence from chlorophyll. Chlorophyll fluorescence spectra show rather wide spectral features over 700 nm region. However, there are several Fraunhofer lines in these spectral areas, and chlorophyll fluorescence makes these lines shallower. From these features, one can retrieve sun-induced chlorophyll fluorescence indices (SIF) (Frankenberg et al. 2011a). These fluorescence spectra result from photosynthetic reaction of vegetation, hence GPP correlates with SIF (Frankenberg et al. 2011b; Guanter et al. 2012; Frankenberg et al. 2014). Figure 48 shows an example of global SIF (Frankenberg et al. 2011a).

The carbon flux is expressed by net ecosystem production (NEP). NEP is calculated from NPP by subtracting under the ground respirations. However, the under the ground respirations cannot be retrieved from satellite data. In order to estimate NEPs, carbon cycle model is necessary. Many kinds of terrestrial carbon cycle model are proposed (Running Gower 1991; Esser et al. 1994; Foley et al. 1996; Tian et al. 1999). Another approach combines ground-based carbon flux estimation with atmosphere-based carbon concentrations. As described in the

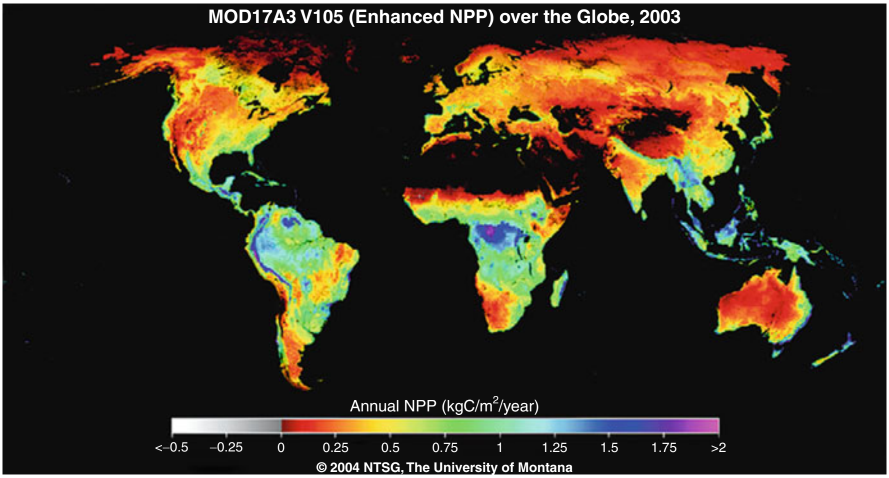


Fig. 47 Global NPP derived from MODIS (Zhao et al. 2005) (Courtesy of NASA)

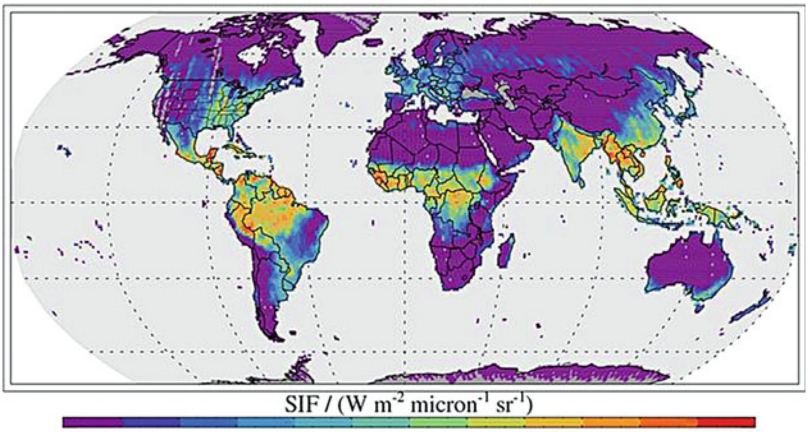


Fig. 48 Global SIF derived from GOSAT (Frankenberg et al. 2011a)

section on "Greenhouse gases" in this chapter, satellite sensors can now achieve good accuracy readings for both atmospheric carbon dioxide and methane gases. Therefore, models which can describe both atmospheric concentrations and ground-based fluxes with assimilation capability may result better accuracy carbon cycle understandings.

Cryospheric Applications

Sea Ice

Sea ice plays an important role for energy circulations of the Earth. Sensible heat and latent heat over sea ice are very different from those over open ocean. It is also important for ship navigations over high latitude areas. Sea ice has been monitored using microwave radiometers for a long time. The geophysical parameter which is retrieved from microwave radiometer is ice concentrations. Ice concentration is the ratio of sea ice covered area to the total area. Figures 49 and 50 show Arctic sea ice concentrations trend retrieved from AMSR-E and AMSR-2.

Microwave scatterometer also can monitor sea ice. However, parameters which can be retrieved from microwave scatterometer are different from radiometers. From scatterometer, areas of sea ice and the discrimination between multiyear ice and new ice can be obtained.

There are several other parameters which are important for monitoring sea ice. One is the thickness of sea ice, but it is very difficult to retrieve sea ice thickness from satellite data. Another parameter which is important is the thickness of thin ice. When the sea ice is very thin, i.e., less than 1 m, the energy flux between sea and

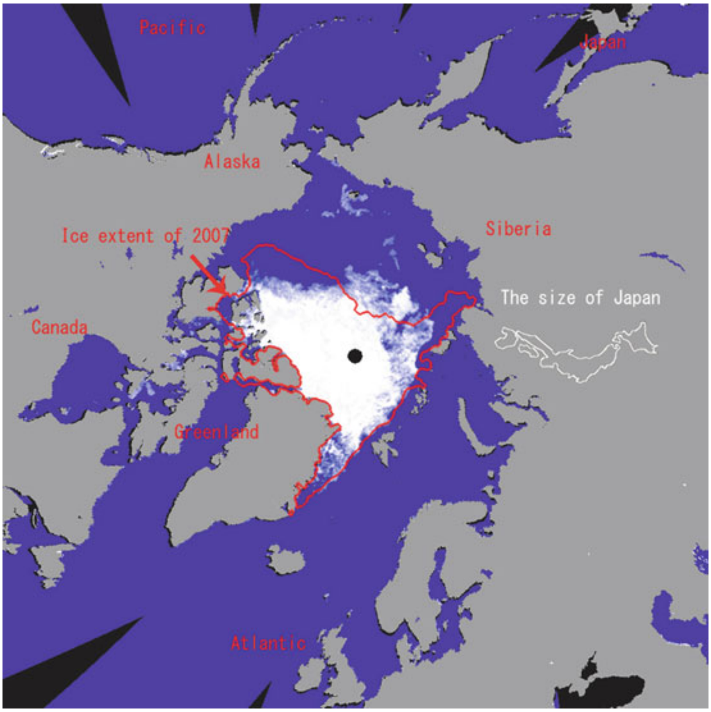


Fig. 49 Arctic sea ice in 2012 summer compared to that of 2007 (Arctic sea ice trend 2012) (Courtesy of JAXA)

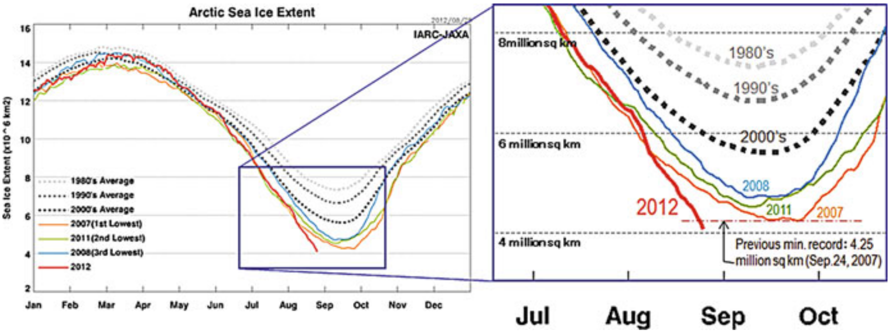


Fig. 50 Arctic sea ice trend from 1980s to 2012 (Arctic sea ice trend 2012) (Courtesy of JAXA)

atmosphere changes drastically according to the thickness. Many attempts have been done to retrieve this parameter, but still accurate algorithms are not present. Another application of sea ice is the monitoring of icebergs. Microwave scatterometer is now used for this purpose as well.

Snow and Glaciers

Snow also perturbs global climate. Snow is also very important for water supply. Several geophysical parameters are important, i.e., snow cover, snow depth, and snow albedo. Snow cover and albedo can be retrieved from optical sensors, while snow depth can be retrieved from microwave radiometer. Figure 51 shows an example of global snow cover from MODIS data, and Fig. 52 shows the global snow depth retrieved from AMSR-E data.

Glaciers are affected by global warming. Many of the existing glaciers are retreating. It is not clear that these retreats are caused by global warming or not. Anyway, it is important to monitor the motion of these glaciers. The forefront of glaciers can be monitored using optical sensors. Another way of monitoring glaciers is to use SAR interferometry (Fatland and Lingle1998; Mohr et al. 1998; Joughin et al. 1998; Rabus and Fatland 2000). Using the SAR interferometry, the retreating speed of glaciers can be obtained.

Operational Applications

NWP and Weather Forecasting

The weather forecasting of developed countries is based on the results of numerical weather projection (NWP) software. Until around 15 years ago, these NWPs used only in situ data for the input. However, these NWPs now use many kinds of satellite data in addition to the in situ data. Most popular data used as input by NWPs are microwave sounder data, thermal IR sounder data, microwave radiometer data, microwave scatterometer data, microwave altimeter data, GPS occultation data,

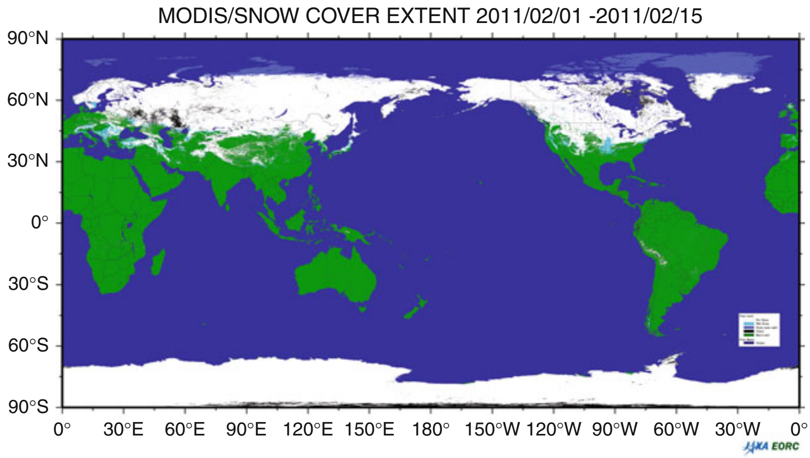


Fig. 51 An example of global snow cover from MODIS data (MODIS Snow Cover Extent 2015) (Courtesy of JAXA)

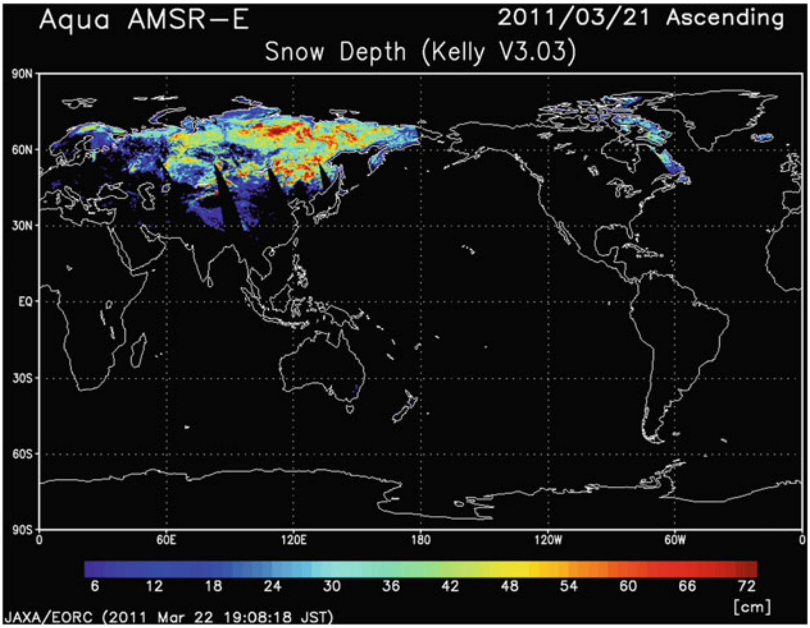


Fig. 52 Global snow depth (shown in cm) retrieved from AMSR-E data (Aqua AMSR-E Snow Depth 2011)

etc. Another very important data for NWP are atmospheric winds obtained from geostationary satellite. From visible and thermal infrared images, motion of clouds is extracted and the winds near the clouds are retrieved. For clear areas, water vapor motion extracted from mid IR is used for winds retrieval. Geostationary satellite imagers can cover only within ±50° latitudinal areas. For higher latitudes, winds extracted from MODIS are now used.

At the first stage of the NWP applications, retrieved geophysical parameters were used for assimilations. However, for thermal IR and microwave sounder and radiometer data, radiances from these sensors are now directly assimilated to the NWP. The impact of satellite data to the NWP is positive, but the extent how much improvements are made is not so clear, because NWP models themselves have been improved. Figures 53 and 54 show impacts of satellite data to NWP estimated by ECMWF. In Fig. 53, baseline is NWP without any satellite data, AMV is NWP with satellite-derived atmospheric winds data, EUCOS is NWP with AMV plus AMSU data, and control is NWP with all satellite data. From these figures, impacts are larger in the Southern Hemisphere than northern hemisphere. In the Northern Hemisphere, satellite data impacts are around 1.6 days at the 6 day forecast, while it is and two and a half days for extended forecasts. From Fig. 54, it is shown that satellite winds, water vapor, optical sounder, and microwave scatterometer have large impacts to NWP.

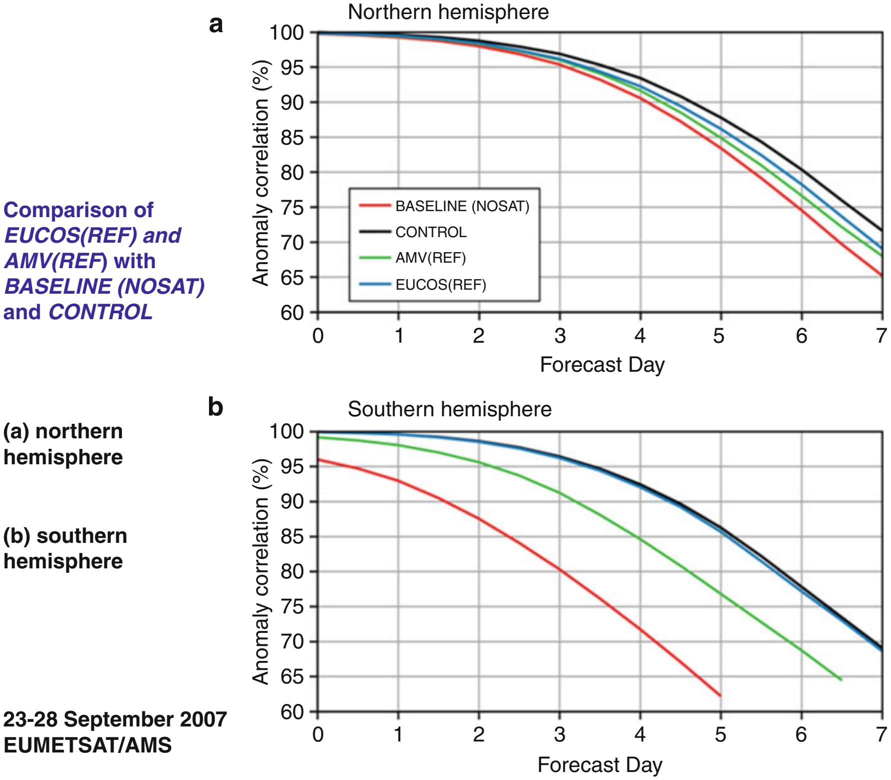


Fig. 53 Impacts of satellite data to NWP for several sensor combinations estimated by ECMWF (Kelly 2007)

Fisheries

Fisheries are one of the largest operational application areas of remote sensing. Satellite data applications to fisheries have begun from early 1980s. At the first stage, SST was used. The accuracy of satellite-derived SST is not enough to directly find out fishing grounds, but from the SST imagery, it is very easy to detect SST front and positions of oceanic current. Especially, some of the oceanic fronts are good fishing grounds.

Another application has been chlorophyll-a. Chlorophyll-a concentrations correspond to phytoplankton concentrations, which further correlate to zooplankton concentrations. Nowadays, fisher men use many other satellite sensor data, e.g., microwave altimeter data, microwave scatterometer data, and microwave radiometer data for finding good fishing grounds.

Figure 55a shows SST distributions and fishing grounds. From this figure, fishing grounds lie along the front of SST. Figure 55b shows the chlorophyll-a distribution and fishing grounds. From this figure, it is shown that fishing grounds lie near high chlorophyll-a concentration areas.

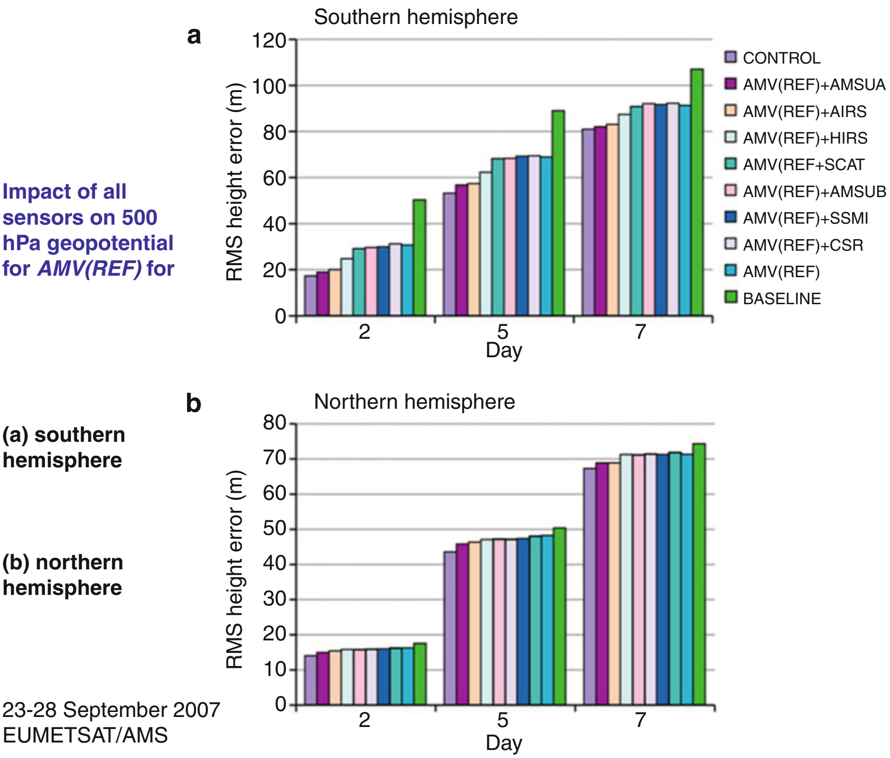


Fig. 54 Impacts of satellite data to NWP for each sensor estimated by ECMWF (Kelly 2007)

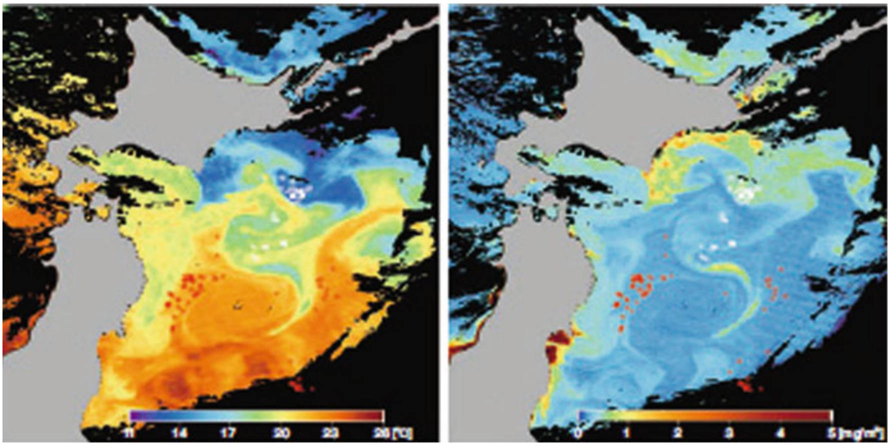


Fig. 55 SST (**a**) and chlorophyll-a (**b**) distribution and fishing grounds of Sanriku Coast, Japan observed by GLI on ADEOS2 and fishing boats information (SST and chlorophyll-a 2003) (Courtesy of JAXA)

This figure presents GLI sea surface temperature and chlorophyll-a concentration images overlaid on fisheries of skipjack and tuna. Fisheries of warmwater skipjack were distributed in areas of relatively high sea surface temperature and low chlorophyll-a concentration. Also, saury is in relatively low sea surface temperature and high chlorophyll-a concentration.

Disasters

Earthquake

An application area of SAR data is the detection of ground movement. Using differential interferometric SAR, the displacement of the ground can be detected with cm order accuracy. Detected land displacement is used for earthquake studies as well as volcano monitoring. Figure 56 shows the land displacement by the 2011 off the Pacific coast of Tohoku Earthquake observed by PALSAR on ALOS.

Biomass Burnings

The number of total biomass burnings in the world is around 1 million times. From these biomass burnings, many kinds of atmospheric constituents are emitted. These gases include, but not limited to, CO_2, CO, CH_4, NO, NH_3, and O_3. The total CO_2 emission amount varies depending upon each year, and rather difficult to estimate, but may range from 2 Gton carbon to 4 Gton carbon. However, as the regrown vegetation absorb CO_2, the net emissions will be much smaller (Levine et al. 1995; Jacobson 2004). It is also one of the largest sources of tropospheric ozone; thus degrading the quality of atmosphere. Large-scale biomass burnings also takes the lives of people and burns household articles.

In order to monitor the global biomass burnings, satellite monitoring is the only mean. Many kinds of satellite sensors are used for this purpose. The most used sensors are AVHRR on NOAA, MODIS on Terra and Aqua, and imagers onboard geostationary meteorological satellites. Infrared channels are used for fire detection. However, 11 and 12 μm regions are saturated quite quickly, so 3.7 μm or shorter wavelength is more appropriate. Global fire maps can be accessed through MODIS Rapid Response System Global Fire Maps (MODIS Rapid Response System Global Fire Maps 2016) of NASA or Current & Archived Significant Global Fire Events and Fire Season Summaries (Global Fire Map 2016) of International Strategy for Disaster Reduction (ISDR), etc. Figure 57 shows a global fire map of 03/21/16–03/30/16 distributed by the above NASA site. Satellite-derived fire monitoring has some disadvantages. One of the problems is that the spatial resolution of infrared sensors are usually not so fine, hence, burnt areas are overestimated. Another disadvantage is that it cannot monitor under thick clouds.

Floods

Flood is one of the most frequent disasters over the world. Figure 58 shows the percentages of disaster events by each category between 2000 and 2008 from two

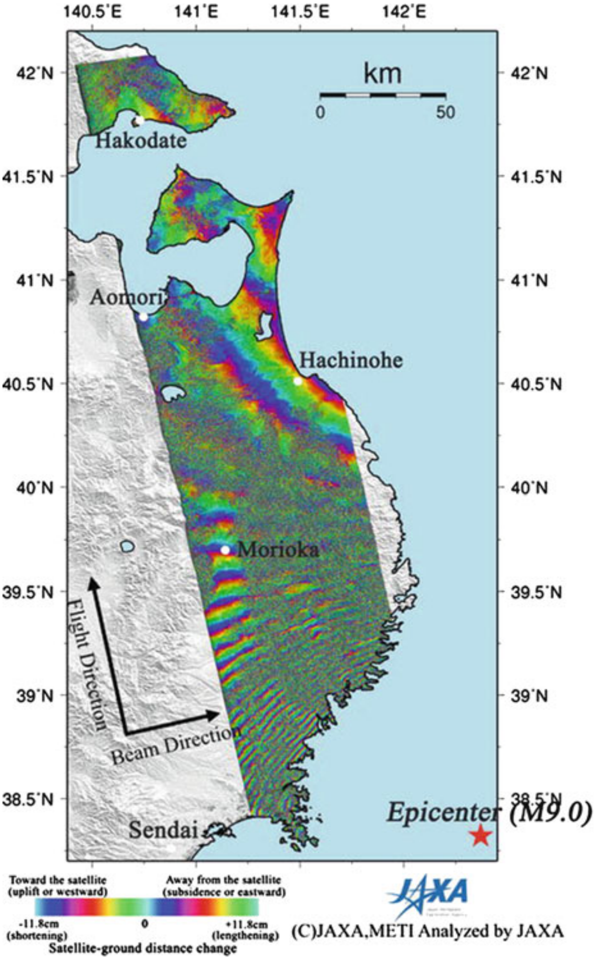


Fig. 56 The land displacement by the 2011 off the Pacific coast of Tohoku Earthquake observed by PALSAR on ALOS (PALSAR on ALOS 2011) (Courtesy of JAXA)

international disaster databases, EM-DAT (EM-DAT, The International Disaster Database 2016) and NatCatSERVICE (Munich RE, NETCATSERVICE 2015). There are some differences between these two databases because of the difference of events registrations, but anyway, floods share highest or second highest disaster of natural disasters. Remote sensing has been used to estimate inundated areas by floods. This is done by comparing two images taken before the flood and after the flood. Both optical sensors and SARs are used for this purpose. Figure 59 shows flood areas caused by a cyclone over Myanmar taken by PALSAR on ALOS. Figure 60 shows a flood over northeastern China taken by GLI on ADEOS2. Optical sensors can detect inundated areas clearly, but it cannot take images under cloudy conditions. SAR can take images in any conditions, but sometimes it is difficult to extract inundated areas.

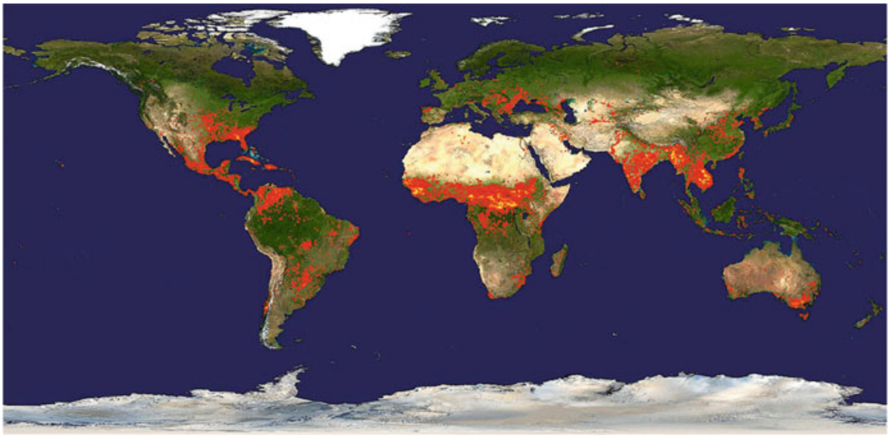


Fig. 57 Global fire map of 03/21/16–03/30/16 (MODIS Rapid Response System Global Fire Maps 2016) (Courtesy of NASA)

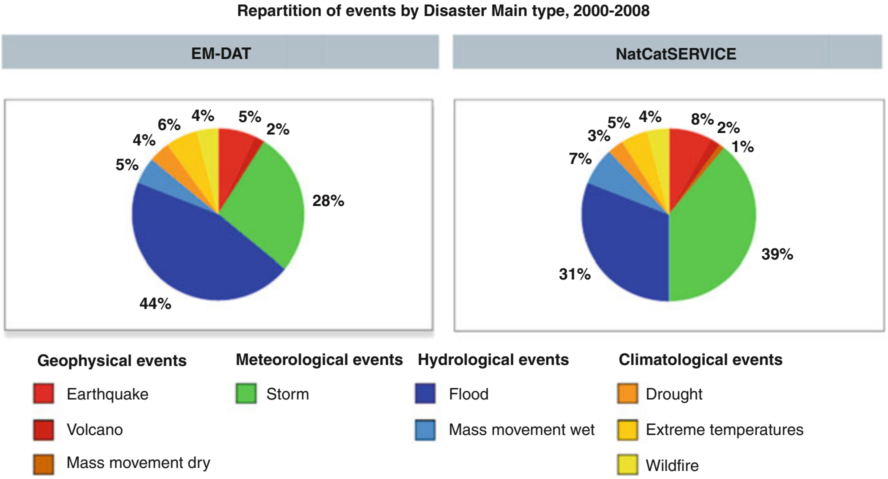


Fig. 58 Events by natural disasters main types, 2000–2008 (Below 2009)

Ship Navigation

Ship routing and navigation in Arctic sea areas was one of the earliest operational applications of SAR. One of the disadvantages of SAR for near real-time applications is the frequency of the observations. However, in high latitude regions, i.e., higher than 68°, scan mode SAR can observe any areas in this region once a day. C-band SAR is thought to be most useful for this application, and there are now more

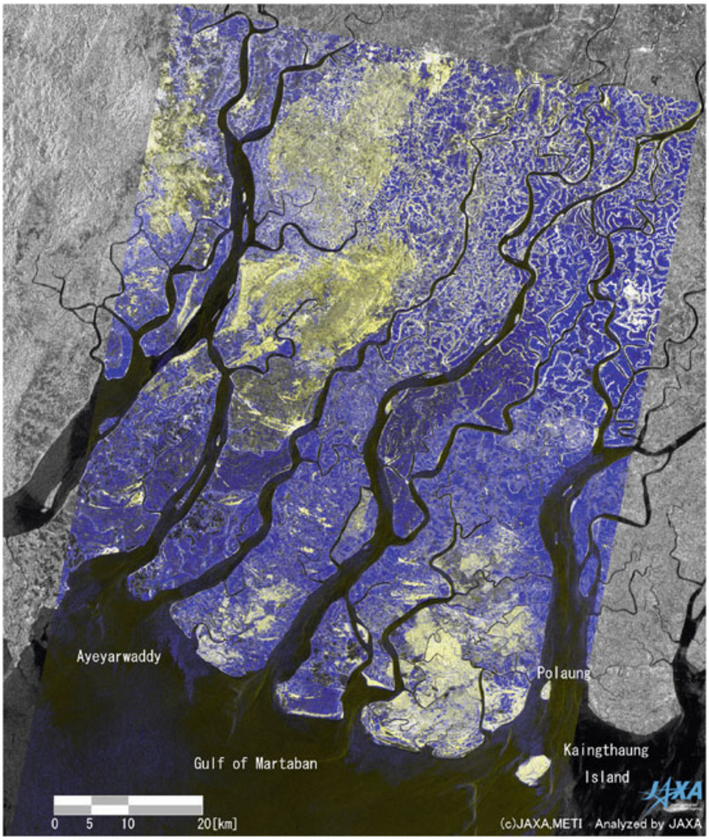


Fig. 59 Images of Ayeyarwaddy, Myanmar, observed by PALSAR on April 24 and May 6, 2008 (Myanmar flood water observation by PALSAR 2008) (Courtesy of JAXA)

than four sensors operating. C-band SAR can distinguish multiyear ice, first year ice, landfast ice, thin ice, leads/polyniyas, and areas of ridges.

Figure 61 shows an example of ship routing map generated by Kongsberg.

Agriculture

Agriculture is one of the largest application areas of remote sensing. There are several applications for agriculture, but largest applications are crop acreage estimation and crop yield estimation. Crop acreage estimation started in USA from 1970s. National Agricultural Statistics Service (NASS) of USDA has started state level crop acreage estimation using Landsat imagery from 1978 (Bailey and Boryan 2010). Now, many countries are using remote sensing for crop acreage estimation.

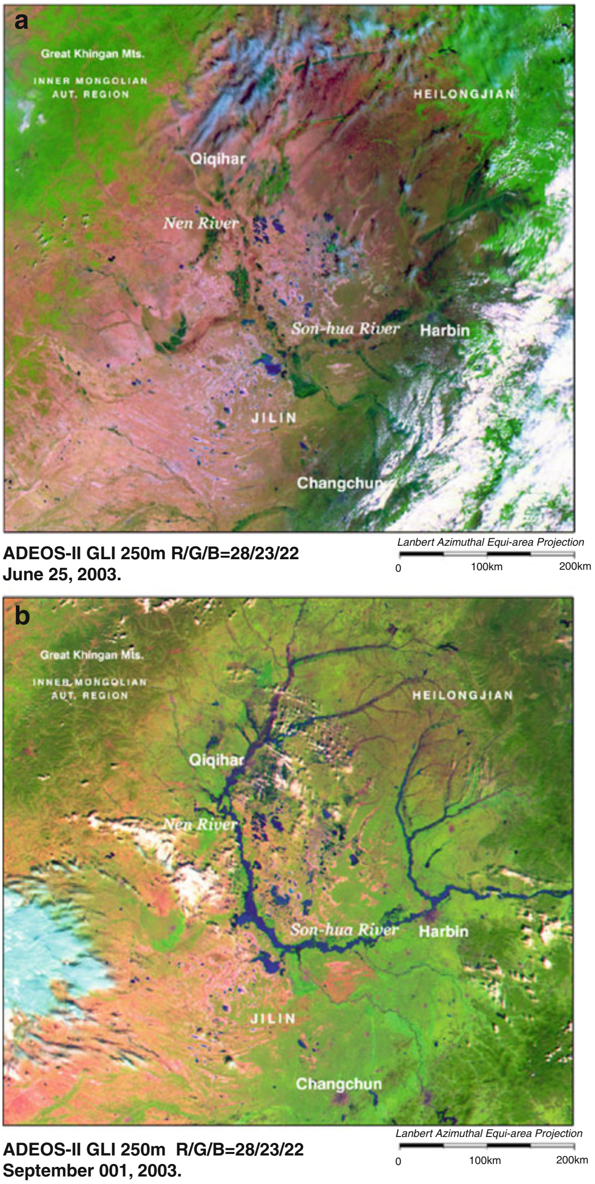


Fig. 60 GLI captured the conditions before and during the flooding in Northeastern China that continued from July to October 2003. (**a**) Before the flood. (**b**) After the flood (Northeastern China Suffers Large-Scale Flooding 2003) (Courtesy of JAXA)

However, there are several problems of using remote sensing for crop acreage estimation on a global scale. The first problem comes from the spatial resolution. For countries like USA or Canada, the dimensions of each crop field is very large, hence spatial resolution of 30 m of Landsat TM is sufficient for these countries.

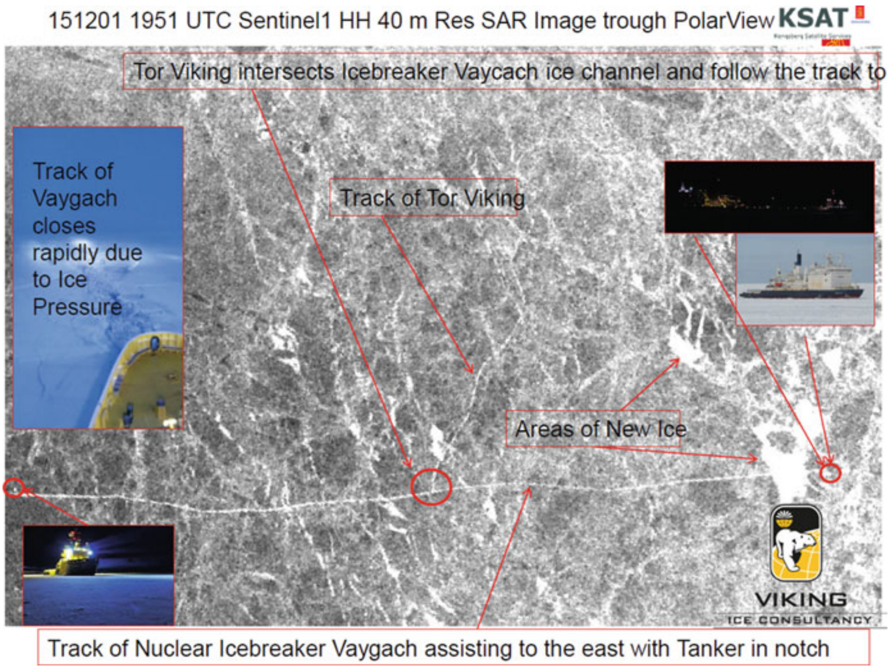


Fig. 61 Ship routing over Arctic region using SAR (Larsen 2016)

However, for Japan and most south East Asian countries, these dimensions are very small, and 30 m resolution cannot discriminate each crop field.

Second problem is the cloud cover. Optical sensors cannot observe under clouds. This problem is most typical for rice fields, most of which resides in Monsoon areas. SAR can penetrate clouds, but its ability to discriminate crop species is very low.

The third problem is the timing which NASS emphasizes (Bailey and Boryan 2010). The discriminability of remote sensing to crop species is highest when crop grows sufficiently, but most statistics needs acreage estimation in earlier stages. The accuracy of crop acreage estimation is rather difficult to estimate. Workshop on best practices for crop area estimation with remote sensing data (Best Practices for Crop Area Estimation with Remote Sensing Data 2007) was held in 2008 under GEO, and each country reports the accuracy of their crop acreage estimate by remote sensing. The accuracies range from 60 % to 95 %, but the real best accuracy will be in the range of around 85 %.

Figure 62 shows a part of land cover map of State of Illinois using Landsat TM data generated by a NASS project, and Fig. 63 shows the results of validation. From Fig. 63, it is shown that high classification accuracies are obtained for some crops (e.g., corn and soybeans are around 98 %), while classification accuracies are low for other crops (e.g., rice, barley, rye, oats are around 50 %). Use of DMC satellite which can provide 30 m resolution images every day may improve these problems.

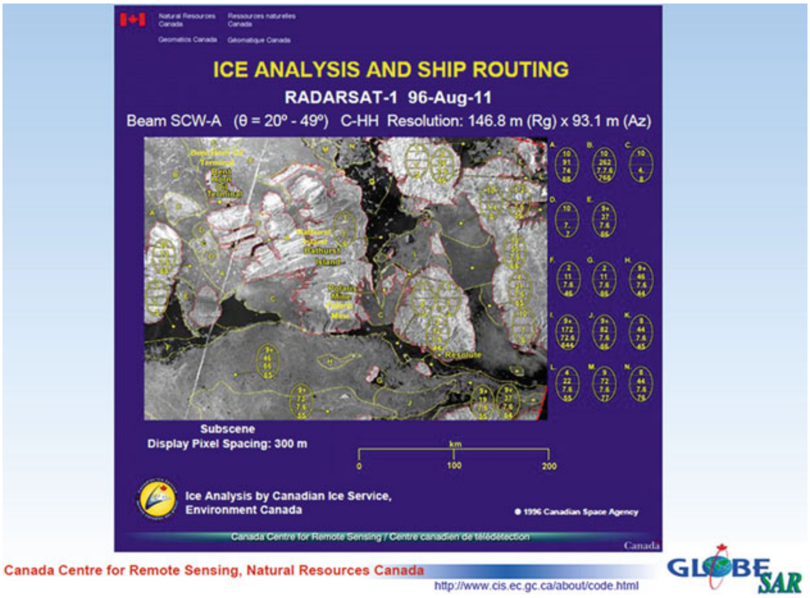


Fig. 62 A part of land cover map of State of Illinois using Landsat TM data generated by a NASS project (Luman and Tweddale 2008)

The second application area is the crop yield estimation. Usually, crop yield estimation is done using regression models with climate variables, like temperature, precipitation, etc. In addition to these variables, addition of parameters derived from remotely sensed data, e.g., NDVI, EVI, and LAI usually gives better results. Only one problem is the timing of remote sensing data acquisition. For instance, in order to estimate rice yield, there are three timings each of which has only 1 week duration. These timings also depend on the kinds of rice and the areas of rice fields. So, it is very difficult to acquire appropriate remote sensing images which can be used for yield estimations.

Conclusion

There are many other application areas, e.g., urban planning, archeology, and water resources, which are not described in this chapter. However, the application areas of remote sensing are spreading rapidly thanks to the new sensors as well as many remote sensing satellites. For global change monitoring, a long-term record is necessary. There are some long time records starting from 1960s (NOAA satellites) and 1970s (microwave radiometer), but most of the sensors for this purpose started from the end of 1990s, and still need further continuous

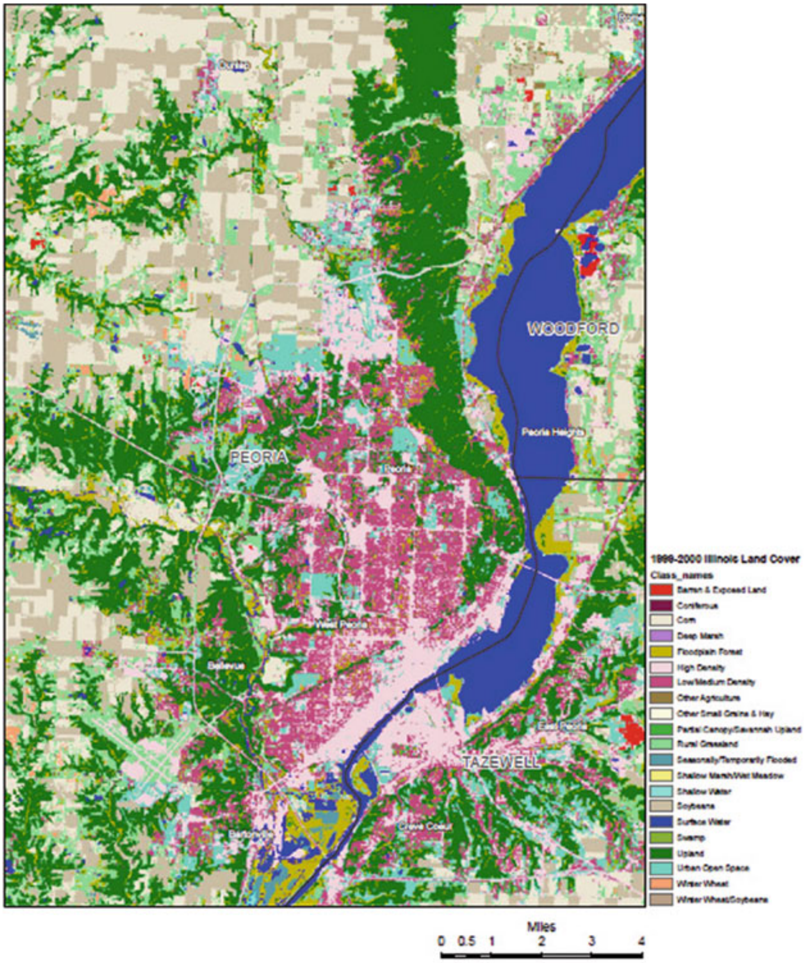


Fig. 63 Results of validation of the total Illinois land cover (Luman and Tweddale 2008)

monitoring. For local applications, high spatial resolution sensors made new applications. Problems in this field are cost of image acquisition and frequency of observations.

Cross-References

- ▶ Fundamentals of Remote Sensing Imaging and Preliminary Analysis
- ▶ Introduction and History of Space Remote Sensing
- ▶ Lidar Remote Sensing
- ▶ Processing and Applications of Remotely Sensed Data

References

3D view inside an extra-tropical cyclone observed off the coast of Japan, March 10, 2014, by GPM's Dual-frequency Precipitation Radar (2014), http://www.nasa.gov/sites/default/files/fig1_rgb_0.png

ADEOS EarthView, Three-Dimensional Distribution of Atmospheric Ozone (1998), http://suzaku.eorc.jaxa.jp/GLI2/adeos/Earth_View/jap/adeos16j.pdf

AEROSOLS RESULTS OVER LAND (1997), https://polder-mission.cnes.fr/en/POLDER/SCIEPROD/ae_val_res_ls.htm, http://smsc.cnes.fr/POLDER/A_produits_scie.htm

AMSR/AMSR-E SST algorithm (2002), http://sharaku.eorc.jaxa.jp/AMSR/doc/alg/9_alg.pdf

AMSR-2 Weekly Image (2014), ftp://suzaku.eorc.jaxa.jp/pub/AMSR/usr/amsrftp/.sample/SST/

AMSR-E 200302 Soil Moisture (2004), http://www.eorc.jaxa.jp/earthview/2004/img/tp040723_01.gif

AMSR-E 200308 Soil Moisture (2004) http://www.eorc.jaxa.jp/earthview/2004/img/tp040723_02.gif

J.R. Anderson, E.E. Hardy, J.T. Roach, R.E. Witmer, A land use and land cover classification system for use with remote sensor data. Geological Survey Professional Paper 964 (1976)

D. Antoine, A. Morel, Relative importance of multiple scattering by air molecules and aerosols in forming the atmospheric path radiance in the visible and near-infrared parts of the spectrum. Appl. Opt. **37**, 2245–2259 (1998)

Aqua AMSR-E Snow Depth (2011), http://sharaku.eorc.jaxa.jp/AMSR/L3_browse/PM/geo/201103/21/P1AME110321A_P3SWEKel303E0.png

Aquarius L3 Image Browser (2015), http://podaac.jpl.nasa.gov/aquarius/gallery

Arctic sea ice trend (2012), http://www.eorc.jaxa.jp/en/earthview/2012/tp120825.html

I. Asanuma, Depth and time resolved primary productivity model examined for optical properties of water, in *Global Climate Change and Response of Carbon Cycle in the Equatorial Pacific and Indian Oeeans and Adjacent Landmasses*. Elsevier Oceanogrhy Series (Elsevier, Amsterdam, 2006), pp. 89–106

Australia and Yasi… New floods? … what is SMOS seeing to help forecasts? (2011), http://smsc.cnes.fr/SMOS/GP_actualites.htm

J.T. Bailey, C.G. Boryan, Remote Sensing Applications in Agruculture at the USDA National Agricultural Statistics Service. (2010), http://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/ICAS5/PDF/ICASV_2.1_048_Paper_Bailey.pdf

M.J. Behrenfeld, E. Boss, D.A. Siegel, D.M. Shea, Carbon-based ocean productivity and phytoplankton physiology from space. Global Biogeochem. Cycles **19**, GB1006, 1–14 (2005)

R. Below, A. Wirtz, D. Guha-Sapir, Disaster Category Classification and peril Terminology for Operational Purposes (2009). p. 9 http://cred.be/sites/default/files/DisCatClass_264.pdf

P.G. Berenfeld, M.J. Falkowski, Photosynthetic rates derived from satellite-based chlorophyll concentration. Limnol. Oceanogra. **42**, 1–20 (1997)

A. Berk, S.L. Bernstein, G.P. Anderson et al., MODTRAN cloud and multiple scattering upgrades with application to AVIRIS. Remote Sensing of Environment **65**, 367–375 (1998)

Best Practices for Crop Area Estimation with Remote Sensing Data (2007), http://www.earthobservations.org/documents/cop/ag_gams/200707_01/Summary_countries_AG.pdf

T. Blaschke, G.J. Hay, Object-oriented image analysis and scale-space: theory and methods for modeling and evaluating multi-scale landscape structures. Int. Arch. Photogram. Remote Sensing **34**(4/W5), 22–29 (2001)

G. Bo-Cai, M.J. Montes, L. Rong-Rong, H.M. Dierssen, C.O. Davis, An atmospheric correction algorithm for remote sensing of bright coastal waters using MODIS land and ocean channels in the solar spectral region. TGARS **45**, 1835–1843 (2007)

I. Buehler, S.A.P. Eriksson, T. Kuhn, A. von Engeln, C. Verdes, ARTS, the atmospheric radiative transfer simulator. J. Quant. Spectrosc. Radiat. Transfer **91**, 65–93 (2005)

Carbon Monoxide SCIAMACHY/ENVISAT 2004 (2004), http://www.iup.uni-bremen.de/sciamachy/NIR_NADIR_WFM_DOAS/SCIA_CO_glo_2004.png

M.-E. Carr, M.A. Friedrichs et al., A comparison of global estimates of marine primary production from ocean color. Deep Sea Res **53**, 741–770 (2006)

P.S. Chavez Jr., An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sensing Environ **24**, 459–479 (1988)

R.M. Chomkoa, H.R. Gordon, S. Maritorenab, D.A. Siegel, Simultaneous retrieval of oceanic and atmospheric parameters for ocean color imagery by spectral optimization: a validation. Remote Sensing Environ **84**, 208–220 (2003)

Cloud vertical distribution from Cloudsat (2010), http://cloudsat.atmos.colostate.edu/2010tcs/20100915julia.jpg

S.A. Clough et al., Atmospheric radiative transfer modeling: a summary of the AER codes. J. Quant. Spectrosc. Radiat. Transfer **91**, 233–244 (2005)

CO_2 flux (2012), https://data.gosat.nies.go.jp/GosatL4Image/dist/browse/L4A/Flux_Map/LatLon/Monthly/CO2/V02.03.000/co2.flux_map.latlon.region.big2.201207.000.v0203.gif

D.P. Craig, T. Thirunamachandran, Third-body mediation of resonance coupling between identical molecules. Chem. Phys. **135**, 37–48 (1989)

P.Y. Deschamps, M. Herman, D. Tanre, Modeling of the atmospheric effects and its application to the remote sensing of ocean color. Appl. Opt. **22**, 3751–3758 (1983)

Earth Observatory, Absorption Bands and Atmospheric Windows (1999), http://earthobservatory.nasa.gov/Features/RemoteSensing/remote_04.php

Earth Probe TOMS Data & Images (1997), http://toms.gsfc.nasa.gov/eptoms/ep_v8.html

El Niño phenomenon being close to the strongest on record (2015), http://www.eorc.jaxa.jp/en/earthview/2015/tp151130.html

EM-DAT, The International Disaster Database (2016), http://www.emdat.be/

EP/TOMS Version 8 Monthly Average Aerosol Index (1996), ftp://toms.gsfc.nasa.gov/pub/eptoms/images/monthly_averages/aerosol/IM_aiavg_ept_199608.png

ERSDAC. ASTER LEVEL 4A01 DATA PRODUCTS SPECIFICATION (GDS Version) Version 1.1 (ERSDAC Tokyo, 2002)

ERSDAC. *ASTER Reference Guide (Version 1.0)* (ERSDAC, Tokyo, 2003)

G. Esser, J. Hoffstadt, F. Mack, U. Wittenberg, High-resolution biosphere model (HRBM)- Documentation model version 3.00.00, in *Mitteilungen aus dem Institut fur Pflanzenokolgie der Justus-Liebig-Universitat Giessen*, ed. by G. Esser, vol 2 (Universitat Giessen, Giessen, 1994)

FAO, FRA 2000 Forest Cover Mapping & Monitoring with NOAA-AVHRR & Other Coarse Spatial Resolution Sensors. *Forest Resources Assessment Programme Working Paper 29* (FAO, 2000)

FAO, Land Cover Classification System (LCCS)." 2000. *FAO Corporate Document Repository.* http://www.fao.org/docrep/003/x0596e/X0596e01i.htm#TopOfPage, http://www.fao.org/docrep/003/x0596e/X0596e01i.htm#TopOfPage

D.R. Fatland, C.S. Lingle, Analysis of the 1993–95 Bering Glacier (Alaska) surge using differential SAR interferometry. J. Glaciol. **4**, 532–546 (1998)

J.A. Foley, I.C. Prentice, N. Ramankutty, S. Levis, D. Pollard, S. Sitch, A. Haxeltine, An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. Global Biogeochem Cycles **10**, 603–628 (1996)

C. Frankenberg, A. Butz, G.C. Toon, Disentangling chlorophyll fluorescence from atmospheric scattering effects in O_2 A-band spectra of reflected sun-light. GRL **31**, (2011), Doi 10.1029/2010GL045896

C. Frankenberg, J.B. Fisher, et al., New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. GRL **31**, (2011), Doi 10.1029/2010GL045896

C. Frankenberg, C. O'Dell et al., Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2. Remote Sens Environ **147**, 1–12 (2014)

K. Fukue, M. Maeda, H. Shimoda, Continental scalse land cover classification using modis surface reflectance products. *The international archives of the photogrammetry, remote sensing and spatial information sciences, Vol. XXXVIII, Part 8* (2010). pp. 953–957

GLAS/ICESat L1 and L2 Global Altimetry Data (2014), http://nsidc.org/data/docs/daac/glas_icesat_l1_l2_global_altimetry.gd.html

Global atmospheric temperature at 700 hPa measured by AIRS on Aqua (2009), http://photojournal.jpl.nasa.gov/catalog/PIA12098

Global atmospheric water vapor (total precipitable water) measured by AIRS on Aqua (2009), http://photojournal.jpl.nasa.gov/catalog/PIA12097

Global distributions of HCN and C2H6 showing biomass burning (2003), http://www.imk-asf.kit.edu/english/673.php

Global Fire Map (2016), https://lance.modaps.eosdis.nasa.gov/imagery/firemaps/firemap.2016081-2016090.2048x1024.jpg

Global land cover map (2002), http://earthobservatory.nasa.gov/Newsroom/view.php?id=22585

Global land cover map using MERIS data (2008), http://www.esa.int/esaEO/SEMZ16L26DF_planet_0.html

GLOBCOVER, Products Description and Validation Report (2008), http://due.esrin.esa.int/files/p68/GLOBCOVER_Products_Description_Validation_Report_I2.1.pdf

R.M. Goldstein, H.A. Zebker, C.L. Werner, Satellite radar interferometry: Two-dimensional phase unwrapping. Radio Science **24**, 713–720 (1988)

H.R. Gordon, D.K. Clark, Clear water radiances for atmospheric correction of coastal zone color scanner imagery. Appl. Opt. **20**, 4175–4180 (1981)

H.R. Gordon, M. Wang, Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. Appl. Opt. **33**, 443–452 (1994)

H.R. Gordon, K.J. Voss, "MODIS Normalized Water-leaving Radiance", Algorithm Theoretical Basis Document (MOD18) (2004)

GOSAT Project (2009), http://www.gosat.nies.go.jp/eng/gosat/zu8.htm

A.W. Gruen, Adaptive least squares correlation: a powerful image matching technique. J. Photogram. Remote Sens. Cartography **14**, 175–187 (1985)

L. Guanter, C. Frankenberg et al., Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements. Remote Sens Environ **121**, 236–251 (2012)

P. Hofmann, Detecting urban features from IKONOS data using an object-oriented approach. RSPS2001 79–91 (2001)

J. Ishizaka, E. Siswanto, T. Itoh, H. Murakami, Y. Yamaguchi, N. Horimoto, T. Ishimaru, S. Hashimoto, T. Saino, Verification of vertically generalized production model and estimation of primary production in the Sagami Bay, Japan. J. Oceanography **63**, 517–524 (2007)

D.J. Jackson, *Classical Electrodynamics*, 3rd edn. (American Association of Physics Teachers, New York, 1962)

M.Z. Jacobson, The short-term cooling but long-term global warming due to biomass burning. J. Climate **17**, 2909–2925 (2004)

N. Jacquinet-Husson, V. Capelle, L. Crépeau, R. Armante, N.A. Scott, A. Chédin,, The GEISA/IASI Spectroscopic database in its 2008 Edition. *ISSWG2-2 IASI Sounding Science Working Group, Darmstadt, Allemagne, 21–22 April 2009* (2009)

Japan Association of Remote Sensing. *Remote Sensing Graphical Explanation*. Japan Association of Surveyors (2001)

Japan Association of Remote Sensing. *Remote Sensing Note*. Sensing, Japan Association of Remote (1985)

Jason, cycle002, Period:25/01/2002 - 04/02/2002 (2002), http://www.jpl.nasa.gov/images/earth/jason_mer.jpg

JAXA GLOBAL RAINFALL WATCH. 4, 13:00 4 2016. http://sharaku.eorc.jaxa.jp/GSMaP/index.htm

JAXA. MODIS near real time RcRefl (2015), http://kuroshio.eorc.jaxa.jp/ADEOS/mod_nrt_new/1129 data/tric/nrt/day/201510/2502/A2GL11510250218OD1_ONLTE_08000_05300_RcRefl.png

Z. Jelenak, T. Mavor, L. Connor, N.-Y. Wang, P. S. Chang, P. Gaiser,. Validation of ocean wind vector retrievals from WindSat polarimetric measurements. *The 4th Int. Asian-Pacific Environmental Remote Sensing Conference* (2004)

W.L. Jones, L.C. Schroeder, F.J. Wentz, Microwave scatterometer measurements of oceanic wind vector. *Proceedings of the Symposium, 26–30 May 1980*. Venice, (1981), pp. 553–562

I.R. Joughin, R. Kwok,M.A. Fahnestock, Interferometric estimation of the three dimensional ice-flow velocity vector using ascending and descending passes. TGARS **36**, 25–37 (1998)

Y.J. Kaufman, The atmospheric effect on remote sensing and its correction, in Asrar, G. ed. *Theory and Application of Optical Remote Sensing* (J.W. Sons, NY, 1989), pp. 336–428

G. Kelly, The relative contributions of the various space observing system. *2007 EUMETSAT Meteorological Satellite Conference*. EUMETSAT (2007)

Y.S. Kotchenova, F.E. Vermote, R. Matarrese, F.J. Klemm Jr., Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path Radiance. Appl. Opt. **45**, 6762–6774 (2006)

E.R. Kursinski, G.A. Hajj, J.T. Schofield, R.P. Linfield, K.R. Hardy, Observing the earth's atmosphere with radio occultation measurements using the Global Positioning System. J. Geophys. Res. **102**, 23429–23465 (1997)

G. Lagerloef, R. Schmitt, Role of ocean salinity in climate and near-future satellite measurements: Meeting Report. EOS. Trans. Am. Geophys. Union **87**, 466–467 (2006)

G.S.E Lagerloef, Introduction to the special section: the role of surface salinity on upper ocean dynamics, air sea interaction and climate. J. Geophy. Res. 107, 8000 (2002)

H.E. Larsen, Ship route planning in Arctic ice infested waters using Near-Real-Time satellite image products (2016), http://www.arcticfrontiers.com/downloads/arctic-frontiers-2016/presentations-4/thursday-28-january-2016/part-iii-technology-needs-1/1301-12-hans-eilif-larsen/file

J.S. Levine, W.S. Cofer, D.R. Jr. Cahoon, E.L. Winstead, The global impact of biomass burning. Environ Sci Technol (1995), pp. 1–55

LIDAR LEVEL1 BROWSE IMAGES - 2011-01-15 01:41:08Z - SECTION2 (2011), http://www-calipso.larc.nasa.gov/products/lidar/browse_images/show_detail.php?s=production&v=V3-01&browse_date=2011-01-15&orbit_time=01-41-08&page=2&granule_name=CAL_LID_L1-V alStage1-V3-01.2011-01-15T01-41-08ZD.hdf

D. Luman, T. Tweddale, Assessment and potential of the 2007 USDA-NASS cropland data layer for statewide annual land cover applications (2008), https://www.ideals.illinois.edu/bitstream/handle/2142/18134/INHS2008_49.pdf?sequenc=1

S.N. Madsen, H.A. Zebker, Imaging radar interferometry. Principles & Applications of Imaging Radar, in *Manual of Remote Sensing*, 3rd edn. (Wiley, New York, 1998)

E.P. McClain, W.G. Pichel, C.C. Walton, Comparative performance of AVHRR-based multichannel sea surface temperatures. J. Geophy. Res. **90**, 11587–11601 (1985)

W.G. Melbourne, et al. *The Application of Spaceborne GPS to Atmospheric Limb Sounding and Global Change Monitoring*. Pasadena: Jet Propulsion Laboratory, California Institute of Technology (1994)

Mie G, Ann. Physik 377 (1908a)

Mie G, Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. Leipzig, Ann. Phys. **300**: 377–445 (1908b)

MODIS Rapid Response System Global Fire Maps (2016), https://lance.modaps.eosdis.nasa.gov/imagery/firemaps/firemap.2016081-2016090.2048x1024.jpg

MODIS Snow Cover Extent (2015), http://kuroshio.eorc.jaxa.jp/JASMES/daily/global/data/CSF10D/201502/MDS20150127_20150205_GLBOD10D_SNWFG_EQ05KM_304.png

J.J. Mohr, N. Reeh, S.N. Madsen, Three dimensional glacial flow and surface elevation measured with radar interferometry. Nature **391**, 273–276 (1998)

J.L. Monteith, Solar radiation and productivity in tropical ecosystems. J. Appl. Ecol. **9**, 747–766 (1972)

Munich RE, NETCATSERVICE (2015), http://www.munichre.com/en/reinsurance/business/non-life/georisks/natcatservice/default.aspx

Myanmar flood water observation by PALSAR (2008), http://www.eorc.jaxa.jp/en/earthview/2008/tp080509.html

F. Naderi, M.H. Freilich, D.G. Long, Spaceborne radar measurement of wind velocity over the ocean – an overview of the NSCAT scatterometer system. Proceedings of the IEEE **79**, 850–866 (1991)

T.Y. Nakajima, A. Higurash, T. Nakajima, S. Fukuda, S. Katagiri, Development of the cloud and aerosol retrieval algorithms for ADEOS-II/GLI mission. Journal of Remote Sensing Society of Japan **29**, 60–69 (2009)

T. Nakajima, et al., GSS Reference Handbook (RSTAR Reference Handbook). *EORC Bulletin, Technical Report No.15, 360p., ISSN 1346–7913, JAXA/EORC* (2004)

NASA RapidScat Proving Valuable for Tropical Cyclones (2015), http://www.jpl.nasa.gov/news/news.php?feature=4562

NASA/JPL, *Shuttle Radar Topography Mission* (2016), http://www2.jpl.nasa.gov/srtm/

NASA/JPL, SRTM DSM (2004), http://photojournal.jpl.nasa.gov/jpegMod/PIA04965_modest.jpg

R.R. Nemani, C.D. Keeling, H. Hashimoto, W.M. Jolly, S.C. Piper, C.J. Tucker, R.B. Myneni, S.W. Running, Climate-driven increases in global terrestrial net primary production from 1982 to 1999. Science **300**, 1560–1563 (2003)

M. Neubert, Segment-based analysis of high resolution satellite and laser scanning data. *The 15th International Symposium Informatics for Environmental Protection* (Marburg Mtropolis, Zurich, 2001), pp. 379–386

Northeastern China Suffers Large-Scale Flooding (2003), http://www.eorc.jaxa.jp/en/earthview/2003/tp031112.html

Office, US CLIVAR, Report of the US CLIVAR Salinity Working Group. US CLIVAR Report 2007–1 (2007)

S. Perry, F. Kruse, ASTER data use in mining applications, in *Land Remote Sensing and Global Environmental Channge*, ed. by B. Ramachandran, C.O. Justice, M.J. Abrams (Springer, New York, 2011), p. 315. http://www.jspacesystems.or.jp/ersdac/GDEM/ver2Validation/Appendix_A_ERSDAC_GDEM2_validation_report.pdf

A.B. Pour, M. Hashim, Geological structure mapping of the Bentong-Raub Suture Zone, Peninsular Malaysia using PALSAR Remote Sensing Data. ISPRS Annals Photogram. Remote Sensing Spat. Inf. Sci. 89–92 (2015), http://www.palsar.ersdac.or.jp/data/kouhou/

Primary Productivity of Phytoplankton (2004), http://suzaku.eorc.jaxa.jp/GLI/doc/GLI_BOOK_CD/PDF/CHAP_6.PDF

R.T. Rabus, D.R. Fatland, Comparison of SAR-interferometric and surveyed velocities on a mountain glacier: Black Rapids Glacier, Alaska, U.S.A. J. Glaciol. **46**, 119–128 (2000)

L. Rayleigh (John William Strutt) On the light from the sky, its polarization and colour. Philos. Magaz. 107–120, 274–279 (1871)

C.D. Rodgers, *Inverse Method for Atmospheric Sounding* (World Scientific Pulishing, 2000)

L.S. Rothman, I.E. Gordon, A. Barbe et al., The HITRAN 2008 molecular spectroscopic database. Journal of Quantitative Spectroscopy & Radiative Transfer **110**, 533–572 (2009)

Route planning for ships in ice (n.d), http://www.ccrs.nrcan.gc.ca/resource/tutor/gsarcd/pdf/ap_ice_e.pdf

Y.-J. Park, K. Ruddick, Model of remote-sensing reflectance including bidirectional effects for case 1 and case 2 waters. Appl. Opt. **44**, 1236–1249 (2005)

D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation, in *Parallel Distributed Processing*, ed. by J.L. McCleland, D.E. Rumelhart, The PDP Research group, vol 1 (The MIT Press, Cambridge, MA, 1986a)

D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors. Nature **323**, 633–636 (1986b)

S.W. Running, S.T. Gower, FOREST-BGC, a general model of forest ecosystem processes for regional applications. II. Dynamic carbon allocation and nitrogen budgets. Tree Physiology **9**, 147–160 (1991)

S.W. Running, R.R. Nemani, F.A. Heinsch, M. Zhao, M. Reeves, H. Hashimoto, A continuous satellite–derived measure of global terrestrial primary production. Bioscience **54**, 547–560 (2004)

S.W. Running, P.E. Thornton, R. Nemani, J.M. Glassy, Global terrestrial gross and net primary productivity from the Earth Observing System, in *Methods in Ecosystem Science*, ed. by O. Sala, R. Jackson, H. Mooney (Springer, New York, 2000), pp. 44–57

C. Schaaf, Recent developments in the MODIS Albedo, Nadir BRDF Adjusted Reflectance (NBAR) and Reflectance Anisotropy Products (MCD43) (2010), http://modis.gsfc.nasa.gov/sci_team/meetings/201001/presentations/land/schaaf.pdf

C. Schaaf, A. Strahler, et. al., MODIS Reflectance Albedo and Reflectance Anisotropy. Land and Vegetation Direct Readout workshop (2007)

B. Schölkopf, A. Smola, K.-R. Mülller, Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Tubingen: Technical Report No. 44. Max-Planck Institut für biologische Kybernetik (1996)

SEA SURFACE TEMPERATURE (1 MONTH - AQUA/MODIS) (2015), http://neo.sci.gsfc.nasa.gov/view.php?datasetId=MYD28M&year=2015

D.A. Siegel, M. Wang, S. Maritorena, W. Robinson, Atmospheric correction of satellite ocean color imagery: the black pixel assumption. Appl. Opt. **39**, 3582–3590 (2000)

SST and chlorophyll-a distribution and fishing grounds of Sanriku coast Japan, (2003), http://suzaku.eorc.jaxa.jp/GLI/doc/GLI_BOOK_CD/PDF/CHAP_6.PDF

A. Strahler, MODIS land Cover Product Algorithm Theoretical Basis Document (ATBD) Version 5.0. (1999), http://modis.gsfc.nasa.gov/data/atbd/atbd_mod12.pdf

T. Tadono, H. Ishida, F. Oda, S. Naito, K. Minakawa, H. Iwamoto, Precise Global DEM Generation by ALOS PRISM. ISPRS Annals Photogramm Remote Sens Spat Inf Sci **II-4**, 71–76 (2014)

M. Takagi, H. Shimoda, *Handbook of Image Analysis, revised ed.* (University of Tokyo Press, Tokyo, 2004a)

M. Takagi, H. Shimoda, *Handbook of Image Analysis (in Japanese)* (University of Tokyo Press, Tokyo, 2004b)

The land displacement by the 2011 off the Pacific coast of Tohoku Earthquake observed by PALSAR on ALOS (2011), http://www.eorc.jaxa.jp/ALOS/en/img_up/l_dis_inf_tohokueq_110315_f2e.htm

The Mineral Distribution Map near Talc Deposit Area of Mt.Fitton (Australia) using Full Band Data of ASTER (n.d.), http://www.science.aster.ersdac.or.jp/en/topic_image/Geology/001.html

H. Tian, J.M. Melillo, D.W. Kicklighter, A.D. McGuire, J. Helfrich, The sensitivity of terrestrial carbon storage to historical climate variability and atmospheric CO_2 in the United States. Tellus **51**, 414–452 (1999)

Total NO2 measured by SCIAMACHY on ENVISAT (2011), http://www.temis.nl/airpollution/no2col/data/scia/2011/07/no2trop201107.gif

V. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 1995)

E.F. Vermote, N. El Saleoql et al., Atmospheric correction of visible to middle-infrared EOS-MODIS data over land surfaces: background, operational algorithm and validation. J. Geophys. Res. Atmos. **102**, 17131–17141 (1997)

E.F. Vermote, A. Vermeulen, Atmospheric correction algorithm: spectral (1999), http://modis.gsfc.nasa.gov/data/atbd/atbd_mod08.pdf, http://modis.gsfc.nasa.gov/data/dataprod/dataproducts.php?MOD_NUMBER=09

T. Meissner, F. Wentz, An updated analysis of the ocean surface wind direction signal in passive microwave brightness temperature. IEEE Trans. Geosci. Remote Sensing **40**, 1230–1240 (2002)

T. Westberry, M.J. Behrenfeld, D.A. Siegel, E. Boss, Carbon-based primary productivity modeling with vertically resolved photoacclimation. Global Biogeochem. Cycles **22**, GB2031 (2008), Doi 10.1029/2007GB003078

Windsat images (2003), http://www.nrl.navy.mil/WindSat/images/wndmap.jpg

XCH_4 distribution level3 (2014), http://data.gosat.nies.go.jp/GosatBrowseImage/browseImage/v02XX_L3/XCH4_L3_201404010430_v02.21.png

XCO_2 distribution level3 (2014), http://data.gosat.nies.go.jp/GosatBrowseImage/browseImage/v02XX_L3/XCO2_L3_201404010430_v02.21.png

H.A. Zebker, R.M. Goldstein, Topographic mapping from interferometric SAR observations. J. Geophys Res **91**, 4993–4999 (1986)

M. Zhao, S. Running, F.A. Heinsch, R. Nemani, Collection 005 change summary for the MODIS land vegetation primary production (17A2/A3) algorithm (2005), http://landweb.nascom.nasa.gov/QA_WWW/forPage/C005_Change_NPP.pdf

Geographic Information Systems and Geomatics

Jesus A. Gonzalez

Contents

Abstract

The role of spatial data for decision making has increased the need for geographic information systems. This chapter starts by briefly describing the theory of geographic information systems. After that we present the interactions of geographic information systems with remote sensing and global navigation,

---

J.A. Gonzalez (✉)
National Institute of Astrophysics, Optics, and Electronics (INAOE)/Regional Center for Space Science and Technology Education for Latin America and the Caribbean (CRECTEALC), Santa María Tonantzintla, Puebla, Mexico
e-mail: jagonzalez@inaoep.mx

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_48

positioning, and timing satellite systems. This is done with the idea to illustrate how geographic information systems (GISs), remote sensing, and global navigation satellite systems work together in order to generate final products. Then, we focus on the capability of GIS to analyze spatial data. We finally present examples of applications of GIS and current trends of research in the area. The examples presented through this chapter capture how GIS can be used for decision-making tasks in different areas of knowledge.

Keywords

Coordinate system • Database management systems (DBMS) • Data mining • Datum • Ellipsoid • Geographic information systems • Geoid • Geomatics • Georeferencing • GIS conceptual framework • Global Positioning System • Location-based service • Map projection • Metadata • Pattern • Remote sensing • Spatial data analysis • Spatial database • Vector and raster models

Introduction

A great part of data that we generate can be related to a location on Earth. The concept of an object's location could be as simple as the place where it was generated or the place where an object can be found. The purpose for keeping track of the location of data varies depending on the application and may change in time. For instance, the location of an object might not be static; it could be dynamic, as in the case of an automobile in motion. This issue can be as complicated as the scope of the functionality of our application.

The inherent relation of geographic information to spatial dimensions attaches it to a location on Earth, with reference to a coordinate system. For this purpose, we require a representation of Earth in order to assign a specific location to an object. An option for this representation is a sphere with a radius of approximately 6,400 km and coordinates that are measured as latitude and longitude. Latitude is measured as the distance from the equator to a point to the north or the south of it. Longitude is measured as the distance from the (commonly) Greenwich meridian to the east or west. Height is another variable to take into account although more difficult to measure with high precision. A variation of this representation is the use of an ellipsoid instead of a sphere in order to obtain a closer representation of Earth for our geographic coordinate system.

Geographic data is produced from different sources. It can be identified by its spatial component. It may come from a source located in space, such as a remote sensing satellite or an airplane taking an aerial image. It could be data coming from a GPS satellite which we use to calculate our location. It could even be data from a sensor on Earth such as an electronic total station, as the tool of a civil engineer. An electronic total station is an electronic distance measurement (EDM) instrument that can be seen as the modern theodolite used to measure angles and directions, among others. No matter what the source of the data is, if it is being acquired, it is important to store it. Database management systems have evolved to make room for spatial data. Nowadays, these systems are able not only to store this type of data but also allow users to formulate queries that evaluate objects on the basis of spatial relations. In this way, we can find important information that is useful for decision making in different applications.

The evolution of spatial databases did not stop with these facilities. It is here where geographic information systems (GISs) were developed to create, visualize, and manipulate spatial data. A GIS is composed of a spatial database, a graphic user interface, and a set of tools to manipulate spatial data. Furthermore, many GISs are created to work in the web environment so that multiple users are able to obtain the benefits of a web application.

Once the GIS has been built and data has been collected, there is a treasure to be exploited, resulting in valuable information that can be used for decision making. There are different ways to analyze the data and information stored on GIS. One of them consists of overlaying of layers of data or information, as a way of organizing different types of data (as we will see in the following sections). Another way could be with the use of spatial queries. One more could be the specific processing of the data, as in the case of the creation of a network flow model. One more could be a data mining analysis.

Geomatics is defined as the field of study related to the gathering, storing, processing, management of spatially georeferenced data, and delivering of geographic information. These topics, among others, will be covered in the following sections. The rest of the chapter describes with more detail the conceptual framework of a GIS, its interactions with remote sensing data, some of its applications, the role of GIS in decision making, future trends of GIS, and conclusions.

GIS Conceptual Framework

In the previous section, we were introduced to the broad picture of GIS. Now we will concentrate in more detail on important concepts that support the theory of this multidisciplinary area. Let us start with by defining a GIS as a collection of components necessary to store spatial data to be manipulated in order to create spatial products (see Fig. 1). Then, the components of a GIS can be classified in three main categories. The first component is computer hardware to store data and software. The second component is computer software, required to manipulate data and create valuable products. Finally, the last and the most valuable component is geographic data. In order to manage geographic data, database management systems (DBMS) have been extended to deal with the spatial component. There are both commercial and open source tools with such an extension. An example of an open source DBMS with such an extension is PostgreSQL, with its spatial component called postGIS. For more information about postGIS, please refer to http://postgis.refractions.net/.

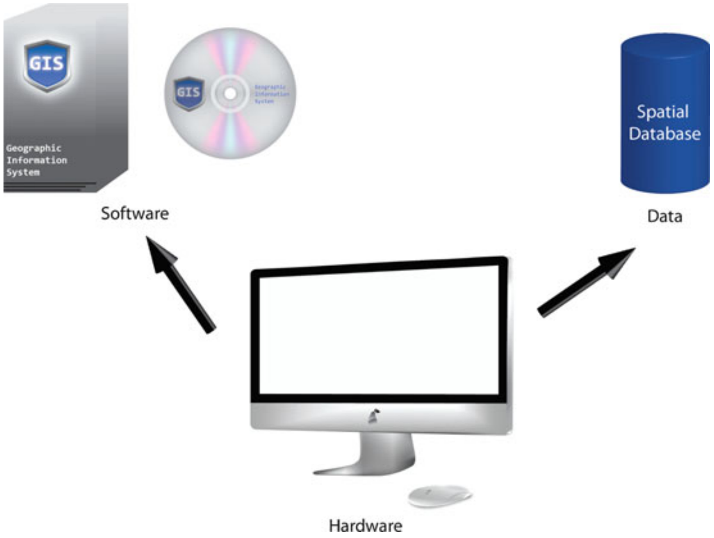


Fig. 1 Main components of a geographic information system: computer hardware, software, and geographic data. Computer hardware is the data and software repository. Software is used to manipulate geographic data and to create valuable products used for decision making. Data is the heart of a GIS, its prime matter

GIS Sources of Data

There are different sources of data that we can store in a GIS (see Fig. 2). Much of the information was already available when GIS came into the market. This was the case of paper maps that were digitized and introduced to them. Data obtained from field work is another important source of data because although it is not an efficient way to obtain it, sometimes it is the only way to get data on a specific variable. Satellite and aerial images are yet another source of data for GIS. This type of data requires different preprocessing steps, depending on the purpose of the GIS. As we discussed before, information coming from global navigation satellite systems, such as GPS, can also be stored in a GIS (Lee 2001a). These are some of the sources of data for a GIS, but there might be more, in which case the other data depends on the spatial components of our GIS. Once we have identified these spatial components, we ask ourselves the question, what is the source of any other information that we require? The answer is directly related to the problem that our GIS application is intended to solve.

Vector and Raster Models

In order to introduce data into a GIS, we need to know how it is organized. There are two models to store data in a GIS. They are known as the vector and raster models. The goal is to model the real world, to represent it in a level of abstraction with the

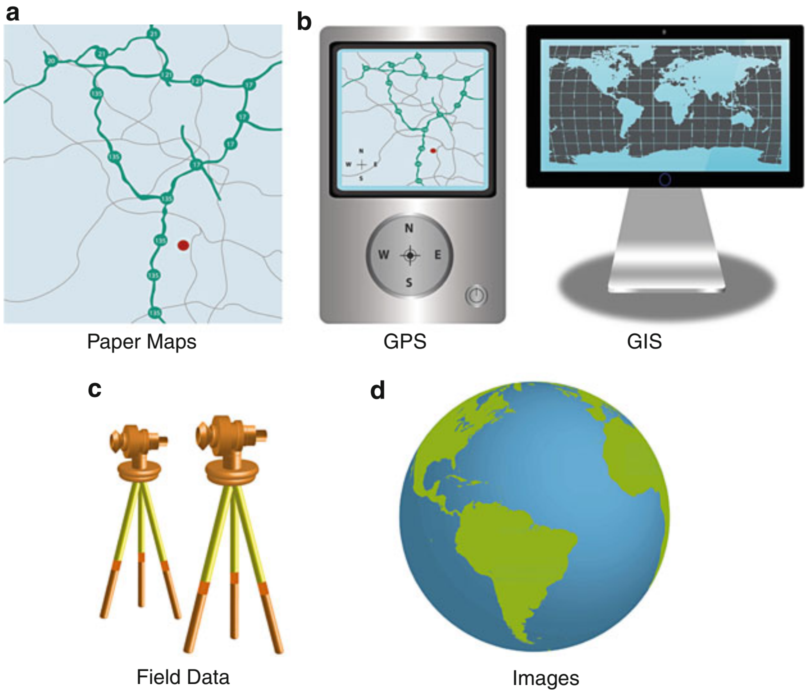


Fig. 2 Some of the sources of data of a geographic information system. (**a**) Paper maps, (**b**) data coming from GPS satellites, (**c**) data coming from field work, and (**d**) data coming from satellite and aerial images

adequate level of detail to be mapped to the GIS. We might not completely store every detail of reality but only as much as we require. Let us assume for now that we are capturing this data from satellite images, but as we know, we could also obtain it from measures in the field or even from other ways.

In the vector model, each object in the real world is represented with one of three possibilities. These are a point, a line, or a polygon. Figure 3 shows an example of a road, a school, and a parcel represented with a line, a point, and a polygon, respectively. An important characteristic of this model is that the spatial relations between different objects can be captured. These are known as topological, metric, and direction relations. More information about the topic can be found in Koperski and Han (1999).

In the case of the raster model, data is represented with a grid of data. Each cell in the grid corresponds to a pixel in the image and is classified as a type of object. In the previous example, the line representing the road fills the points in the grid that overlap with the road. The school is represented with one pixel and the parcel for sale is represented by a set of pixels (an area) as can be seen in Fig. 4.

Up to now we know how data is structured to be stored in a GIS. However, we are missing one of the fundamental concepts that make a GIS so useful: its capability to

Fig. 3 GIS vector model. Partial representation of the downtown of a city. A parcel for sale is represented with a polygon, main road with a polyline, and central school with a point

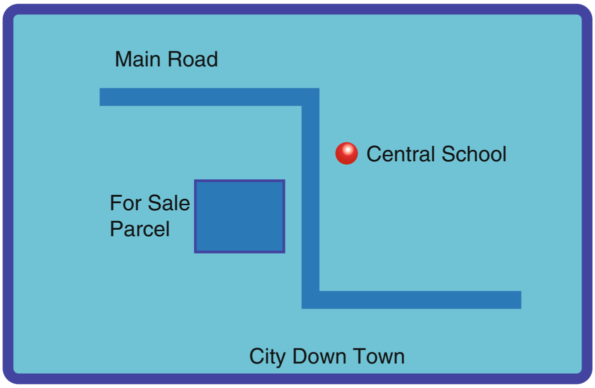
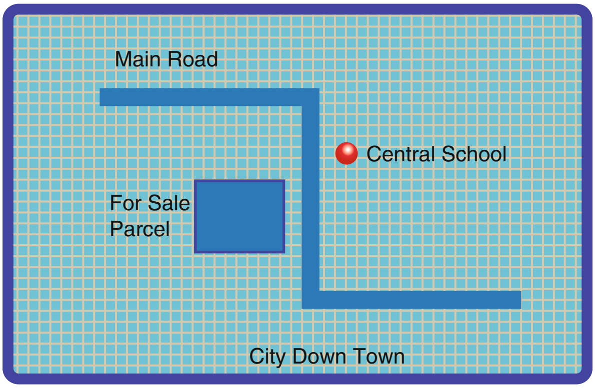


Fig. 4 GIS raster model. Partial representation of the downtown of the city modeled with the GIS vector model of Fig. 3 with the raster model. In this case, the parcel for sale can be identified by the cells filled in blue. The main road can be identified by the cells filled as well in *blue*. Finally, central school corresponds to the cell filled in *red*



manage locations. We need to be able to attach location to the objects that we store. For this, we require the use of a georeference system as we describe in the following section.

Georeferencing

When we load an image in a GIS, we want to identify the location of the objects in it. In order to do it, we need to georeference the image. We do this by finding a correspondence of the locations and a map projection using a coordinate system. For this, we introduce some concepts about Earth. We are first concerned with the measurement of Earth (Gelati 2006). There is a division of science known as geodesy or geodetics that does this. The important pieces of geodesy that concern GIS are the reference Earth shape or ellipsoid geodetic positioning or geodetic datum and coordinates, the true Earth shape or geoid and vertical datum, and the practical representation of the Earth or map projections.

The Ellipsoid

We use an ellipsoid as a good approximation to represent the shape of the Earth. This ellipse rotates around one of its axes. There have been different proposals of ellipsoids. One of the most common is the WGS84, which parameters are an equatorial axis of 6,378,137.00 m, a polar axis of 6,356,752.3142 m, and an inverse flattening of 298.257223563. As we know, the ellipsoid does not accurately represent the shape of the Earth. There are mountains that are higher than the line of the ellipse and there also exist places below sea level, which are below the line of the ellipse. This proves that this is not an accurate representation of Earth.

Mean Sea Level (Geoid)

As we can note, the shape of the Earth is very irregular. There is not any geometric body that has the exact shape of the Earth. This is the reason why the Earth's shape received the name of the geoid. The geoid is an irregular equipotential surface that coincides with mean sea level over the oceans. It has also an imaginary continuity across the continents, which have undulations on the surface (the topography) because of the irregular distribution of the gravitational mass forces of the planet. The geoid is used as a reference surface for leveling, that is, we measure elevation relative to the geoid (Li and Götze 2001). A more detailed description about the geoid concept can be found in Mok and Chao (2001). Figure 5 illustrates how the geoid differentiates from the ellipsoid.

A geodetic datum is defined as a reference model that associates a geodetic reference ellipsoid (the ellipsoid parameters: equatorial axis, polar axis, and inverse flattening) to a coordinate system (defined by a geodetic space through orientation, position, and scale). A geodetic datum is a mathematical model of Earth (Gelati 2006). There are two types of datums. These can be either geocentric or local.

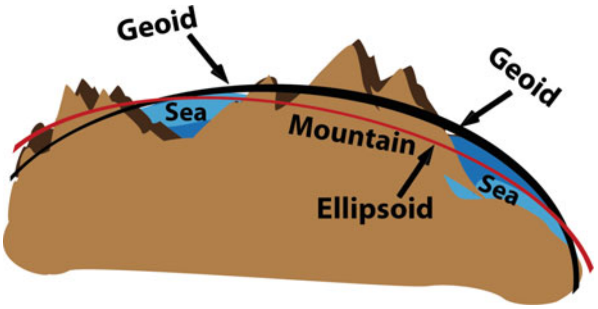


Fig. 5 Difference between the geoid and the ellipsoid. The ellipsoid is a smooth geometric shape. The geoid passes over or under the ellipsoid depending on the irregular distribution of the gravitational mass forces of Earth. We can also appreciate how the topography of Earth differs from that of the geoid and the ellipsoid

A geocentric datum is globally centered and is a good approximation of the whole Earth. In this case, the center of the reference ellipsoid coincides with the Earth's center of mass (Gelati 2006). In Fig. 6 we can see how the center of the ellipsoid and the center of mass of the Earth coincide. We can also see that the geoid (or mean sea level) and the reference ellipsoid are in general a good approximation.

A geocentric datum is best suited for global applications, just as GPS uses the WGS84 geocentric datum. In contrast, a local geodetic datum better suits a particular region where the reference ellipsoid has better adjustment with the Earth's shape. In this case, the center of the ellipsoid does not always coincide with the Earth's center of mass. Because of this, a local geodetic datum does not provide a good global representation of the Earth. Figure 7 shows how the ellipsoid's adjustment to the Earth is better in a local region than in the rest of it.

Once we described the ellipsoid, the geoid, and the datum, we will describe the geographic coordinate system, which is based on the Earth's rotation around its

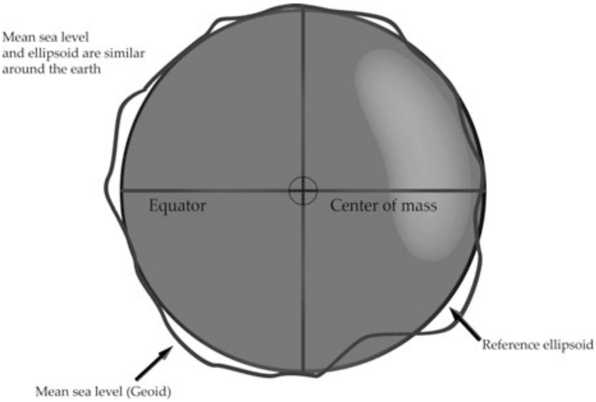


Fig. 6 The geocentric datum. In a geocentric datum, the center of mass coincides with the center of the reference ellipsoid. It is a good approximation of the whole Earth, but there can be more accurate approximations for particular regions (This figure was adapted from Fig. 7.1 in Gelati (2006))

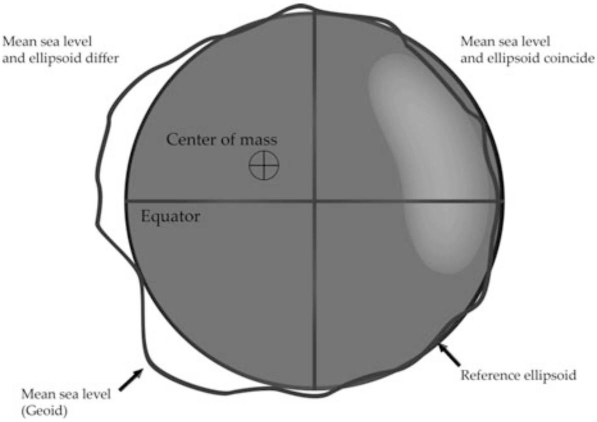


Fig. 7 The local geodetic datum. In a local geodetic datum, the center of mass does not coincide with the center of the reference ellipsoid. It is a good approximation of a region of the Earth, but it is not for the whole Earth. We can see this in the figure because the geoid and the ellipsoid have better adjustment in a particular region of Earth than the rest of it (This figure was adapted from Fig. 7.2 in Gelati (2006))

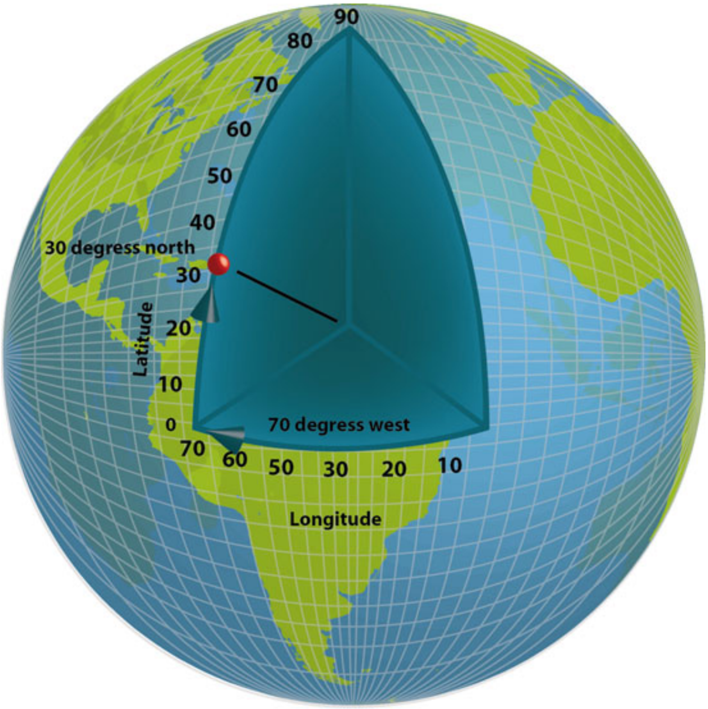


Fig. 8 Geographic coordinate system. Latitude is measured north or south from the equator, while longitude is measured east or west from the Greenwich meridian (Figure adapted from (http://services.arcgisonline.com/arcgisexplorer500/help/latlong_from_globe_center.png))

center of mass. We can determine the geographical coordinates of any point on the Earth's surface based on its latitude and longitude. The Earth's center of mass is on its rotation axis. The plane that passes through the center of mass and perpendicular to the axis defines the equator. Latitude is defined as the angle from the meridian between the equator and the reference parallel. This will always be north (N) or south (S). Then, the maximum latitude will be of 90°. Longitude is a geographic coordinate that defines the east (E) or west (W) position of a point on the Earth's surface. It is the angle measured east or west between the plane containing the prime meridian (Greenwich) and a plane containing the North Pole, the South Pole, and the location in question (Longley et al. 2005) (Fig. 8).

Even when an ellipsoidal representation of Earth has the advantage of being realistic, it has some disadvantages. For example, it is impossible to observe the entire terrestrial surface at the same time; we can only see one of its faces but not the other. This does not happen with a map. In a map, we can show the whole world. The ellipsoidal representation is not easy to manage as a map is. We cannot change the scale of the terrestrial globe in a practical way as we do with our paper map. Of course, there are many disadvantages of paper maps that we do not have with a terrestrial globe such as the geometric deformations suffered during the map projections that were used to create the maps. Now we will describe what a map projection is.

Map Projection

A map projection is defined as a mathematical transformation between the geographic coordinate system (in latitude and longitude) and a system on the plane surface. There are two common planar coordinate systems. One of them is the Cartesian system (with X, Y coordinates) and the other is the polar system (with range and angle coordinates). The problem with this transformation is that three main types of distortions are introduced. These are length, area, and angular distortions. This means that length, area, and angles cannot be preserved by a single map projection at the same time. As an example, a length distortion means that length measured on a map does not correspond to the length of the same feature measured in the real world. This is the distortion introduced by the map projection. That is, the use of plane geometry and trigonometry involving Cartesian coordinates to perform the calculations does not lead to correct results after the map projection. There is also an error when we measure angles in the map. For more information about map projections, please refer to Lee (2001b) (Fig. 9).

Until now, we have studied the basics of GIS. We know what a GIS is, its sources of data, and the vector and raster models. We are able to recognize spatial data, and we learned the difficulties involved with approximating the shape of the Earth. We can also find the coordinates of a location on Earth. Now, we will learn about the

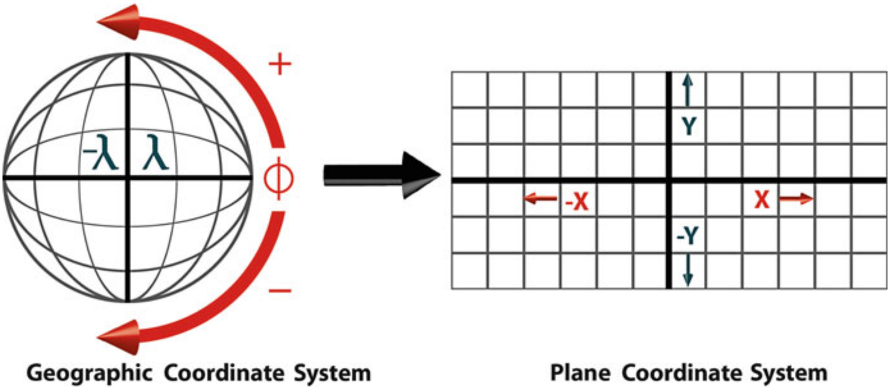


Fig. 9 Map projection. The geographic coordinate system using latitude (ϕ) and longitude (λ) is projected into a plane coordinate system using X and Y coordinates, in this case, a Cartesian coordinate system (This figure was adapted from the online course of geography, lesson 7: A deeper understanding of coordinate systems and projections (https://www.e-education.psu.edu/geog486/l7_p9.html))

interactions between remote sensing and geographic information systems as a way to improve each other's capabilities.

Metadata

Metadata is commonly referred as data about the data. This is the information that we use to document our data. In this way, metadata describes all the parameters necessary to work with spatial data: the data owner, source, resolution, and scale. A metadata framework can be described in different formats such as ASCII, HTML, Extensible Markup Language (XML), Standard Generalized Markup Language (SGML), and Resource Description Framework (RDF). In order to create compatibility among geospatial products and tools, a great effort to create standards over geospatial data has been done. Some of the available standards for geospatial metadata are (Gelati 2006; ISO 2011):

- ISO 19139 Geographic Information Metadata XML Schema Implementation
- ISO 19115 Geographic Information Metadata
- Content Standard for Digital Geospatial Metadata
- Dublin Core Metadata Element Set
- Australian Government Locator Service
- UK GEMINI Discovery Metadata Standard

Interactions of GIS with Satellite Systems

In this section, we describe in more detail how GIS interacts with remote sensing and the Global Positioning System platforms. The integration of these interactions has made possible what we have today and most of what we are creating for the future.

Geographic Information Systems and Remote Sensing

Remote sensing (RS) and GIS are two areas that interact with each other. There are three main ways in which these interactions can be combined to enhance each other (Wilkinson 1996). In the first one, RS is used as a tool to obtain data to be used in a GIS. Second, GIS data can be used as auxiliary information to improve products created from RS sources. Finally, RS and GIS are usually used together for modeling and analysis processes (Weng 2010).

RS contributes to the information that is stored in a GIS in different ways. One of the most important contributions is the extraction of thematic information from satellite images to create GIS layers. RS images are used to extract cartographic information to be the input to GIS, as in the case of the production of base maps. A very important application that requires the use of RS is the update of GIS databases. In this case, RS images are used to detect changes in thematic information to update

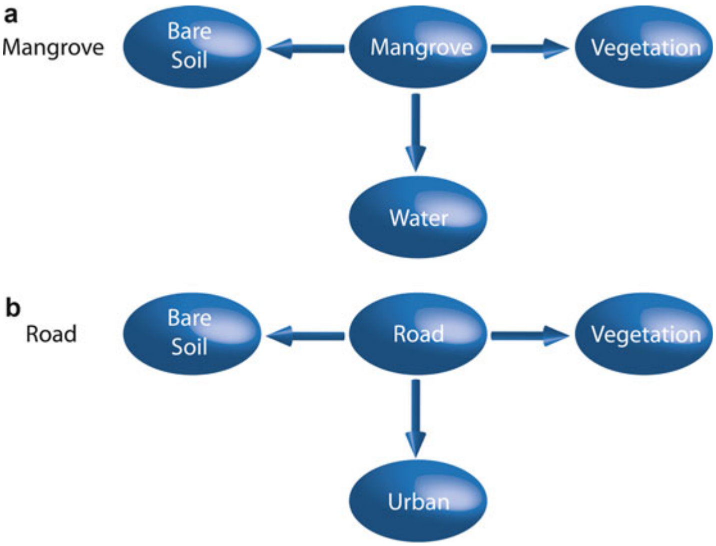


Fig. 10 Two patterns found by a graph-based spatial mining algorithm. (**a**) A graph pattern describing the mangrove class. This pattern says that a mangrove is usually found adjacent to regions of the classes bare soil, vegetation, and water. (**b**) A pattern describing the road class. This pattern says that a road is usually found adjacent to regions of the classes bare soil, vegetation, and urban

GIS databases. RS images have also been used as background for GIS representations. This is the case of visualization tools for digital elevation models, which are very important for different applications (Weng 2010).

On the other hand, GIS data is used to improve some of the processes used in RS. These processes are, among others, the selection of the area of interest, its preprocessing, or its classification (Weng 2010). It is of great interest how GIS context information can be used to post-process the classification results of a statistical RS classification algorithm to improve its accuracy (Gonzalez et al. 2008). Another interesting approach is the use of a structural data representation (i.e., a graph-based representation) in order to use both types of information at the same time (nonspatial and spatial). In this way, the classification algorithm takes advantage of all the available information at the moment that it is performing the classification task (Pech et al. 2004). Figure 10 shows two patterns found through the Subdue system (Cook and Holder 1994), a graph-based spatial data mining process. Pattern (a) corresponds to the description of the class “mangrove” and tells us that, in general, a region of interest (ROI) that belongs to this class is adjacent to other regions of the classes “bare soil,” “vegetation,” and “water.” Pattern (b) describes the class “road” and tells us that, in general, a ROI that belongs to this class is adjacent to other regions of the classes “bare soil,” “vegetation,” and “urban.” This information is used in a post-processing step to validate the class assigned by a statistical classification algorithm in order to improve its classification accuracy.

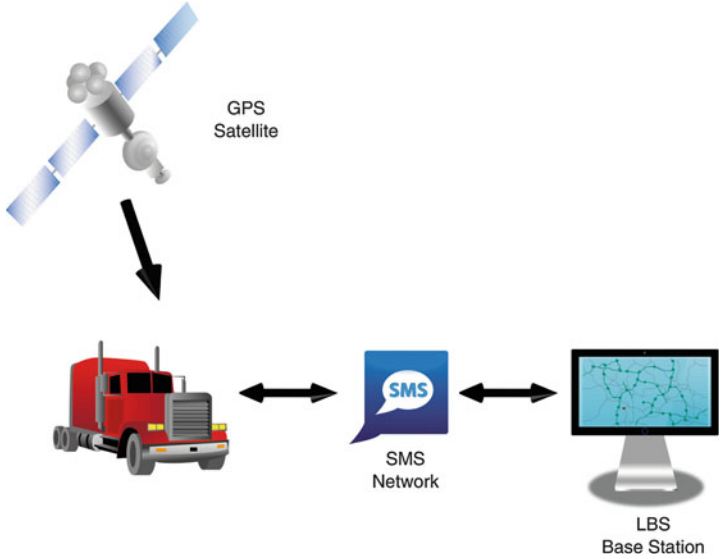


Fig. 11 GPS satellites updating a location-based service application. The satellites update the position of vehicles being monitored by the base station. The base station receives the position via SMS messages. The base station sends useful information (depending on their location: addresses of gas stations, restaurants, hotels, etc.) to the vehicles via SMS messages

Geographic Information Systems and Intelligent Positioning

GIS has also a mature level of integration with positioning systems, as in the case of the Global Positioning System (GPS) of the USA. In this case, we can say that there are four main levels of integration. In the first one, a GIS only takes the information reported by GPS and displays it in a map. In a second level, there are more functions. The GIS can manage WGS84 coordinates and different layers of the map (i.e., boundaries, counties, roads, rivers, and more). It is also possible to zoom in and out to take a look to a specific location. In a third level of integration, it allows entering waypoints (the coordinates of important reference points) describing interesting features. This allows the GPS–GIS system to create a GIS database that can be used to make a map. In the last level of integration, the map, an intelligent map, is associated to a set of logical rules that are used to improve the accuracy of the reported position (Taylor and Blewitt 2006). Figure 11 depicts a set of satellites sending positioning information to vehicles.

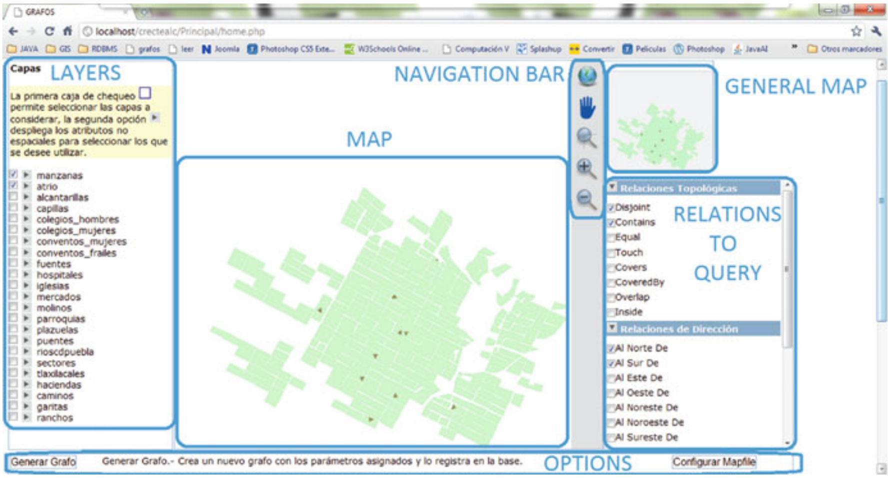
Spatial Data Analysis

The organization of spatial data in spatial databases is a plus that does not only facilitate the way a GIS accesses information but also provides users powerful analysis tools. The extension of a relational database into a relational spatial database requires adding geometry information to the spatial objects stored in it. This includes the coordinates that define both the shape of the objects (as points, lines, or polygons) and their coordinates in space. This information is commonly stored in a table related to another table that stores the nonspatial information describing the spatial object. Indexes over these tables are created in order to access the data in an efficient way.

A GIS connected to a spatial database provides different tools to analyze the data stored in such a database. It allows organizing data of the same type (i.e., roads are represented by lines, trees by points, parcels by polygons, county divisions with polygons, a satellite image of each county with polygons) in layers. Because of this, a form of visual analysis consists of the overlaying of several layers to allow the user to identify how different features of distinct layers interact in space. We can give transparency to any of the layers (i.e., a satellite image) so that we can appreciate important features with more detail. In this level of spatial data analysis, the user interacts with the GIS to create a useful product, name it a map, that can be used for decision making.

Another level of spatial data analysis is known as spatial data mining (SDM). In this case, data is usually extracted from the GIS or spatial database and transformed into a data representation that can be managed by the spatial data mining system. There are different spatial data mining tasks: clustering (Ng and Han 1994), spatial association rules (Koperski et al. 1996), co-location patterns (Xiong et al. 2004), and outliers detection (Shekhar et al. 2002), among others. Some of the more interesting data representations (and useful for spatial data) are those able to deal with structural data, such as inductive logic programming (Muggleton 1995) and graph-based learning (Cook and Holder 1994). Spatial clustering methods find patterns that share a spatial component. Spatial association rules try to associate spatial objects to neighboring objects. An approach known as co-location patterns states that we usually find in a nearby region instances of a set of spatial features. That is, when a subset of such spatial features are commonly located together (in a nearby region), it can be considered a co-location pattern.

In Fig. 12 we show the integration of a GIS with a graph-based data mining tool. The GIS loads the spatial data stored in a postGIS spatial database and presents the base map located at the center of the interface. The GIS allows the user to analyze the data by presenting the spatial layers contained in the spatial database (the option to perform this function is located in the upper left of the graphic user interface; see Fig. 12). It is also possible to perform spatial queries using topological, distance, and direction relations (as we can see to the right of Fig. 12). The interface has an option to transform the queried data into its graph-based representation in order to send it as input to a graph-based data mining system, for instance, the option called Subdue. Subdue performs the data mining task and finds spatial patterns. The instances of the patterns found can then be visualized in the main map so that the domain expert can interpret the mined results. This interface integrates a GIS system with a set of tools for decision making.



Capas

La primera caja de chequeo ☐ permite seleccionar las capas a considerar, la segunda opción ▸ despliega los atributos no espaciales para seleccionar los que se desee utilizar.

- ☑ ▸ manzanas
- ☑ ▸ atrio
- ☐ ▸ alcantarillas
- ☐ ▸ capillas
- ☐ ▸ colegios_hombres
- ☐ ▸ colegios_mujeres
- ☐ ▸ conventos_mujeres
- ☐ ▸ conventos_frailes
- ☐ ▸ fuentes
- ☐ ▸ hospitales
- ☐ ▸ iglesias
- ☐ ▸ mercados
- ☐ ▸ molinos
- ☐ ▸ parroquias
- ☐ ▸ plazuelas
- ☐ ▸ puentes
- ☐ ▸ rioscdpuebla
- ☐ ▸ sectores
- ☐ ▸ tlaxlacales
- ☐ ▸ haciendas
- ☐ ▸ caminos
- ☐ ▸ garitas
- ☐ ▸ ranchos

Relaciones Topológicas

- ☑ Disjoint
- ☑ Contains
- ☐ Equal
- ☐ Touch
- ☐ Covers
- ☐ CoveredBy
- ☐ Overlap
- ☐ Inside

Relaciones de Dirección

- ☑ Al Norte De
- ☑ Al Sur De
- ☐ Al Este De
- ☐ Al Oeste De
- ☐ Al Noreste De
- ☐ Al Noroeste De
- ☐ Al Sureste De

Generar Grafo.- Crea un nuevo grafo con los parámetros asignados y lo registra en la base.

Fig. 12 A GIS data mining graphical user interface. This is a GIS graphical user interface created to analyze data from the city of Puebla, in Mexico (this is the reason why the labels are written in Spanish). The GIS has three main components. The first component allows the user to analyze data overlaying layers. In the second level, the user can perform spatial queries. In a third level, the user can perform the graph-based spatial data mining task, having the opportunity to visualize the resulting patterns (subgraphs) in the map

Applications

The high capability of GISs to store spatial data, process it, analyze it, and create final products such as thematic maps makes them a powerful tool to apply to any field where spatial data plays a role. GISs are used for applications in industry, in government, in health care, in environment protection, and in many other areas. In the rest of this section, we briefly describe a couple of applications as examples of GIS applications.

In the area of medicine, GISs are very useful for epidemiology studies. This type of application allows physicians to keep track of how a disease expands geographically. If a different type of treatment is being applied in different counties or states, and the statistics are shown in real time in the GIS, the efficiency of each treatment can be appreciated in real time in the GIS graphical user interface. This application could be used for any type of disease. It could be the swine influenza A (H1) in humans, malaria, aids, cancer, or any other disease.

More dynamical GISs are those that receive signals from different sensors such as GPS, as in the case of navigation consoles for automobiles or electronic chart display and information system (ECDIS) for vessels. An ECDIS is commonly connected to a GPS, a radar system, a meteorological station, a gyroscope, and other sensors of the vessel. The GIS presents the navigation charts and, with the help of the GPS, it plots

Fig. 13 An electronic chart display and information system. This system is composed of a touch screen monitor (*bottom*) to control the GIS functionality. The *top* monitor displays the information received through the internal network from all the sensors of the vessel connected to the navigation system

the position of the vessel. The GIS allows drawing the path that the vessel should follow in the navigation chart. In addition, the radar system communicates with the navigation console (the ECDIS) and transmits the objects that it detects so that they can be plotted in the navigation chart and can be considered as dangerous or being in the middle of the path of the vessel. In this case, the ECDIS should play an alarm so that the vessel's captain considers a maneuver to avoid the blocking object (perhaps another vessel) (Fig. 13).

Another important area of application for GIS is that dedicated to disaster management. These GIS tools are created for any of the four phases of disasters: mitigation, preparedness, response, or recovery (UN-SPIDER 2011). Examples of information in GISs in this application may include floods, earthquakes, oil spills, storms, fires, tsunamis, volcano eruptions, epidemics, and droughts, among others. This, being an important area for any government, is an area of opportunity for GIS. For more information about the area of disaster management, please refer to the United Nations Portal of Knowledge at http://www.un-spider.org/knowledge-base. Figure 14 shows the emergency response cycle that considers its four phases: preparedness, response, recovery, and mitigation.

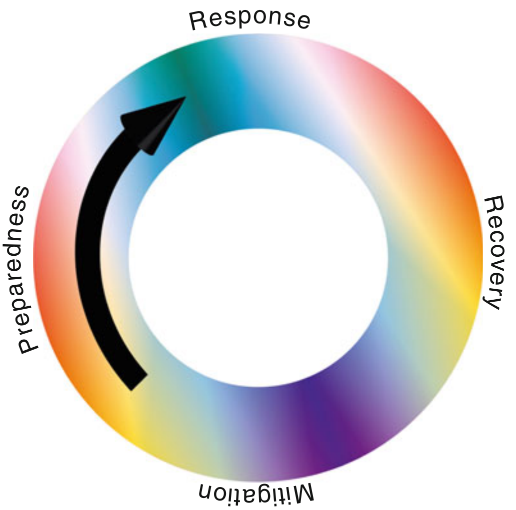


Fig. 14 The emergency management cycle (Adapted from Wikipedia)

Examples of Current Trends

GISs are tools that can be used in any field of study. They are being used to make more efficient processes as part of any industry or government. This multidisciplinary work demands different research areas to meet and innovate. Some of these current topics are the following:

- Augmented reality and GIS are being combined in different applications. The goal is to simulate how the real world would look like if we added artificial objects to it. Examples of augmented reality applications are its integration to landscape visualization (Ghadirian and Bishop 2008). Another example is the use of augmented reality for underground infrastructure visualization (Schall et al. 2009).
- Another important application is the integration of semantic information to the 3D reconstruction of city models as we can see in recent research (Kolbe 2008). In this work, the author uses GML3 to represent the shape, the graphical appearance of the city models, the semantics, the representation of the thematic properties, and the taxonomies of the objects. GML is the Geography Markup Language, an eXtensible Markup Language (XML) grammar created to express geographical features. In Wolf and Asche (2010), the authors create a 3D tactical intelligence surveillance map for a group of crime experts who study spatiotemporal patterns of residential burglary crimes.
- The integration of artificial intelligence techniques, such as fuzzy theory, is not new but is being more and more useful. An example of such a case can be seen in

Kanjilal et al. (2010), in which the authors find an appropriate implementation approach to fuzzy regions.

- These are some examples of both current research areas and applications of GIS that are used to solve real-world problems.

Conclusion

Geographic information systems are an advanced technology that allows developing applications in any area of study. Their power to analyze data enables them to create tools ideal for decision making. The advances in the development of satellite technology as sources of data for GIS enhance the quality of data as well as its analysis capability. In the current and future years, more research in this area will contribute to the development of more technology to solve more real-world problems, either in the industry, government, academia, or social areas.

Cross-References

- ▶ Fundamentals of Remote Sensing Imaging and Preliminary Analysis
- ▶ Global Navigation Satellite Systems: Orbital Parameters, Time and Space Reference Systems and Signal Structures
- ▶ Processing and Applications of Remotely Sensed Data
- ▶ Remote Sensing Data Applications

References

D.J. Cook, L.B. Holder, Substructure discovery using minimum description length and background knowledge. J. Artif. Intell. Res. **1**, 231–255 (1994)

S.R. Gelati, *Geographic Information Systems Demystified* (Artech House, Boston, 2006)

P. Ghadirian, I.D. Bishop, Integration of augmented reality and GIS: a new approach to realistic landscape visualisation. Landsc. Urban Plan. **86**(3–4), 226–232 (2008)

J.A. Gonzalez, L. Altamirano, J.F. Robles, Data mining with context information for satellite image classification. Ambiencia **4**, 147–158 (2008)

ISO, International Organization for Standardization (2011), http://www.iso.org/iso/home.htm. Last visited 7 Apr 2011

V. Kanjilal, H. Liu, M. Schneider, Plateau regions: an implementation concept for fuzzy regions in spatial databases and GIS, in *Proceedings of the Computational Intelligence for Knowledge-Based Systems Design, and 13th International Conference on Information Processing and Management of Uncertainty (IPMU'10)*, (Springer, Berlin/Heidelberg, 2010), pp. 624–633

T.H. Kolbe, *Representing and Enhancing 3D City Models with CityGML. Lecture Notes in Geoinformation and Cartography (LNGC* (Springer, Berlin/Heidelberg, 2008)

K. Koperski, J. Adhikary, J. Han, Spatial data mining: progress and challenges, in *SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD96)*, Montreal, 1996

A. Koperski, J. Han, *Spatial Data Mining: Progress and Challenges. Research Issues on Data Mining and Knowledge Discovery*, Montreal, 1999

Y.-C. Lee, Geographical data and its acquisition, in *Geographic Data Acquisition*, ed. by Y.-Q. Chen, Y.-C. Lee (Springer, Wien, 2001a), pp. 1–8

Y.-C. Lee, Map projections, in *Geographic Data Acquisition*, ed. by Y.-Q. Chen, Y.-C. Lee (Springer, Wien, 2001b), pp. 43–63

X. Li, H.-J. Götze, Ellipsoid, geoid, gravity, geodesy, and geophysics. Geophysics **66**, 1660–1668 (2001)

P.A. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind, *Geographical Information Systems and Science*, 2nd edn. (Wiley, Chichester, 2005)

E. Mok, J.C.H. Chao, Coordinate systems and datum, in *Geographic Data Acquisition*, ed. by Y.-Q. Chen, Y.-C. Lee (Springer, Wien, 2001), pp. 11–24

S.H. Muggleton, Inverse entailment and progol. N. Gener. Comput. **13**, 245–286 (1995)

R.T. Ng, J. Han, Efficient and effective clustering methods for spatial data mining, in *Proceedings of the 20th International Conference on Very Large DataBases*, Santiago, 1994, pp. 144–155

M. Pech, A. Tchounikine, R. Laurini, J. Gonzalez, D. Sol, Graph-based representation for spatial data mining: a proposal, in *Proceedings of the CASSINI-SIGMA Conference*, Grenoble, 2004

G. Schall, E. Mendez, E. Kruijff, E. Veas, S. Junghanns, B. Reitinger, D. Schmalstieg, Handheld augmented reality for underground infrastructure visualization. Pers. Ubiquit. Comput. **13**(4), 281–291 (2009)

S. Shekhar, C.-T. Lu, P. Zhang, Detecting graph-based spatial outliers, in *Intelligent Data Analysis*, vol. 6 (IOS Press, Amsterdam, 2002), pp. 451–468

G. Taylor, G. Blewitt, *Intelligent Positioning: GIS-GPS Unification* (Wiley, Hoboken, 2006)

UN-SPIDER, United Nations platform for space-based information for disaster management and emergency response, http://www.un-spider.org/. Last visited 10 Apr 2011

Q. Weng, *Remote Sensing and GIS Integration, Theories, Methods, and Applications* (McGraw-Hill, New York, 2010)

G.G. Wilkinson, A review of current issues in the integration of GIS and remote sensing data. Int. J. Geogr. Inf. Syst. **10**, 85–101 (1996)

M. Wolf, H. Asche. Towards 3D tactical intelligence assessments for crime scene analysis, in *Computational Science and Its Applications: ICCSA 2010: International Conference, Fukuoka, Japan March 23–26, 2010: Proceedings Part I*, ed. by D. Taniar, O. Gervasi, B. Murgante, E. Pardede, B.O. Apduhan. Springer Lecture Notes in Computer Science (LNCS), vol. 6016 (Springer, Berlin, 2010), pp. 346–360

H. Xiong, S. Shashi, H. Yan, K. Vipin, M. Xiaobin, S.Y. Jin, A framework for discovering co-location patterns in data sets with extended spatial objects, in *Proceedings of the 2004 SIAM International Conference on Data Mining (SIAM2004)*, Florida, 2004, pp. 78–89

Developments in Hyperspectral Sensing

Su-Yin Tan

Contents

Abstract

Hyperspectral remote sensing is a relatively new development in remote sensing technologies, effectively measuring both spatial and high spectral information from surface materials and constituents within a single system. Compared to multispectral remote sensing, hyperspectral imagery can provide more accurate and detailed spectral information of the Earth's surface, measuring hundreds of bands from the visible to the near infrared. Hyperspectral data can be obtained using either space-based or airborne platforms, with expanding applications on unmanned aerial vehicles (UAVs). This chapter discusses hyperspectral technologies and their vast applications focusing primarily on airborne and spaceborne systems, reviewing past and future directions of sensor technology developments. Hyperspectral imaging is a rapidly growing field of space-based remote sensing

---

S.-Y. Tan (✉)
Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada
e-mail: su-yin.tan@uwaterloo.ca

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_101

and will continue to expand in utility for various civilian and public-good applications. Various nations are planning hyperspectral remote sensing missions, which will see increased acquisition of hyperspectral data of the Earth's surface on a more frequent and timely basis in the near future.

Keywords

Hyperspectral • Remote sensing • Imaging spectroscopy • Spectral feature analysis • Applications • Sensors • AVIRIS • CASI • Hyperion • CHRIS

Introduction

> Satellite imagery can be used retrospectively, meaning that the data collected by satellites today will probably help solve issues we are not currently even aware of—an advantage which is invaluable.
>
> – Nathalie Pettorelli, Zoological Society of London, Methods Blog, June 10, 2015

Hyperspectral remote sensing has become a powerful analytical tool for applications in environment, ecology, forestry, agriculture, and geoscience. Such sensors offer the ability to detect molecular absorption and spectral signatures of surface materials and constituents. Accurate estimation of land surface characteristics is necessary for a wide variety of applications. For example, quantitative estimation of vegetation biochemical and biophysical characteristics can be used in agricultural, ecological, and meteorological applications. Due to its global coverage, repetitiveness, and nondestructive observation of land surfaces, space-based hyperspectral remote sensing has been recognized as a reliable method of measuring specific land cover variables that would be difficult to assess using conventional multispectral sensors.

Hyperspectral remote sensing is a relatively new development in remote sensing technologies. It effectively combines imaging, spectroscopy, and remote sensing technologies within a single system. Compared to multispectral remote sensing, hyperspectral imagery can provide more accurate and detailed spectral information of the Earth's surface in a narrow wavelength of light. In general, hyperspectral remote sensors measure hundreds of bands from the visible to the near infrared. Hyperspectral imagery is often applied for object detection, geological mapping, and land cover classification.

This chapter discusses hyperspectral technologies and their vast applications. Hyperspectral imaging systems will first be explained, including the advantages they offer over multispectral imagers. We review both airborne and space-based systems and platforms, although more focus is on the latter. Finally, various applications where hyperspectral remote sensing has been effectively used are described, as well as future directions and developments of such technologies. Hyperspectral remote sensing is a rapidly growing field of satellite applications. Many applications can potentially take advantage of hyperspectral data for improved understanding of surface material spectral characteristics.

What Is Hyperspectral Sensing?

Hyperspectral remote sensing is a relatively new technology, but imaging spectroscopy has been in existence for over 100 years for identifying materials and their composition. Spectroscopy refers to the measurement of electromagnetic radiation intensity as a function of wavelength. It originates from the seventeenth century through the study of visible light dispersed by a prism according to its wavelength. Sir Isaac Newton first demonstrated that white light could be split into different component colors. Significant achievements in imaging spectroscopy were attributed to airborne instruments in the early 1980s and 1990s (Vane et al. 1984). Experimental advancements led to spectral measurement devices, which are now referred to as spectrometers, spectrophotometers, spectrographs, or spectral analyzers. The first space-based imaging spectrometer was launched in 1999 with the NASA Moderate-Resolution Imaging Spectroradiometer (MODIS).

In general, hyperspectral sensors provide detailed spectral information from every pixel in an image. The systems technology collects images of a scene in tens to hundreds of contiguous spectral bands nearly simultaneously and in a relatively narrow bandwidth (<10 nm) (Fig. 1). This is in contrast with multispectral sensors, which acquire about 5–10 spectral bands at a relatively wide spectral interval (>100 nm). Hyperspectral remote sensors acquire spatial and spectral information simultaneously from a distance, with the aim of providing the radiance and detailed spectrum for each pixel in an image. This allows for the construction of reflectance spectra that closely resemble those measured on laboratory instruments.

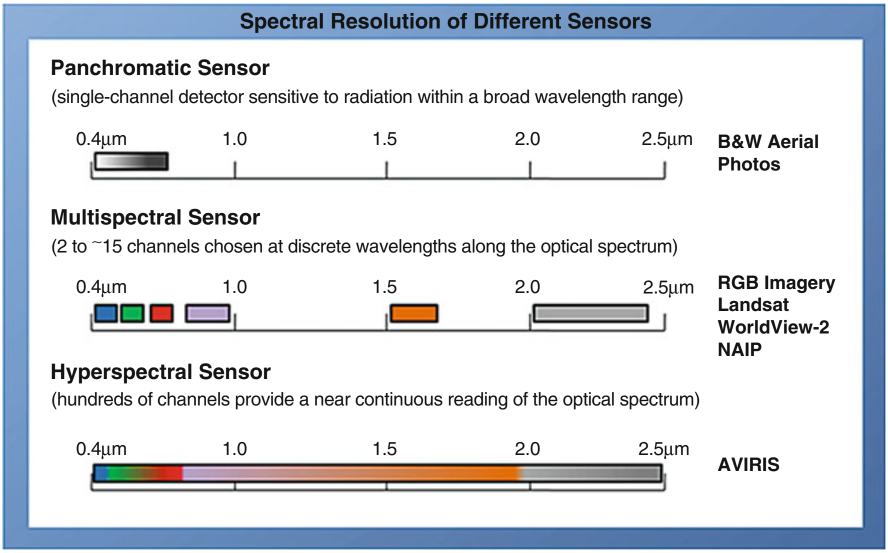


Fig. 1 Differences between hyperspectral and multispectral sensors (Exelis 2013)

Spectral-based analytical tools and software can then be used to interpret data collected from hyperspectral sensors. This allows for the direct identification and, sometimes, abundance determinations of individual materials based on their reflectance characteristics. This information enables targets to be identified based on the spectral behavior of its surface materials, including minerals, atmospheric constituent gases, vegetation, and water quality. While processing and evaluating information, it is also necessary to conduct ground measurements and to collect reference data for spectral libraries using conventional sampling.

Alexander Goetz developed the first truly portable field spectrometer in 1974 that used a charge-coupled device (CCD) for spectral applications. This was subsequently designed for aircraft and spacecraft, eventually operating successfully from the orbit in the Landsat program. A classical definition for hyperspectral remote sensing by Goetz et al. (1985) remains relevant today: “The acquisition of images in hundreds of continuous registered spectral bands such that for each pixel a radiant spectrum can be derived.” This definition essentially encompasses all spectral regions, including visible, near infrared, shortwave infrared, midwave infrared, and longwave infrared; all spatial domains and platforms, including ground, air, and space platforms; and all targets, including gas, liquid, and solid.

Hyperspectral imaging is not only characterized by a high number of bands but also by its high spectral resolution. This means that the sensor defines a narrow wavelength range for a particular channel or band, ultimately determining the ability of a sensor to resolve spectral features. The original accepted bandwidth for hyperspectral remote sensing was approximately 10 nm based on initial geological applications (Goetz 1987). However, narrower bandwidths have since become available, which has broadened hyperspectral imaging capabilities. New applications, such as assessing vegetation fluorescence, now require bandwidths of less than 1 nm (Guanter et al. 2006).

Hyperspectral remote sensing has been defined as “spatial spectrometry from afar,” acquiring large quantities of high-quality spectral data from both airborne and spaceborne platforms. The main aim is to extract physical information from raw data collected across the spectrum, which can be easily converted to describe inherent properties of surface targets, such as reflectance and emissivity. This technology has become an interdisciplinary field, including atmospheric science, computer science, aviation, engineering, statistics, and applied mathematics. High spectral resolution data is acquired simultaneously both spatially and temporally, thus providing a new dimension to remote sensing data.

Hyperspectral imagery has often been described as an “expert” geographic information system (GIS). It involves multiple layers of geocoded datasets that are used to generate thematic layers. Hyperspectral data contains both spatial and spectral information from materials within a given scene captured simultaneously. Each pixel across a sequence of continuous, narrow spectral bands contains both spatial and spectral properties. Pixels are sampled across many narrowband images by a scanning system at a particular spatial location, resulting in a “hyperspectral data cube” represented in Fig. 2. A spectral reflectance curve can be plotted as wavelength versus radiance or reflectance. This information can then be used to identify and characterize a particular feature within the scene, based on unique spectral signatures (Fig. 3).

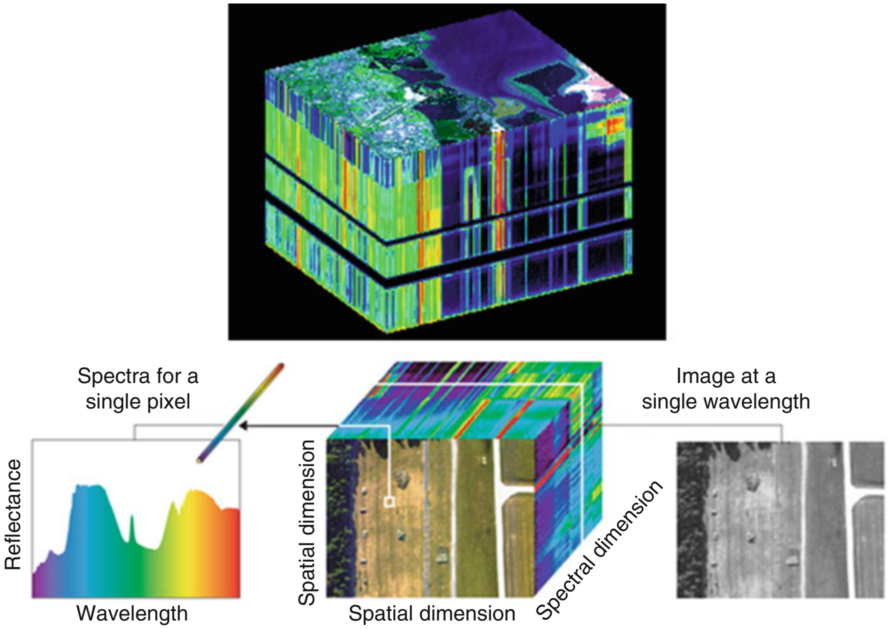


Fig. 2 Hyperspectral data cube containing all geographical and spectral data for each image pixel (Tamas Janos 2008)

Hyperspectral Sensors

Similar to other remote sensing systems, imaging spectrometers can be carried on satellites, aircraft, unmanned aerial vehicles (UAVs), or other platforms. This chapter focuses mainly on satellite hyperspectral sensors with some discussion of airborne sensors. Airborne hyperspectral remote sensing systems can provide a higher spatial and spectral resolution image. Unlike sensors on aircraft, sensors on satellites have the capacity to provide global coverage at regular intervals.

In general, the spectral range of airborne hyperspectral sensors is 380–12,700 nm and for satellite sensors is 400–1,400 nm. The AVIRIS airborne hyperspectral imaging sensor obtains spectral data over 224 continuous channels, each with a bandwidth of 10 nm over a spectral range from 400 to 2,500 nm. An example of an operational space-based hyperspectral imaging platform is the Air Force Research Lab's TacSat-3/ARTEMIS sensor, which has 400 continuous spectral channels, each with a bandwidth of 5 nm.

Ultraspectral sensors represent future developments in hyperspectral imaging technologies, having 1,000 s of spectral channels, each with a bandwidth narrower (less than 5 nm) than those of hyperspectral sensors. Such sensors allow for the quantitative assessment of scene materials, ranging from solids, liquids, to gases. For example, the abundance of gases or effluents can

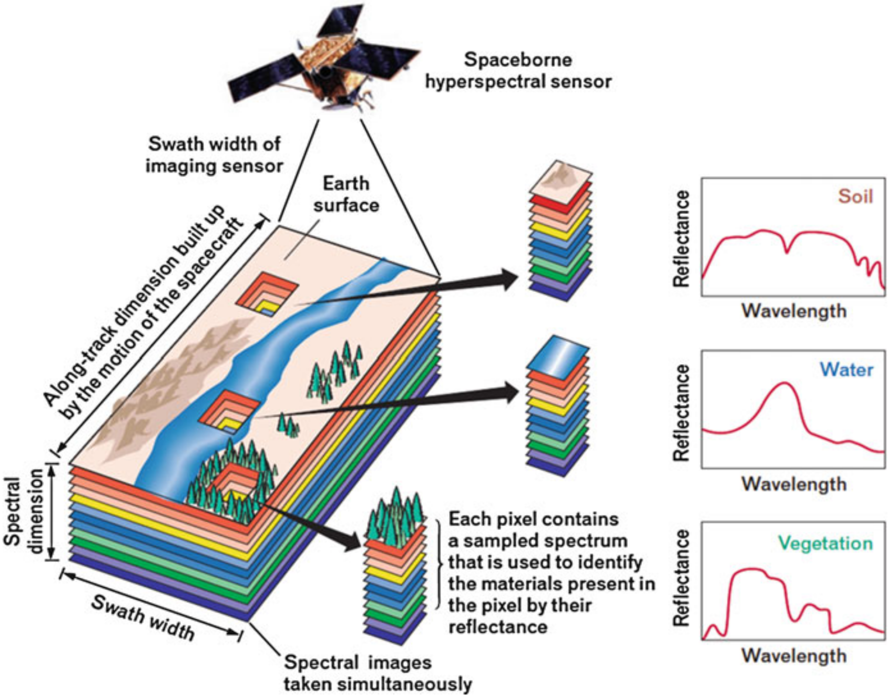


Fig. 3 Conceptual diagram of hyperspectral imaging (Shaw and Burke 2003)

potentially be determined from the width and strength of absorption features in a given spectrum (Fig. 4).

Airborne Hyperspectral Sensors

Airborne hyperspectral remote sensing systems carry an imaging spectrometer and measure electromagnetic radiation reflected from the Earth's surface. Most hyperspectral sensors are mounted on aerial platforms and less on satellite platforms. Airborne hyperspectral sensors offer the advantages of acquiring high spatial and spectral resolution imagery, although the spatial resolution depends on the height of the flight path and the spectral resolution depends on the imaging spectrometer used. Aircraft platforms also offer flexibility in adjusting image acquisition to account for weather conditions, solar illumination conditions, and cloud cover. Additional revisits may be planned for change detection, while sensor maintenance, repair, and configuration adjustments can easily be made to aircraft platforms compared to satellites.

Sensors scan the ground either in pushbroom or whiskbroom modes. Pushbroom scanners (also referred to as along-track scanners) consist of a line of sensors arranged perpendicular to the flight direction (Fig. 5). The one-dimensional sensor array captures an entire scan line at once and is often lighter, smaller, and less

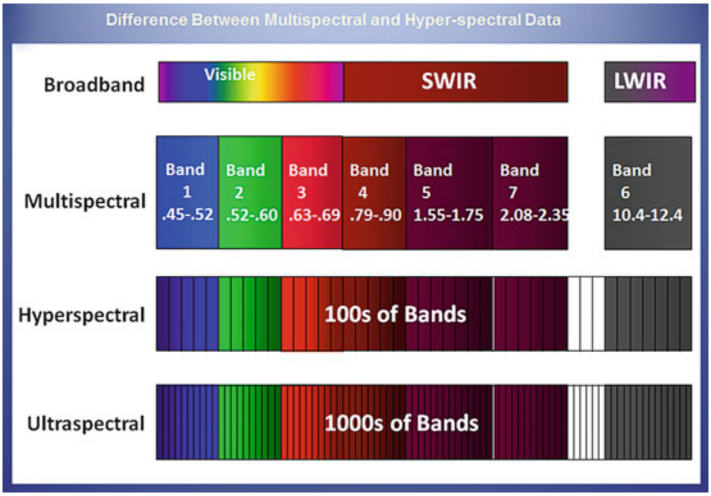


Fig. 4 A comparison of multispectral, hyperspectral, and ultraspectral remote sensing imaging (Sunil Kumar 2006)

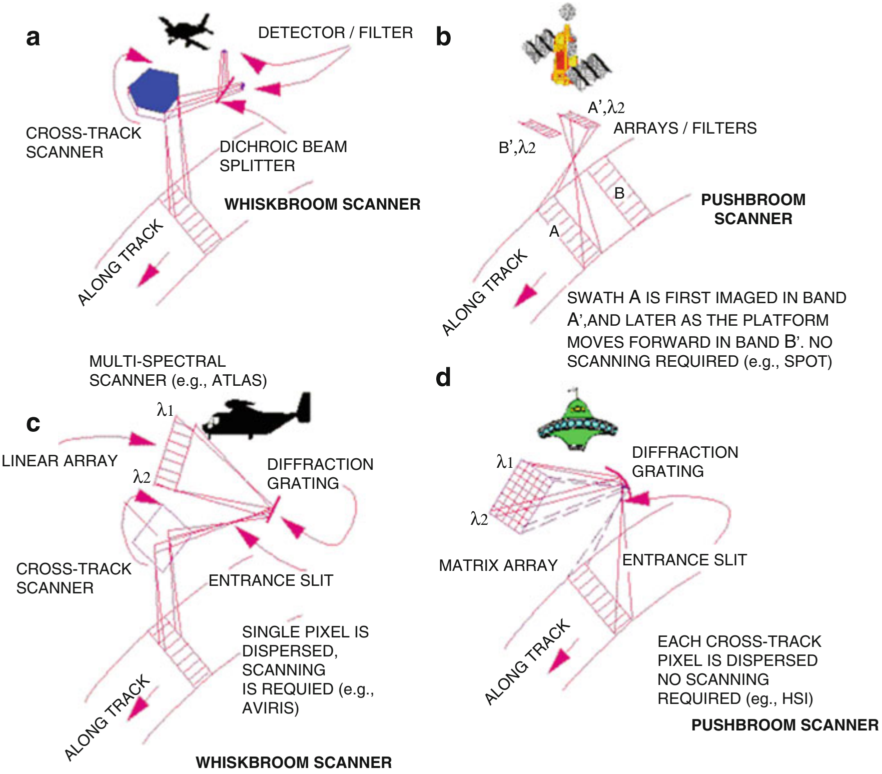


Fig. 5 Scanner platform systems for hyperspectral image acquisition (Ames Remote 1998)

complex than whiskbroom scanners. They tend to also have superior radiometric and spatial resolution, although calibration is required for the large number of detectors comprising the sensor system.

In comparison, whiskbroom scanners (or across-track scanners) use rotating mirrors to reflect light into a single detector that collects data one pixel at a time, scanning from side to side perpendicular to the direction of the sensor platform. The moving parts make such sensors expensive, large, and complex to build. Spatial distortions can result and rotating mirrors are prone to wearing out. Since whiskbroom scanners focus on a subsection of the full swath at any time, a higher image resolution can typically be achieved for the same size of scan swath of a pushbroom scanner. Furthermore, such systems have fewer sensor detectors that require calibration as compared to other scanner sensor systems.

Many airborne hyperspectral sensors have been developed and operated worldwide. The first airborne tests were by the airborne imaging spectrometer (AIS) conducted by NASA Jet Propulsion Laboratory (JPL) flown in November 1982. This program was spearheaded by Alex Goetz and Greg Vane. This first instrument consisted of a 32 x 32 mercury cadmium telluride detector array with a spectrometer grating system that provided 128 spectral bands operating in the 0.9–2.4 μm spectral range.

The first AIS flights were taken over the Cuprite mining district, Nevada, achieving unambiguous identification of kaolinite and alunite minerals, thus confirming that spectral reflectance data could successfully identify geologic minerals. Perhaps more importantly, during a second flight, a spectral "unknown" was discovered and identified as the mineral buddingtonite, which was thought to be an indicator mineral for gold deposits (later proven to be untrue). This marked the first discovery of a previously unknown mineral occurrence using a remote sensing technology.

These achievements led NASA to fully support research and development in imaging spectrometers, which began in 1983 under the leadership of Alex Goetz. This resulted in the development of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), which first flew in 1987. Furthermore, two space-based instruments were designed, the Shuttle Imaging Spectrometer Experiment (SISEX) and the High-Resolution Imaging Spectrometer (HIRIS). However neither instruments were actually built due to the Challenger disaster in 1986 and financial constraints, respectively. AVIRIS, however, became highly successful and a principal source of data for most of the hyperspectral research programs we know today. The AVIRIS instrument also continued to improve through major upgrades in response to advancing technology and new knowledge.

AVIRIS is a whiskbroom imager that uses a scanning system for acquiring data on the transverse direction of advancement. It can operate from a variety of aircraft including the high-altitude NASA ER-2 (Fig. 6). AVIRIS acquires 224 spectral bands or channels in the range of 0.40–2.50 μm with a spectral resolution of about 10 nm. Four off-axis double-pass Schmidt spectrometers capture light from fore optics using optical fiber that is sent to four linear panels, one for each spectrometer. The AVIRIS sensor can fly on NASA ER-2 and take images from an altitude of 20 km with a spatial resolution of 20 m and a swath width of 10.5 km. Starting in

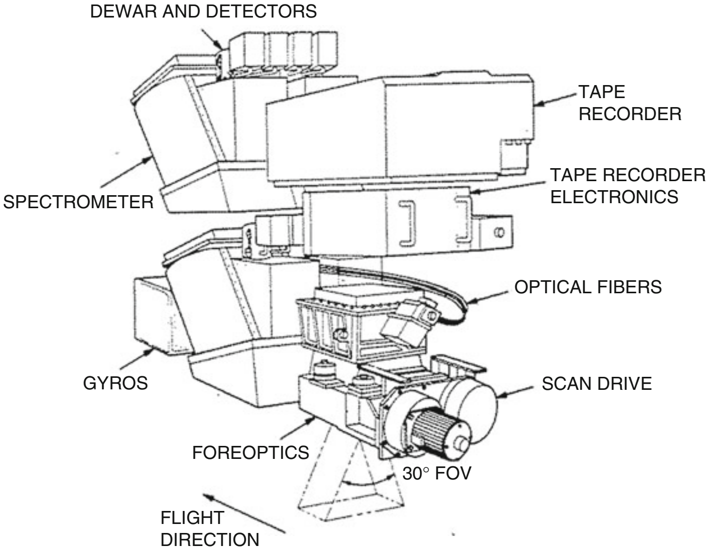


Fig. 6 The AVIRIS whiskbroom scanner (Ames Remote 1998)

1998, the sensor was mounted on a Twin Otter aircraft flying at low altitude, taking images with a spatial resolution that ranged between 2 and 4 m.

With the success of AVIRIS, many private companies began to develop their own imaging spectrometers (Table 1). Developed in 1989, the Canadian Compact Airborne Spectrographic Imager (CASI) is considered to be the first commercial imaging spectrometer and is now used in many countries (Fig. 7). Developed by the ITRES Corporation (www.itres.com), the imaging spectrometer covers the visible and near-infrared region with a spectral range of 0.4–1.0 μm. With a 1,500 pixel field of view, CASI can achieve spatial resolutions as high as 25 cm, depending on the altitude of the aircraft. The spectral bands that the instrument measures and the bandwidths used are all programmable to meet user specifications and requirements.

There has been a multitude of airborne hyperspectral imaging systems developed. This includes the Hyperspectral Digital Imagery Collection Experiment (HYDICE) developed by the US Naval Research Laboratory (NRL), which was a pioneering airborne imager first flown in 1995. HyMap (also known as Probe-1) is also a higher-quality airborne imager (www.hyvista.com) from Australia. Other commercial imaging spectrometers, which cover the full hyperspectral range (0.4–2.5 μm), are AISA from Finland, Probe-1 from Canada, and DAIS-7915 from Germany.

Experimentation with airborne imaging spectrometers has helped guide the development of hyperspectral sensor systems for advanced satellite systems. In fact, a satellite hyperspectral sensor usually has an airborne counterpart, which can be used for calibration and data processing purposes, as well as to develop application data products before the satellite is launched. Examples include the NRL

Table 1 Summary of airborne hyperspectral sensors

Name	Available date	Spectral region (μm)	Number of bands	Spectral interval (nm)	FOV (degrees)	Developer
FLI	1981	0.4–1.0	288	2.5	N/A	DFO, Canada
AIS	1983	1.2–2.4	128	10	90	NASA/JPL, USA
AVIRIS	1987	0.4–2.5	224	10	30	NASA/JPL, USA
CASI	1988	0.4–1.0	288	3	35.4	ITRES, Canada
AMSS	1989	0.4–1.05 2.05–2.4	20 44	32 8	92.16	Geoscan, Australia
AHS	1991	0.43–0.83 1.605–2.405	20 15	20 50	85.92	Daedalus, USA
ASAS	1992	0.4–1.0	68	10	19	NASA, USA
CHRISS	1992	0.425–0.85	125	3.4	10.3	SAIC, USA
ROSIS	1992	0.43–0.88	256	5	±16	DLR, Germany
AAHIS-1	1994	0.44–.835	108	11	193 mrad	SETS, USA
HYDICE	1994	0.4–2.5	210	~10	8.94	USA
AISA	1995	0.4–2.5	488 254	4.5–14	40	SPECIM, Finland
HyMap	1998	0.4 –2.5	128	15	61.3	Australia
Probe-1	1998	0.4–2.5	128	VNIR:11 SWIR:18	10	Canada
DAIS-7915	2000	0.4–2.5	72	0.9–60	78	GER, USA, and DLR, Germany
ARES	2005	0.4–2.5 8–12	128 32	15 130	N/A	Australia
APEX	2014	0.38–0.97 0.94–2.50	114 199	0.45–7.5 5–10	28	ESA, Switzerland, Belgian

PHILLS airborne sensor that supports the COAS imaging spectrometer on the NEMO satellite and the European Airborne PRISM Experiment (APEX) sensor that supports the satellite-borne PRISM sensor on the LSPIM mission.

Spaceborne Hyperspectral Sensors

Endeavors for spaceborne hyperspectral remote sensing systems came comparatively later than the development of airborne hyperspectral remote sensing systems. Unlike airborne sensors, sensors on satellites have the capacity to provide global coverage at regular intervals. In November 2000, NASA successfully launched the

Hyperion EO-1 sensor with the purpose of taking hyperspectral images from space for mineralogical mapping.

Hyperion is a hyperspectral satellite sensor that acquires 242 spectral channels, working in the spectral range of 0.40–2.50 μm with a spectral resolution of about 10 nm and a spatial resolution of 30 m. Hyperion is a pushbroom scanner that captures images at an altitude of 705 km with a swath width of 7.5 km and a high radiometric resolution of 8 bits. The system consists of two spectrometers to improve signal-to-noise ratio (SNR) with one functioning in the visible/near infrared (VNIR) (0.4–1.0 μm) and the other functioning in the shortwave infrared (SWIR) (0.9–2.5 μm). Hyperion was based on the heritage of the LEWIS Hyperspectral Imaging Instrument (HIS) and has demonstrated a wide range of applications in mining, geology, forestry, agriculture, and environmental management. The sensor has generated detailed classification of land assets, aiding remote mineral exploration, crop yield prediction and assessments, and containment mapping, thus demonstrating that space-based imaging spectroscopy enables a wide range of scientific applications (Fig. 7).

In 2001, the European Space Agency launched the Compact High-Resolution Imaging Spectrometer (CHRIS) flown on the PROBA-1 mission. The objective of the mission was to collect bidirectional reflectance distribution function (BRDF) data for better understanding of spectral reflectances, as well as testing the

Fig. 7 Hyperion hyperspectral imagery has enabled researchers to differentiate between minerals and rocks in identifying ancient copper mines and smelting sites near Khirbat en-Nahas, Jordan. The top image is a natural color image, and the lower image is a false-color image, identifying different rock types (NASA 2016)

Fig. 8 CHRIS image of Dongting Hu (Dongting Lake) in China, acquired on February 12, 2015 from the PROBA-1 satellite (ESA 2015)

capabilities of imaging spectrometers on agile small satellite platforms (Fig. 8). CHRIS provides 19 spectral bands in the VNIR range only (0.40–1.05 μm) at a ground sampling distance of 17 m. CHRIS is a unique sensor due to the fact that its spectral resolution can be modified or reconfigured from 19 bands up to 63 bands or channels. As the number of bands increases, the spatial resolution of the imagery decreases (i.e., the greater the number of bands, the bigger the pixel size). The finest spatial resolution produced by CHRIS is 18 m, while the coarsest spatial resolution reaches 36 m. Another important characteristic of this sensor is that each nominal image is observed by five consecutive pushbroom scans by single-line array detectors, acquiring up to five images of each target during every acquisition sequence.

Hyperion and CHRIS were significant milestones. They demonstrated the value of working with complex hyperspectral data to the scientific community and influenced the development of subsequent space-based hyperspectral remote sensing missions. As technology demonstrators, however, both Hyperion and CHRIS were plagued by poor data quality issues and did not collect global coverage. Experience with spaceborne hyperspectral remote sensing has fueled arguments that hyperspectral imaging may be best left to airborne platforms, which fly below the atmosphere. A period of technological stagnation followed Hyperion and CHRIS with only renewed interest in hyperspectral spaceborne missions occurring during the mid-2000s. Successful missions have since included the Indian HySI, Chinese HJ-1A, and NASA's HICO, which all operate over the spectral region from 0.4 to 0.95 μm (VNIR). Launched in April 2008, ISRO's HySI hyperspectral imager is a prototype pushbroom instrument that provides a total of 64 spectral bands and is used principally for resource characterization and detailed studies.

The Chinese Pushbroom Hyperspectral Imager (PHI) and Operative Modular Imaging Spectrometer (OMIS) are two representative hyperspectral imagers developed by the Chinese Academy of Sciences. China launched HJ-1A and HJ-1B small satellites and a new hyperspectrometer HIS onboard the HJ-1A satellite on September 5, 2008. This sensor conducts repeated global monitoring at a $\pm 30^{\circ}$ side-viewing angle with a 96 h revisiting cycle. Imagery is captured at a 50 km swath width and 115 spectral bands covering a spectral range of 0.45–0.95 μm with a spectral resolution of 4.32 nm and 100 m spatial resolution. Although narrower in spectral range when compared to the EO-1 Hyperion, the HJ-1/HIS spectrometer has improved spectral resolution for better ground feature identification and information extraction. It provides another valuable tool for developing quantitative research and application, such as atmospheric composition detection, water environment monitoring, and vegetation growth monitoring.

NASA's Hyperspectral Imager for the Coastal Ocean (HICO) is an imaging spectrometer based on the PHILLS airborne imaging spectrometers and was the first spaceborne imaging spectrometer designed for coastal ocean research. Sponsored initially by the Office of Naval Research and then by NASA during its final 2 years of operation, HICO was developed to demonstrate improved coastal remote sensing products, including bathymetry, bottom types, water optical properties, and onshore vegetation maps. The sensor has a spatial resolution of $<100 \times 100$ m, spectral range of 0.40–0.9 μm, and spectral resolution of 5.7 nm. HICO provides enhanced products at a reduced cost by adapting proven aircraft imager architecture and using commercial off-the-shelf components. HICO was installed on the International Space Station (ISS) on September 23, 2009 and collected over 10,000 images during its first 5 years of operation until an X-class solar storm resulted in its permanent failure in September 2014.

In the next decade, various nations are planning missions for hyperspectral remote sensing of the Earth's surface on a timely and frequent basis. This includes Germany's EnMAP satellite led by the German Research Centre for Geosciences and managed by DLR, which is envisaged for launch in 2018. Other examples include the Brazilian-American Flora Hiperspectral satellite, Italy's PRISMA satellite, India's TWSat with the HYSI-T coarse hyperspectral imager (currently in operation), US Air Force TacSat-3, South Africa's MSMI sensor, NASA JPL HyspIRI mission, ESA FLEX mission, and Canada's HERO imaging system for developing and delivering hyperspectral products.

Future plans may also include incorporating hyperspectral capabilities in future satellites of the Landsat program. Future EO constellations, such as the UK Disaster Monitoring Constellation (DMC) and Germany's RapidEye constellation, may also include hyperspectral sensors. Finally, wide-area synoptic sensors derived from MODIS and MERIS may incorporate hyperspectral capabilities. China, Israel, and other countries, including Canada, have potential to operate hyperspectral EO sensors and have missions currently under construction and to be launched in the future (Table 2).

Table 2 Summary of space-based hyperspectral remote sensing sensors

Name	Available date	Spectral region (μm)	Number of bands	Spectral interval (nm)	GSD (m)	Swath width (km)	Developer
HIRIS	1994	0.400–2.500	192	11	30	30	NASA/JPL, USA
Hyperion	2000	0.356–2.576	242	10	30	7.5	NASA/JPL, USA
CHRIS	2001	0.415–1.050	19/63	5–12	17/34	13	ESA, UK
HySI	2008	0.400–0.950	64	8	506	130	ISRO, India
HJ-1A	2008	0.459–0.950	115	5	100	51	CAST, China
HICO	2009	0.380–0.960	128	5.7	90	42	NASA/ONR, USA
Flora	2016	0.400–2.500	200	10	30	150	NASA/JPL, USA, and INPE, Brazil
PRISMA	2017	VNIR:0.40–1.01 SWIR:0.92–2.05	VNIR:66 SWIR:171	10	30	30	ASI, Italy
EnMAP	2018	VNIR:0.42–1.00 SWIR:0.90–2.45	VNIR:89 SWIR:155	VNIR:8.1 + 1 SWIR:12.5 + 1.5	30	30	GFZ/DLR, Germany
HISUI-ALOS-3	2018	VNIR:0.40–0.97 SWIR:0.90–2.50	VNIR:57 SWIR:128	VNIR:10 SWIR:12.5	30	30	Japan
HYPXIM-CB	2018	0.400–2.500	N/A	14	15	15	CNES, France
HYPXIM-CA	2020	0.400–2.500	N/A	10	1	30	CNES, France
HyspIRI	2020	0.380–2.500	>200	10	30	30	NASA/JPL, USA
HERO	>2016	0.400–2.500	>200	10	30	30	CSA, Canada
MSMI	>2016	0.400–2.350	200	10	15	15	SunSpace, South Africa

Applications and Future Developments

There are a multitude of applications in which hyperspectral images can be used, including agriculture, forestry, geology, and environmental monitoring. Hyperspectral remote sensing imagery can be applied for detecting specific known target materials, such as vegetation species, soil properties, and geologic minerals. Target detection of ground features in remote sensing imagery depends primarily on the features' spectral characteristics; therefore, an understanding of the spectral signatures of ground features is critical. By acquiring high spectral resolution data, hyperspectral imagery enables for detailed applications on (a) target detection, (b) material mapping, (c) material identification, and (d) mapping details of surface properties.

In many cases, hyperspectral sensors can obtain more detailed information about surface materials that is not possible to achieve with multispectral satellite images. For example, hyperspectral imagery can be used to determine the chemical concentrations in leaves, identify vegetation stress, map the occurrence of plant species, detect surface contamination by mining waste, and map the presence of microorganisms and pollution in water bodies. This can sometimes involve distinguishing targets from very similar backgrounds or locating examples of targets that are smaller than the nominal pixel size. For example, a multispectral sensor can map the presence or absence of vegetation in a forest scene, whereas hyperspectral imagery can potentially map the distribution of vegetation species and the health or abundance of vegetation biomass.

Hyperspectral sensors are characterized by its bandwidth and continuous nature of spectral information. The contiguous spectral signatures enable acquisition of detailed data on materials and other details of ground features. Note that a sensor acquiring imagery in 20 bands may be hyperspectral if all bands are adjacent and with a 10 nm width. In contrast, if the sensor operates in 20 bands with a width of 100 nm or nonadjacent bands, the sensor would not be considered to be hyperspectral. Continuous spectral reflectance curves record the reflectance of soil, water, and vegetation, as well as details of absorption, allowing for rigorous analysis of surface composition.

As shown in Fig. 9, many hyperspectral imaging applications exist over different regions of the electromagnetic spectrum, ranging from the VNIR to LWIR. Military applications and target detection projects may use hyperspectral sensors operating in both NIR and SWIR spectral bands to discriminate camouflage material from background vegetation. For example, hyperspectral imagery used by military personnel to detect military vehicles under partial vegetation canopy may make use of significant differences between camouflage and plant material in the SWIR based on differences in moisture content, which may not be apparent in other parts of the spectrum.

Similarly, spectral characteristics of oil seeps and oil-contaminated soils are generally too subtle to be detected by traditional multispectral sensors. Hyperspectral sensors have sufficient spectral resolution to identify small amounts of hydrocarbon-based material through their spectral signatures. Hyperspectral remote sensing has

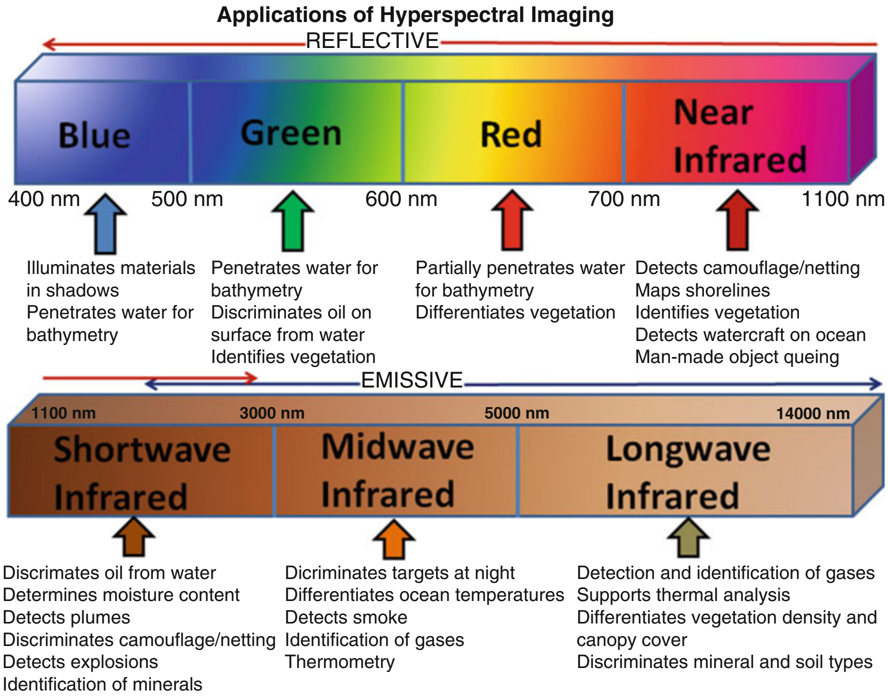


Fig. 9 Applications of hyperspectral imaging (Elowitz 2015)

also been used to map heavy metals and other toxic wastes within active and historic mining districts, including superfund sites. Both airborne and space-based hyperspectral remote sensing have demonstrated potential for precision agriculture applications, such as detecting crop stress and diagnosing crop disease that are not visible to the naked eye.

Different methods of analyzing the spectral information in hyperspectral data exist, such as comparing the pixel spectrum with a set of spectra taken from a well-known spectral library. This allows the user to identify specific surface materials, such as chlorophyll, dissolved organics, atmospheric constituents, and environmental contaminants. Therefore, the spectral recognition of targets using their spectral signature as a footprint and on the spectral analysis of absorption features enables quantitative assessment of surface materials in sites of interest (Fig. 10). With extensive availability of airborne sensors, this is offering huge potential in the areas of hydrology, disaster management, urban mapping, atmospheric studies, fisheries and oceans, and national security. These are only a few of the application areas, where the use of hyperspectral imagery is becoming more commonplace.

Traditional methods for landscape-scale vegetation mapping have required conducting labor and time intensive surveys and fieldwork. Vegetation tends to exhibit strong absorptions in the blue (0.45 μm) and red (0.67 μm) wavelengths due to the presence of chlorophyll. Variable leaf structure and canopy shadows may

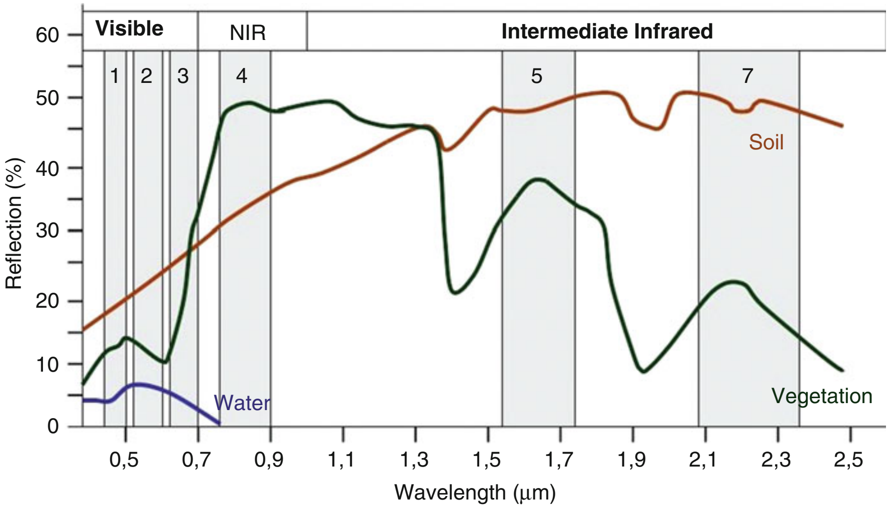
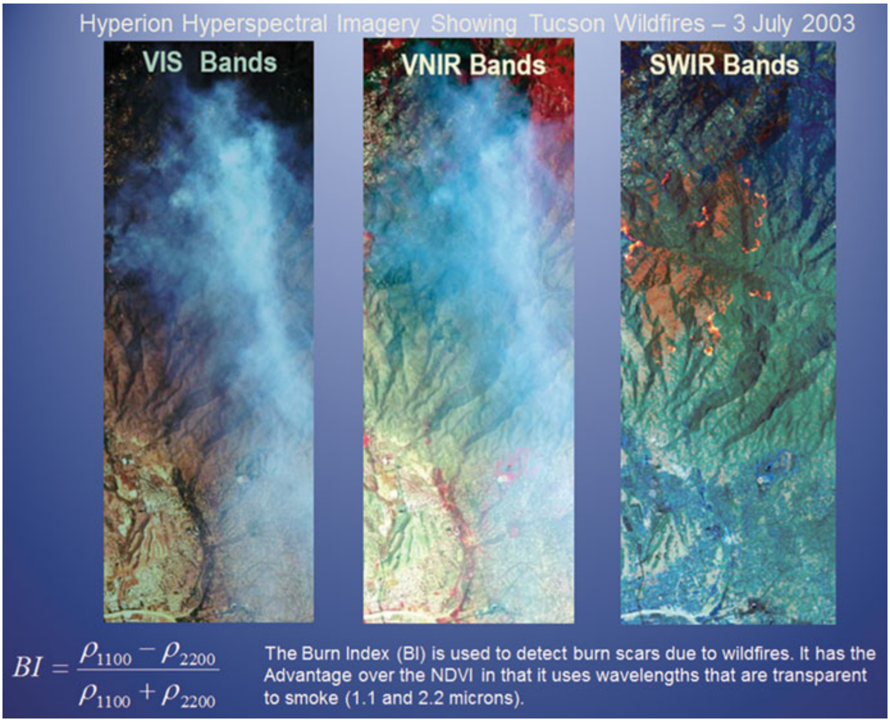


Fig. 10 Spectral signatures of water, vegetation, and soil targets (EARSeL 2016)

also result in different spectral signatures that enable vegetation species to be distinguished. Hence, hyperspectral data have enhanced potential for classifying and mapping land use and vegetation, providing detailed accurate products in a time- and cost-effective manner. For example, the airborne CASI sensor has been used for precision agriculture applications, mapping the chlorophyll status of corn crops using the 550, 670, 701, and 800 nm bands. Hyperion space-based imagery has been used to assess burn scars and hot spots through smoke resulting from forest wildfires (Fig. 11). Smoke tends to be more transparent in SWIR bands than in VNIR bands, and a burn index can be used to assess the severity of burn scar and vegetation damage.

Hyperspectral remote sensing imagery has been used for the detection and mapping of a wide range of soil properties, including moisture, salinity, and organic content. For example, Hyperion imagery has been employed for detecting soil organic carbon content based on 152 bands sensing from 0.43 to 2.36 μm. Detailed geospatial reflectance data of soil properties at field and landscape scales is important for understanding the dynamics of agricultural ecosystems. High-resolution maps of soil properties can be developed, thus overcoming the inaccuracies associated with interpolating in-situ soil test data. Hyperspectral technology can forecast natural hazards, such as mapping the variability of soil properties and linking them to landslide events and environmental disturbances.

Another application of hyperspectral remote sensing imagery is geologic mapping, such as mineral and lithological map production, which is made possible from the collection of continuous spectral bands. Maps of primary rock-type indicators and mineralized environments can be produced. Figure 12 summarizes the spectral absorption of minerals. Airborne hyperspectral remote sensing systems, such as AVIRIS and HyMap, are frequently used for mineral mapping. Space-based



$$BI = \frac{\rho_{1100} - \rho_{2200}}{\rho_{1100} + \rho_{2200}}$$

The Burn Index (BI) is used to detect burn scars due to wildfires. It has the Advantage over the NDVI in that it uses wavelengths that are transparent to smoke (1.1 and 2.2 microns).

Fig. 11 Hyperion hyperspectral imagery showing Tucson, Arizona, forest wildfires on July 3, 2003 (Elowitz 2015)

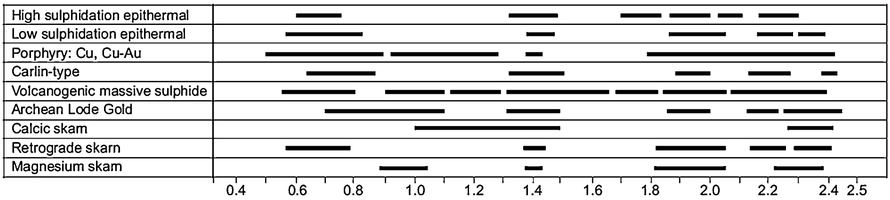


Fig. 12 Spectral absorption (indicated by the bars) of minerals (Van der Meer et al. 2012)

hyperspectral remote sensing can also provide valuable information for mineral mapping, especially for geographic areas that are inaccessible by aircraft sensors, but can also be hindered by lower signal-to-noise ratios.

Separation of signal from noise is important for spectral data processing. Hence, correction of atmospheric effects is the most critical processing step in hyperspectral data analysis. Poor atmospheric correction can result in false positives when the data is analyzed using various techniques. One of the first processing steps is to separate noisy spectral bands from the data and to eliminate highly redundant spectral bands typical of hyperspectral data. This can be accomplished with minimum noise fraction transformations, which is essentially based on a principle component transformation. This process reduces the dimensionality of the hyperspectral datasets, thus facilitating faster processing by computer software. Furthermore, geometric adjustments are often required to derive a geospatially representative map of target reflectance spectra. The large amount of data collected by a hyperspectral sensor often poses as a hindrance to its processing and analysis. Therefore, improvements in data mining techniques may benefit spectral mapping methods for producing final hyperspectral products in the future.

In summary, several challenges exist today, which prevent hyperspectral imaging technologies from moving toward more frequent operational use. This includes lack of reliable data sources with a high signal-to-noise ratio in order to retrieve the desired information and temporal coverage of a region of interest. Although analytical tools and software are now readily available, robust automated procedures for data processing are still lacking.

Hyperspectral imaging is also becoming a big data issue due to the large amounts of data and bands collected by sensors. There is no standardization for data quality, validation, or accuracy assessment. Currently, the processing and analysis of hyperspectral data are too labor intensive and difficult to derive. Spaceborne hyperspectral datasets are also expensive to attain and not freely available, while airborne hyperspectral data is more available commercially. Access to hyperspectral remote sensing data may change with new inexpensive sensors and upcoming unmanned aerial vehicle (UAV) platforms, which potentially offer an inexpensive and easy option for hyperspectral image acquisition.

Conclusion

In conclusion, hyperspectral imaging is a relatively recent remote sensing technology with a growing user community and rapidly expanding applications. It has the ability to detect molecular absorption and particle scattering signatures of surface constituents, which cause reflective or emissive signatures. It is now possible to acquire high-quality hyperspectral data from both aircraft and space-based systems. Global coverage from hyperspectral satellite systems is on the way, as well as improved atmospheric correction, reduced signal-to-noise ratios, and more sophisticated analysis software to assist scientists in using hyperspectral data for various applications.

The availability of hyperspectral data has overcome the constraints and limitations of low spectral and spatial resolution imagery and discreet spectral signatures. Hyperspectral images can now provide continuous high spectral resolution data collecting information from many narrow spectral bands, thus enabling detailed applications and analysis for target detection, material mapping, and identification of surface properties. Future developments in hyperspectral imaging include moving toward using active hyperspectral imaging techniques, where the imaging system provides its own source of controlled illumination, eliminating illumination artifacts and shadows that occur in passive systems. Big data analytical methods and modeling will also benefit the analysis and processing of hyperspectral data. There is also a movement toward even higher spectral resolution systems, such as ultraspectral imagers, which will define 1,000 s of spectral channels and allow for even finer quantitative assessment of scene materials, including solids, liquids, and gases.

Almost every developed country now has invested interest in having hyperspectral sensors operating in space. China, Israel, and other countries, including Canada, are currently designing and developing new hyperspectral sensors operating in space. New sensors operating within the next decadal time frame will likely include higher spectral resolution, more spectral bands, the whole Earth imaging, and image acquisition achieved within a short time period or even in real time. It is anticipated that hyperspectral imaging will be achieved with a variety of platforms, including aircraft, unmanned aerial vehicles (UAVs), satellites, and other platforms as it continues to expand in utility for various civilian and public-good applications.

Cross-References

► Electro-Optical and Hyperspectral Remote Sensing
► Remote Sensing Data Applications

References

Ames Remote, Precision Agriculture remote Sensing Information (1998), http://www.amesremote.com/. Accessed 6 Mar 2016

EARSeL (European Association of Remote Sensing Laboratories), Science Education through Earth Observation for High Schools (SEOS) (2016), http://www.seos-project.eu/modules/remotesensing/remotesensing-c01-p05.html. Accessed 6 Mar 2016

M. Elowitz, What is Imaging Spectroscopy (Hyperspectral Imaging)? (2015), http://www.markelowitz.com/. Accessed 6 Mar2016

ESA, Mission News: Proba-1 back in operation (2015), https://earth.esa.int/web/guest/missions/mission-news/-/article/proba-1-back-in-operation. Accessed 6 Mar 2016

Exelis, Vegetation analysis using vegetation indices in ENVI (2013), http://www.exelisvis.com/Learn/WhitepapersDetail/TabId/802/ArtMID/2627/ArticleID/13742/Vegetation-Analysis-Using-Vegetation-Indices-in-ENVI.aspx. Accessed 6 Mar 2016

A. Goetz, The portable instant display and analysis spectrometer (PIDAS), in *Proceedings of the Third Airborne Imaging Spectrometer Data Analysis Workshop*, vol. 87–30 (JPL Publication, Pasadena, California, 1987), pp. 8–17

A. Goetz, G. Vane, J. Solomon, B. Rock, Imaging spectrometry for Earth remote sensing. Science **228**, 1147–1153 (1985)

L. Guanter, V. Estellés, J. Moreno, Spectral calibration and atmospheric correction of ultra–fine spectral and spatial resolution remote sensing data. Application to CASI–1500 data. Remote Sens. Environ. **109**, 54–65 (2006)

NASA, Hyperion (2016), http://earthobservatory.nasa.gov/Features/EO1Tenth/page3.php. Accessed 6 Mar 2016

G.A. Shaw, H.K. Burke, Spectral imaging for remote sensing. Lincoln Lab. J. **14**(1), 1–28 (2003)

M. Sunil Kumar, *Multispectral and Hyperspectral Remote Sensing and Its Applications*. Bapatla Class Seminar, Agricultural College, by Medida Sunil Kumar BAD-14-06 1 (2006)

F.T. Tamas Janos, Geoinformatics (2008), http://www.tankonyvtar.hu/en/tartalom/tamop425/0032_terinformatika/index.html. Accessed 6 Mar 2016

F.D. Van der Meer, H.M.A. van der Werff, F.J.A. van Ruitenbeek, C.A. Hecker, W.H. Bakker, M.F. Noomen, M. van der Meijde, E.J.M. Carranza, J. Boudewijn de Smeth, T. Woldai, Multi- and hyperspectral geologic remote sensing: a review. Int. J. Appl. Earth Obs. Geoinf. **14**(1), 112–128 (2012)

G. Vane, M. Chrisp, H. Enmark, S. Macenka, J. Solomon, Airborne Visible/Infrared Imaging Spectrometer (AVIRIS): an advanced tool for earth remote sensing, in *Proceedings of the 1984 IEEE International Geoscience Remote Sensing Symposium, SP215* (IEEE, New York, 1984), pp. 751–757

Part IV

Space Systems for Meteorology

Introduction to Space Systems for Meteorology

Joseph N. Pelton, Scott Madry, and Sergio Camacho-Lara

Contents

Abstract

The world's meteorological satellite systems are today vital to every nation in the world not only for reliable weather forecasts but also for key storm warnings and potential disaster alerts in the case of hurricanes, tornadoes, typhoons, monsoons, floods, and other violent and potentially lethal meteorological events. Since the advent of polar-orbiting and geosynchronous meteorological satellites, the ability to predict weather accurately and reliably, forever longer periods of time, has increased to a remarkable extent. With a diverse suite of sophisticated instruments, meteorological satellites also gather essential data for climate change studies as well. The first systems were pioneered by the USA and by NASA

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

S. Madry
Global Space Institute, Chapel Hill, NC, USA
e-mail: Scottmadry@mindspring.com

S. Camacho-Lara
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Santa María Tonantzintla, Puebla, México
e-mail: sergio.camacho@inaoep.mx

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_52

experimental satellites, but over time Europe, Russia, Japan, China, and India have evolved increased capabilities to monitor weather systems using increasingly sophisticated imaging systems. Today, this allows effective sharing of meteorological data on a global scale. The combination of polar orbiting and geosynchronous satellites has allowed higher resolution images to be combined into accurate regional and global displays to see broad patterns of weather formations. New capabilities such as lightning intensity mapping have allowed greater capability to predict storm patterns in near real time and thus identify rapidly where the most violent and more energetic storm fronts are headed.

The global sharing of meteorological data and the combined imaging of international meteorological satellites have not only greatly contributed to effective long range weather forecasting on land and in the oceans but have also allowed the collection of data to monitor longer-term conditions associated with climate change, global warming, and increases in aridity in some regions and increased rainfall in others. In short, meteorological satellites have evolved, particularly in the last decade, to serve an important role in not only short- to medium-term weather forecasting but also to provide important data with regard to national, regional, and global environmental assessment and analysis including the major ocean conditions known as La Niña and El Niño.

The value of meteorological satellite systems and other Earth observation satellite systems for measuring the internationally recognized essential climate variables (ECVs) and for monitoring changes on land, oceans, and atmosphere is greatly increased when the acquired data is made available to national and international user organizations. A study by the International Academy of Astronautics recommends, among other things, that space agencies, companies, universities and nongovernmental organizations, and other international bodies already acting for the coordination of space agencies in the area of climate monitoring should work together to guarantee over time the continuous operational availability of the space sensors and datasets that are necessary for the monitoring of the space-observable ECVs (International Academy of Astronautics, in *Study on Space Applications in Climate Change and Green Systems: The Need for International Cooperation*, November 2010, eds. J.C. Mankins, M. Grimard, Y. Horikawa, ISBN EAN 9782917761113, pp. 15–21). Such cooperation is aimed at reinforcing the programmatic coordination of the Earth science programs worldwide, in the frame of institutions such as the Group on Earth Observations (GEO) and the Committee on Earth Observation Satellites (CEOS), with the goal of guaranteeing the continuous long-term availability for all nations of all space-observable ECV, as defined by the global climate observing system (GCOS); and contribute to the elaboration and implementation of GEO data sharing principles (http://www.earthobservations.org/art_015_002.shtml).

Despite patterns of data sharing and international cooperation with regard to data collection by meteorological satellite systems, there are limits to full disclosure of all satellite data. In particular, there remain certain areas of strategic concern in the context of possible military or defense-related use of meteorological and remote sensing data in the case of hostilities. It is partially because of these strategic and national defense-related concerns that so many different meteorological satellite systems have been deployed and why some parts of the imaging might be encrypted in a manner so that all data that is collected may not be in all instances fully shared. Despite these strategic concerns and interests, most meteorological satellite imaging data is today widely shared and global cooperation is quite universal.

The areas of satellite communications, remote sensing, and satellite navigation have all – to some degree – evolved toward more commercialized economic models and thus have become more oriented to competitive markets. This is not to say that these other space applications services are fully commercialized since there are some well-defined military, defense, and governmental services for satellite communications, remote sensing, and satellite navigation that remain as "publicly provided services." In the case of meteorological satellites, however, these space applications remain almost entirely as governmental services.

Despite discussions and analysis of how meteorological services might transition to commercial service providers within the USA, the provision of space-based meteorological services seems likely to remain as essentially a "public good" and not completely commercialized in any spacefaring nation. Although many countries, private businesses, and individuals derive major benefits from meteorological satellite images, no viable economic model has yet evolved whereby these services might be fully commercialized.

The US National Oceanic and Atmospheric Administration (NOAA), however, is currently evolving a new commercial space policy that mentions possible conditions whereby either "hosted payloads" on commercial satellites or certain routine, operational space functions might be transferred to the commercial space sector – perhaps on a temporary basis. The particular focus in this regard refers to potential future gaps in US polar orbiting meteorological satellite coverage that could occur in 2017. This potential commercialization of some meteorological satellite services in the U.S. is discussed later in this section.

In the chapters that follow specific information about various national and regional meteorological satellite systems will be presented. This introductory chapter provides a quick overview of the various systems that have evolved over time and some information perspective on how these systems are coordinated and work together through the World Meteorological Organization (WMO) and the World Weather Watch (WWW) program (World Meteorological Organization http://www.wmo.int/pages/index_en.html also see "Lessons Learned about the Integrated Global Observing Strategy through the World Weather Watch" http://www.un.org/earthwatch/about/docs/igusland.htm).

Keywords

Chinese Feng-Yang system • Committee on Earth observation satellites (CEOS) • Data and information service (NESDIS) • Essential climate variables (ECVs) • EumetsatMetOps polar orbiting system • European organisation for the exploitation of meteorological satellites (Eumetsat) • Geostationary lightning mapper (GLM) • Geostationary operational environmental satellites (GOES) • Indian

INSAT Japanese GMS or Himawari system • Japanese MTSAT satellites • Joint polar satellite system (JPSS) • Landsat • METEOR satellites • Meteosat • National Aeronautics and Space Administration (NASA) • National Atmospheric and Oceanic Administration • National environmental satellite • NIMBUS • Polar orbiting operational environmental satellites (POES) • Radiometer • Russian geostationary operational meteorological satellite (GOMS) system • TIROS system • US Defense Weather Satellite Service • World Meteorological Organization (WMO) • World weather watch (WWW)

Introduction

The first application satellites were communication satellites that quickly expanded to fulfill a strong demand – especially for international and overseas telecommunications. The advent of meteorological satellites quickly followed. The TIROS satellites, developed and launched by NASA, quickly demonstrated that meteorological satellites could be used to detect weather patterns and to forecast weather much more accurately. This was followed by the Nimbus Landsat and GOES satellite systems deployed by the USA to capture more and more detailed and up-to-date weather information.[1]

The success of these first meteorological satellite systems led to the design and deployment of both polar-orbiting (i.e., close to Earth and sun-synchronous) satellites and geostationary satellites that provide synoptic overviews of broad aspects of the Earth's surface by a number of the spacefaring nations. The increase in the number of meteorological satellites in both of these orbits plus the advancing of sensor technology – to obtain higher resolution images in the infrared and in multispectral frequency ranges – have allowed meteorologists to develop more sophisticated modeling of weather patterns. Over time the advent of meteorological satellites has allowed the development of more accurate short-, medium-, and even longer-term weather forecasts. Most recently, the evolution of this technology and its interpretation has allowed meteorological satellites to be applied to not only weather forecasting but to an array of environmental purposes that include monitoring of atmospheric and oceanic pollution, changes to the polar ice caps and glacial coverage, changes to the protective ozone layer, and broad patterns of climatic changes and global warming. Thus, today's meteorological satellites are in many cases environmental monitoring satellites that can provide crucial information as to methane release from the frozen peat fields in Siberia, pollution in the wetlands of the US Atlantic coast, increases in desertification around the world, or changes to the "ozone holes" in the polar regions of the planet.

---

[1] See chapter "▶ United States Meteorological Satellite Program" for further details on the early history of satellite communications.

Advances in Meteorological Satellite Technology

The initial polar-orbiting satellites that revolved around the Earth on a sun-synchronous basis provided updated information on weather conditions approximately every 90 min, but the coverage was for only a narrow strip of the Earth's surface and data in a particular location was updated only once every 24 h. The advantage was much higher resolution than the geostationary satellites that are about 40 times further away from the Earth than the polar-orbiting satellites in 800-km-high orbits. Modern computer processing techniques were increasingly able to process the data from these two types of meteorological satellites to create useful composite images.

In time, the multispectral and infrared cameras were able to produce higher and higher resolution. Over time the technology has continued to develop and improve. The evolving sensor technology has seen the development of advanced baseline imagers (ABIs), advanced microwave sounding units (AMSUs), advanced very high resolution radiometer (AVHRRs), and high resolution infrared radiation sounders (HIRS). The very latest technology is the development of a geostationary lightning mapper (GLM) which can monitor the intensity of lightning strikes. This mapping allows meteorologist to see where the most intense part of a storm actually is and to "see" the directionality of the moving storm front on a near-real-time basis.

The combination of imagers, radiometers, infrared radiation sounders, and microwave sensors allows for both environmental and meteorological monitoring. The complex suite of sensors on board meteorological satellites combines to follow not only short-term weather phenomena but also enables the observation of longer-term environmental changes, including the monitoring of pollution. Military and civilian technology in this area – both in terms of onboard sensors for Geostationary and polar-orbiting systems as well as computer capacity for processing of much, much greater volumes of data – has evolved quickly in the last two decades. In the process, warnings based on meteorological satellites data have allowed for hundreds of thousands of lives to be saved both by being able to provide short-term warnings with regard to dangerous weather systems and through longer-term forecasts that allow the creation of improved emergency response capabilities involving the most destructive hurricanes, typhoons, and tropical storm systems.

International Cooperation in the Field of Meteorological Satellite Services

Just as NASA research efforts in the 1960s and 1970s with the Television InfraRed Observation Satellite (TIROS), NIMBUS, and Landsat-1 gave rise to the operational meteorological satellites of the National Oceanic and Atmospheric Administration (NOAA) known as GEOS and Advanced TIROS (TIROS-N or ATN) in the USA, the European Space Agency with its Meteosat in the 1970s gave rise in the 1980s to the European Organisation for the Exploitation of Meteorological Satellites (Eumetsat) and its operational METEOSAT and MetOps geostationary and polar-orbiting programs for Europe. [2]In 1977, Japan launched its first meteorological satellite (GMS) into geostationary orbit with an orbital location to cover the western Pacific and East Asia. Similarly, from 1982, the Indian Space Research Organization (ISRO) has deployed INSAT geostationary meteorological satellites and in 1994, building on its experience with the METEOR series of polar-orbiting satellites, the Russian Federation launched the first of its Geostationary Operational Meteorological Satellite (GOMS), later renamed Elektra-1. The 1988 launch by China of the FY-1 (Feng-Yun 1), its first polar sun-synchronous orbit meteorological satellite, led in 1997 to the launch of its FY-2A geostationary meteorological satellite. Through a joint development by the Korea Aerospace Research Institute (KARI) and EADS Astrium, the Communication, Ocean, and Meteorological Satellite (COMS-1), the latest of the geostationary meteorological satellites was launched in June of 2010.[3]

There has been a close relationship between the US and European meteorological satellite system for many years, and this has most recently resulted in the Joint Polar Satellite System with the USA developing and operating the Polar Orbiting Operational Environment Satellites (POES) and the Europeans operating the polar-orbiting MetOps system on a coordinated basis.

JPSS, in its current form, was reconstituted by the White House signing an Executive Order in February 2010. This presidential executive order set in motion the dissolution of the so-called National Polar-orbiting Environmental Satellite System that was a joint program involving NASA, NOAA, and the US Department of Defense. In the original form of the NPOESS, the National Ocean and Atmosphere Administration (NOAA) was to operate the polar satellite for the afternoon orbit, while the Defense Weather Satellite System (DWSS) would operate the morning orbit. The earlier program had experienced schedule slips and cost overruns and thus the JPSS seeks to consolidate management, still meet civil and strategic meteorological data requirements, and also bring forward the objective of numerical weather projections. Under the restructured program NOAA, with NASA support, would operate the overall program for the morning and afternoon orbiting satellites, but the Defense Weather Satellite System would still be able to access the data. [4]These systems provide meteorological data to meteorological researchers and to defense-related agencies in the USA and Europe. The POES and MetOps systems, known as the Joint Polar Satellite System (JPSS), are not currently a part of the World Weather Watch global WMO Space Programme network (Fig. 1).

---

[2] European Organisation for the Exploitation of Meteorological Satellites(Eumetsat) http://www.eumetsat.int/Home/index.htm.

[3] See chapter "▶ International Meteorological Satellite Systems" for more information on these satellite systems.

[4] Joint Polar Satellite System http://www.nesdis.noaa.gov/pdf/jpss.pdf.

Fig. 1 One of the three European Space Agency developed MetOps satellite prior to launch into polar orbit (Graphic courtesy of the European Space Agency)

Development of the Meteorological Satellite Systems

The Japanese government has also had a strong interest in developing and operating meteorological satellites. This interest is driven by the fact that there are an annual series of typhoons and monsoons that threaten the Japanese islands with highly destructive storms. This has led the Japanese government to undertake the development and deployment of the Geostationary Meteorological Satellite system of Japan also known as the MTSAT or Himawari ("Sunflower in English") Satellite System. The MTSAT satellite series that were constructed by the Space Systems Loral company had difficulties with launch failures that occurred with the Japanese II launch vehicle. After two successive launch failures, the MTSAT 1R was successfully launched. During the period 1999–2005, the time of the two MTSAT failed launch attempts and MTSAT 1R, temporary arrangements were made with the USA for the lease of the GEOS 9 satellite to provide meteorological services to Japan and surrounding areas.[5]

---

[5] Japanese Geostationary Meteorological Satellite (GMS) "Himawari" (Sunflower) system by JAXA www.jaxa.jp/projects/sat/gms/index_e.html.

Other countries have continued to develop and launch meteorological satellite systems as well. These systems have included a modernization of the Russian Geostationary Operational Meteorological Satellite (GOMS) and METEOR satellites with the Elektro-L and METEOR-3 series of satellite systems, the Chinese Feng-Yang FY-2 and FY-3 systems, and the Indian INSAT-2 satellite system. The highly capable Russian and Chinese systems are dedicated geostationary systems developed to provide detailed meteorological monitoring. Under a joint agreement with NASA, a data server to provide GOMS data has been established in cooperation with the Russian Ground Microprocessing Information Systems SRC "PLANETA" and the Space Monitoring Information Support Laboratory (IKI RAN).[6]

The INSAT-2 satellite system – also a geo satellite – was designed in an unconventional manner in that some of these spacecraft were designed with two different payloads. One payload provides telecommunications and television broadcasting services, while the other payload supports a meteorological package. In one of these designs, a long boom was extended to create overall equilibrium to the solar cell arrays that was deployed on only one side of the spacecraft so as to not block the meteorological imaging (see Fig. 2 below on the left). In the latest INSAT 3 and INSAT 4 series, however, a common spacecraft bus was used, but the spacecraft were designed either to provide telecommunications in one edition or as meteorological satellites in the alternative design.[7]

Present and Future Coordination of Meteorological Satellite Systems

The World Meteorological Organization (WMO), a specialized agency of the United Nations, has recognized the importance of coordination of the data produced by the various national meteorological satellite systems. In May 2003 at the Fourteenth WMO Congress, a "WMO Space Programme" was created during the meetings in Geneva, Switzerland. The purpose of the WMO Space Programme was to coordinate all environmental satellite matters around the world – including both meteorological programs but also remote sensing for related environmental purposes such as hydrology, oceanic, and glacial monitoring.[8]

With this WMO Space Programme, there is now what is called the World Weather Watch (WWW) Global Observing System (GOS). The GOS includes three different types of satellites. These are operational polar-orbiting satellites, operational

---

[6] (Present and Future Chinese Meteorological Satellite Systems www.fas.org/spp/guide/china/agency/nsmc.htmData server to provide Russian GOMS satellite data http://gcmd.nasa.gov/KeywordSearch/Metadata.do?Portal=GCMD&KeywordPath=&NumericId=4140&MetadataView=Full&MetadataType=0&lbnode=mdlb2.

[7] Indian National Satellite System (Insat) ISRO designed Insat E, http://isro-news.blogspot.com/search/label/INSAT%20-%202E.

[8] Donald Hinsman, *Implementation Activities within the WMO Space Programme* World Meteorological Organization (WMO).

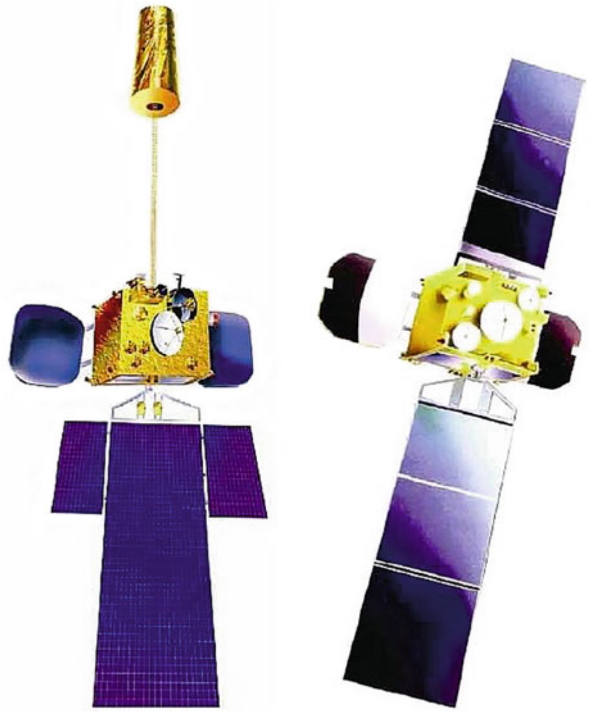


Fig. 2 Insat 2 unconventional asymmetrical design alongside a more conventional later Insat satellite (Graphics courtesy of ISRO)

geostationary satellites, and research satellites that collect important data but are not considered operational satellites.

The three different types of satellites provide imagery in optical and infrared frequencies, microwave and other types of soundings, data collection, and data distribution. In particular, the present operational meteorological satellites include the following geostationary and polar-orbiting satellites: GOES-10, GOES-12, NOAA-15, NOAA-16, and NOAA-17 operated by the USA; GMS-5 operated by Japan; GOMS N-1, METEOR 2–20, METEOR 2–21, and METEOR 3–5 operated by the Russian Federation; METEOSAT-5, METEOSAT-6, METEOSAT-7, and METEOSAT-8 (formerly MSG-1) operated by EUMETSAT; and FY-2B, FY-1C, FY-1D operated by China. Additional satellites in orbit include GOES-8, GOES-9, and GOES-11 operated by the USAs. It should be noted that most space agencies contributing operational polar-orbiting and geostationary satellites have in place contingency plans for satellite systems that guarantee the continued daily flow of satellite data, products, and services WMO Members have come to depend on (Donald Hinsman, *Implementation Activities within the WMO Space Programme* World Meteorological Organization (WMO)).

This has in some instances allowed countries to call into service on a lease or sharing agreement the meteorological satellites of partner countries, such as the above noted agreement between Japan and the USA for the interim lease of the GOES-9 satellite by Japan.

Conclusion

The evolution of meteorological satellites in the half century since the 1960s has shown remarkable technological growth and development. Satellites can detect atmospheric temperatures, cloud cover, oceanic current flows and pollution, monitor storm development and energy levels. Remote satellite sensing provides imaging and soundings over a wide range of different frequencies that range from microwave, infrared, multi-spectral imaging (in the optical range) up to the ultraviolet spectra. These diverse imaging platforms provide data from polar and geo orbits and produce an enormous amount of information now measure in petabytes. Modern high speed computers, expert systems, and artificial intelligence allow this data to be processed in near real time to produce valuable data on which aviation, transportation industries, agriculture, forestry, fishing, construction, and virtually every business and individual on the planet to some extent now depend.

Over the past decades, several key shifts have occurred with regard to the mission of meteorological satellites. The first shift was to expand the mission from short-term weather forecasts to include medium- and longer-term forecasts as well. The second, more recent, shift has been to expand the mission from weather forecasting to environmental monitoring. This has become increasingly important as the significance of atmospheric, oceanic, and glacial pollution has become better understood and environmental issues such as holes in the ozone layer that protects us from radiation and the potential longer-term implications of climate change, the buildup of greenhouse gases, and global warming have become clearer. The various eyes in the sky that are our meteorological and remote sensing satellites give us the best perspective to “see” the changes that are occurring to our planet and how we can best cope with these shifts in the health and vitality of planetary environment.

Cross-References

- ▶ International Meteorological Satellite Systems
- ▶ United States Meteorological Satellite Program

United States Meteorological Satellite Program

Sergio Camacho-Lara, Scott Madry, and Joseph N. Pelton

Contents

S. Camacho-Lara (✉)
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Santa María Tonantzintla, Puebla, Mexico
e-mail: sergio.camacho@inaoep.mx

S. Madry
Global Space Institute, Chapel Hill, NC, USA
e-mail: Scottmadry@mindspring.com

J.N. Pelton
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_54

Abstract

Over the past half century, weather and sophisticated environmental imaging satellites have evolved providing an increasing ability to monitor a wide range of conditions on Earth. A long-term and effective partnership between the National Aeronautics and Space Administration, the US space agency, NASA, and the National Oceanic and Atmospheric Administration, NOAA, has worked to design, launch, and operate a series of environmental monitoring satellites. These environmental monitoring satellites have grown in their technical capabilities to monitor cloud coverage, temperature, and wind velocity over the oceans and seas, lightning intensity, and storm formations. Interactive capabilities have for some time allowed these satellites to assist with the monitoring of climate change, space weather, and search and rescue activities. In short, the expanded technical capabilities of these satellites, and particularly of the Geostationary Operational Environmental Satellite system, have allowed the development of an ever increasing range of applications and functionality.

The initial US meteorological or weather satellite program that began with TIROS created a specific type of remote-sensing satellite that could assist in monitoring weather conditions for the continental USA. Today's GOES and Polar-Orbiting Environmental Satellites (POES) have now grown to become global in scope. These US satellites allow the development of an increasingly wide range of knowledge of the oceans and the Polar region, allow for more accurate mathematical models of meteorological conditions, help to monitor "space weather" conditions, assist with rescue of distressed ships and aircraft, aid transportation systems, and help with monitoring atmospheric pollution and conditions associated with climate change.

The National Ocean and Atmospheric Administration, through its National Environmental Satellite, Data, and Information Service (NESDIS), continuously operates a global network of satellites to achieve these goals. NOAA works closely with NASA in the design of environmental satellites and cooperates with the US Department of Defense in obtaining and distributing environmental information. Data obtained from US environmental spacecraft as well as from other spacecraft around the world are used for a wide range of applications. Currently these applications relate to the oceans and seas, coastal regions, agriculture and resource recovery, detection of forest fires, detection of volcanic ash, monitoring the ozone hole over the South Pole, and even the space environment in terms of the so-called space weather such as solar flares.

Each day NOAA's NESDIS processes and then distributes more than 3.5 billion bits of data. The processed images are distributed to weather forecasters in the USA and globally so that various users, for instance, disaster managers, and the general public can see weather patterns via television or on computer or smart phone displays. The timeliness and quality of the combined polar and geostationary satellite data have been greatly improved by enhanced computer installations, upgraded ground facilities, and international data sharing agreements as well as by military weather services.

Keywords

Advanced baseline imager (ABI) • Advanced microwave sounding unit (AMSU) • Advanced very high resolution radiometer (AVHRR) • Data EUMETSAT • Geostationary lightning mapper (GLM) • Geostationary operational environmental satellites (GOES) • High resolution infrared radiation sounders (HIRS) • Improved TIROS operating system (ITOS) • Information service (NESDIS) • Initial joint polar-orbiting satellite (IJPS) • Joint polar satellites system (JPSS) • MetOp satellites National oceanic and atmospheric administration (NOAA) • National polar-orbiting operational environmental satellite system (NPOESS) • National weather service (NWS) • Nimbus polar-orbiting operational environmental satellites (POES) • TIROS TIROS-N series

Introduction

For more than 50 years, US environmental satellites have been an integral key to lifesaving weather and climate forecasts for the USA and many other countries. NOAA, created in October 1970 to consolidate atmospheric- and oceanic-related activities and to operate environmental satellites, has been carrying out this mission on behalf of the USA for nearly 45 years. Its activities include operation of the Geostationary Operational Environmental Satellite (GOES) system and the Polar-Orbiting Environmental satellites (POES). It is also currently involved in developing the next generation Joint Polar Satellite System (JPSS) with EUMETSAT, the European Organisation for the Exploitation of Meteorological Satellites, and the next generation of GOES spacecraft with NASA. Since its establishment, NOAA has cooperated with NASA in the design of new environmental spacecraft that have allowed more sophisticated monitoring capabilities to be deployed with increasing resolution, new types of sensors, and increasing levels of global coverage.

NOAA has, on an on-going basis, coordinated with the US Department of Defense in identifying environmental monitoring requirements and developing increasingly capable environmental spacecraft for monitoring and scientific research. International agreements have allowed NOAA, through its National Environmental Satellite, Data, and Information Service (NESDIS) and its National Weather Service (NWS) to perform its responsibilities more effectively and at an increased level of cost effectiveness. NOAA has, in particular, been able to obtain an augmented amount of information from other satellite systems operated by other countries via various data sharing arrangements. This has not only allowed more accurate weather forecasts, but it has also increased knowledge related to climate change and global patterns of atmospheric, oceanic, and pollution over the polar region.

Today’s US capabilities in environmental monitoring and reporting grew out of the International Geophysical Year (IGY) that was carried out during the 1950s during the Eisenhower administration and, in fact, gave birth to the Space Age. The launch of such a first “weather satellite” was committed to by the USA in 1953 as part of the planning process for IGY.

Table 1 The TIROS 1 to TIROS 10 satellite network (Gary Davis, NOAA, History of the NOAA Satellite Program, June 2011 http://goes.gsfc.nasa.gov/text/history/History_NOAA_Satellites.pdf)

TIROS satellites	Launch date	Operating life (days)[a]	Orbit	Features
TIROS-1	1 Apr 60	89	Inclined	2 TV cameras
TIROS-2	23 Nov 60	376	Inclined	2 TV cameras, radiometer
TIROS-3	12 Jul 61	230	Inclined	2 TV cameras, radiometer[b]
TIROS-4	8 Feb 62	161	Inclined	1 TV camera, radiometer[b]
TIROS-5	19 Jun 62	321	Inclined	2 TV cameras
TIROS-6	18 Sep 62	389	Inclined	2 TV cameras
TIROS-7	19 Jun 63	1,809	Inclined	2 TV cameras, radiometer[b]
TIROS-8	21 Dec 63	1,287	Inclined	1 TV camera, APT[c]
TIROS-9	22 Jan 65	1,238	Sun Synch[d]	Global coverage, 2 TV cameras
TIROS-10	2 Jul 65	730	Inclined	2 TV cameras

[a] Number of days until satellite was turned off or failed
[b] Radiometer (visible and infrared channels)
[c] Automatic picture transmission for direct readout locally
[d] Sun-synchronous

The TIROS 1 satellite (i.e., Television Infrared Observation Satellite) that inaugurated US environmental space capabilities was actually launched on April 1, 1960, and this became the first of ten such TIROS satellites. A US space-based weather and environmental monitoring capability has continued ever since this event over 50 years ago. Since that time, many dozens of environmental monitoring satellites have been launched. This chapter reviews these past, current, and planned space assets as well as describes cooperative relationships with other environmental spacecraft monitoring systems around the world.

Historical Background

The TIROS series of "weather" satellites initially provided infrared images of weather conditions in the USA, providing these capabilities through the 1960s. The experimental meteorological spacecraft proved the feasibility of using spacecraft to monitor and predict weather systems and even to create mathematical models of Earth-based weather conditions. Later satellites in the series provided increased capabilities that included radiometers that operated in the visible as well as the infrared frequencies, that is, in TIROS 3 and 7, while TIROS 8 provided for direct local reception and readout (Table 1).

After the TIROS series, there was a follow-on TIROS program known as ITOS (Improved TIROS Operational System). These satellites represented a step up from a research and development phase into a fully operational program. ITOS-1 was launched in January 1970 and greatly surpassed the performance of the earlier satellites by providing both direct transmission and storage of television and infrared imagery. Later ITOS spacecraft also supplied vertical profiles of atmospheric temperature. ITOS satellites remained in service through 1979. The problem of the ITOS and the TIROS series before them was that their near-polar orbit required the piecing together of a mosaic of television images to create a unified image of the Earth. This required elaborate and labor-intensive ground processing of the satellite images and then redistribution of the processed data.

The initial TIROS system was essentially a “weather” satellite and the ITOS that followed was a much improved system but still essentially a television camera imaging system focused on weather monitoring and forecasting. TIROS and ITOS, however, served as important precursors to the much more complex and capable space-based environmental monitoring system that is operated by the USA today. Another key step in the evolution was the Nimbus satellites designed and launched by NASA as part of its environmental research program from the mid-1960s through much of 1970s.

Nimbus Satellite Program

NASA’s Nimbus satellites were flown from 1964 through 1978, as advanced research satellites that tested new sensing instruments and data-gathering techniques rather than as operational weather satellites. These satellites served to test new sensors and to augment the understanding the Earth’s environment, including the dynamics of the ozone layer that protects the Earth from solar and cosmic radiation. The Environmental Science Services Administration (former name for the National Weather Service), however, did become a routine user of Nimbus data. These data were valuable for their coverage of conditions over oceans and other areas where few other upper atmospheric measurements were then being made. Instruments on the Nimbus satellites included microwave radiometers, atmospheric sounders, ozone mappers, the Coastal Zone Color Scanner, and infrared radiometers and provided significant global data on sea-ice coverage, atmospheric temperature profiles, atmospheric chemistry (i.e., ozone distribution), the amount of radiation in the Earth’s atmosphere, and sea-surface temperature. The Total Ozone Mapping Spectrometer (TOMS) instrument aboard the final Nimbus, Nimbus-7, mapped the extent of the phenomenon known as the “ozone hole.” A total of seven Nimbus spacecraft were launched by NASA into near-polar and sun-synchronous orbits. The first of these satellites was Nimbus 1 launched on August 28, 1964. See Fig. 1.

The Nimbus series of satellites provided a key capability for satellite remote sensing of the Earth, atmospheric data collection, and weather forecasting. The seven Nimbus satellites actually provided key data from orbit through 1995 – a remarkably long 30-year period for an experimental program. The technology and lessons learned from the Nimbus missions underlie most of the Earth-observing satellites NASA, NOAA, and other space programs have launched over the past three decades. The basic imaging and sounder systems and the technologies used within the Nimbus are still very much a part of today’s systems although current imaging and sounder systems operate at much higher levels of accuracy and frequency of

Fig. 1 The sun-synchronous near-polar-orbiting nimbus meteorological satellite (Graphic courtesy of NASA-GSF and NOAA GOES)

image update. Nimbus operated in near-polar and sun-synchronous orbits, but experiments with geosynchronous orbits that began with the Syncom systems described in earlier chapters related to the history of satellite communications suggested that these orbits could also be effectively used for meteorological and environmental monitoring purposes.

The start of a process that led to a new type of environmental monitoring satellite that would operate from geostationary orbit and provide a near real-time image of the complete "global disk" actually started in 1964, the year the first Nimbus satellite was launched.

In January 30, 1964, a formal agreement was reached between NASA and the US Weather Bureau to work together to establish a National Operational Meteorological Satellite System. Under this agreement, the Weather Bureau would establish overall requirements and performance characteristics and would also operate command and data acquisition stations. As a consequence of this agreement, NASA was formally tasked with designing the spacecraft and conducting their launch. With the later formation of NOAA, this cooperative agreement continued essentially as initially agreed.

NASA Experimental Programs and the Birth of Geostationary Systems

Part of the answer about how a geostationary satellite might be used for environmental purposes came on December 6, 1966, with NASA's launch of the first Applications Technology Satellite (ATS-1). This experimental satellite demonstrated the value of using the geostationary orbit for maintaining a continuous watch over one spot on the globe – in this case the continental land mass of the USA. ATS-1's spin-scan cloud camera, invented by Verner Suomi and Robert Parent at the University of Wisconsin, was capable of providing full disk visible images of the Earth, providing an updated cloud cover image every 20 min.

Professor Suomi commented about this totally new capability after the first ATS-1 images were received. He noted, "Now the clouds move and not the satellite." Research by meteorological scientists began almost immediately. They were able to track clouds and produce models of wind products using image sequences. This was clearly a better way to research cloud formation and atmospheric conditions than having to piece together mosaics of images obtained from sun-synchronous, near-polar orbit satellites.

ATS-3, a larger version of ATS-1 that followed shortly after, was the first spacecraft to routinely transmit full disk Earth-cloud images in color. Earlier series spacecraft had peak sensitivity in the green region of the visible spectrum. However, the ATS-3's Multicolor Spin Cloud Cover Camera provided new peak sensitivity to the red, blue, and green visible spectra using three photo-multiplier light detectors.

The NASA-funded ATS experimental satellite series continued development through six spacecraft. The main focus of this ATS-series was on environmental monitoring spacecraft although the ATS-6 was essentially an experimental communications satellite. NASA's office of international affairs worked actively to allow a wide range of international participation in these ATS experiments.

By the early 1970s, ATS imagery was being routinely used in operational forecast centers, with the first movie loops being used at the National Severe Storm Forecast Center (NSSFC) in the spring of 1972. Atmospheric motion depiction from geostationary satellite image loops was transferred into routine operations at the national forecast centers, and the resulting cloud motion vectors evolved into an important data source of meteorological information, especially over the oceans. These data supplemented the data that were provided by the Nimbus program.

The success of the meteorological experiments carried aboard the ATS-series of satellites led to NASA's development of a new satellite specifically designed to make atmospheric observations in a geostationary orbit, 35,786 km (22,230 miles) above the equator. NOAA's Geostationary Operational Environmental Satellite (GOES) program thus sprang out of this cooperative experimental period.

In particular, NASA designed, built, and launched the first two geosynchronous meteorological satellites: Synchronous Meteorological Satellite-1 (SMS-1) in May 1974 and SMS-2 in February 1975. These two spacecraft were the prototypes for the NOAA GOES program. GOES-1 was launched on October 16, 1975, followed by GOES-2 and 3, which were similar in design and provided continuity of service. The primary instrument on the SMS-1 and 2 and GOES-1 to 3 spacecraft was what was called the Visible/IR Spin Scan Radiometer (VISSR), and was based on Professor Soumi's original conceptual design. The VISSR, a true radiometer, provided day and night observations of cloud and surface temperatures, cloud heights, and wind fields.

The fundamental shift from experimental projects to operational programs occurred in the early 1970s. The initial TIROS and the ATS programs, which ended around 1970 and in the mid 1970s, respectively, were largely replaced by the ITOS (i.e., TIROS-N series) and NOAA's new Geostationary Operational Environmental Satellite (GOES) system starting in 1975. The GOES network has continuously evolved over the past 35 years into what is now a full-scale environmental monitoring network that also supports all of NOAA's goals and scientific functions with new capabilities being systematically added. The US GOES network and its other spacecraft work closely with the international community, especially with EUMETSAT and the World Meteorological Organization.

The Geostationary Operational Environmental Satellite (GOES) Network

The USA operates two meteorological satellites in geostationary orbit: one over the East Coast, GOES EAST around West Longitude – 75.0°, and one over the West Coast, GOES WEST around West Longitude – 135° with overlapping coverage over the USA. These satellites are replaced when they reach the end of their operation lifetime by satellites with improved or additional new instruments.

A total of 15 GOES Satellites have been deployed by NOAA since 1975. These GOES Satellites as described below have become increasingly sophisticated over the past few decades.

These space assets play a key role in supporting the major operational and strategic goals of NOAA. These goals today include (NOAA History, 2011):

- Supporting the needs of the USA and those of world society to obtain up-to-date and reliable weather, water, and other related information, including data on atmospheric and oceanic pollution, on a timely basis
- Protecting, restoring, and managing the use of coastal and ocean resources through an ecosystems approach to water use and management
- Understanding climate variability and enhancing society's ability to plan and respond to climate change
- Supporting US' commerce by supplying accurate and timely information for safe, efficient, and environmentally sound transportation

The various spacecraft assets described in the next section (i.e., GOES and polar-orbiting environmental monitoring spacecraft) now play a critical role in meeting these strategic objectives. Increasingly capable environmental monitoring satellites operated by the USA and other entities – particularly EUMETSAT – allow NOAA to accomplish much more than provide weather data.

GOES-8 to 12 Satellites

The history of the early GOES satellites and their technical characteristics will not be covered in detail here. This is because these earlier spacecraft, although they once performed a vital service, are now fully retired from service and their historical specifications are only currently relevant in that they helped to prove technologies that are incorporated in today's operational spacecraft.

Let's thus start with a review of the GOES-8 to 12 (known as GOES I to M prior to launch). These spacecraft are currently being phased out and are being replaced by the GOES-13, 14, and 15 that will be described in the next section.

Beginning with launches in the mid-1990s, the GOES-8 to 12 series spacecraft provided weather and environmental monitoring for the USA for over a decade. These spacecraft were equipped to perform the following specific functions:

- Acquisition, processing, and dissemination of imaging and sounding data
- Acquisition and dissemination of Space Environment Monitor (SEM) data
- Reception and relay of data from ground-based Data Collection Platforms (DCPs) that are situated in selected urban and remote areas to the NOAA Command and Data Acquisition (CDA) station
- Continuous relay of Weather Facsimile (WEFAX) and other data to users, independent of all other functions
- Relay of distress signals from people, aircraft, or marine vessels to the search and rescue ground stations of the Search and Rescue Satellite-Aided Tracking (SARSAT) system

These satellites, as exemplified by GOES-12 (M) in Fig. 2, have two major subsystem capabilities in the form of the GOES I/M Imager and a Sounder (GOES I/M Brochure):

The GOES I/M Imager is a multichannel instrument designed to sense radiant and solar-reflected energy from sampled areas of the Earth. The multielement spectral channels simultaneously sweep east–west and west–east along a north-to-south path

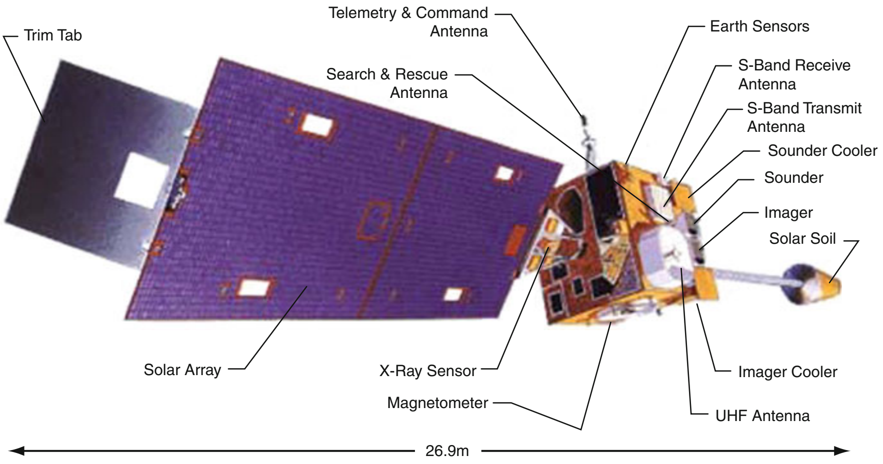


Fig. 2 GOES-12 (M) Launch date: July 23, 2001 (Graphic courtesy of NASA-GSF and NOAA GOES)

Table 2 GOES I/M imager monitoring capabilities

Imager channels and capabilities					
Channel	1	2[a]	3[a]	4	5[a]
Wavelength (μm)	0.65	3.9	6.7	11	12
Product					
Clouds	x	x	x	x	x
Water vapor[a]			x	x	x
Surface temperature		o		x	o
Winds	x		x	x	
Albedo + IR Flux	x		o	x	o
Fires + Smoke	x	x		o	o

Key: [a]new operational data, *x* primary channel, *o* secondary channel

by means of a two-axis mirror scan system. The instrument can produce full-Earth disk images, sector images that contain the edges of the Earth, and various sizes of area scans completely enclosed within the Earth scene using a new flexible scan system. Scan selection permits rapid continuous viewing of local areas for monitoring of mesoscale (regional) phenomena and accurate wind determination. Table 2 above indicates the specific technical performance characteristics of the GOES I/M Imager (GOES I/M Imager).

The GOES I/M Sounder is the other major component of this type NOAA spacecraft. The "GOES I/M Sounder" is a 19-channel discrete-filter radiometer. This radiometer covers the spectral range from the visible wavelengths up to 15 μm. It is designed to provide data from which atmospheric temperature and moisture profiles, surface and cloud-top temperatures, and ozone distribution can be deduced by mathematical analysis. An engineering sketch of the GOES Sounder (or Radiometer) is provided in Fig. 3 below.

The GOES I/M series Sounder or Radiometer operates independently of and at the same time as GOES I/M Imager. The Sounder, in fact, uses a similar flexible scan system as is used in the Imager. The Sounder's multielement detector array assemblies simultaneously sample four separate fields or atmospheric columns. A rotating filter wheel, which brings spectral filters into the optical path of the detector array, provides the infrared channel definition. These capabilities are listed in Table 3 below (GOES I/M Sounder).

GOES 13, 14, and 15

The current generation of GOES satellites will be replaced by a new generation of satellites post 2016. These spacecraft were known as GOES N/O/P prelaunch and now known as GOES-13, GOES-14, and GOES-15 post launch. These spacecraft (See Fig. 4 below) provide important new capabilities in terms of weather monitoring, environmental monitoring, and active search and rescue operations that are described below. In addition to an imager and sounder with expanded capabilities,

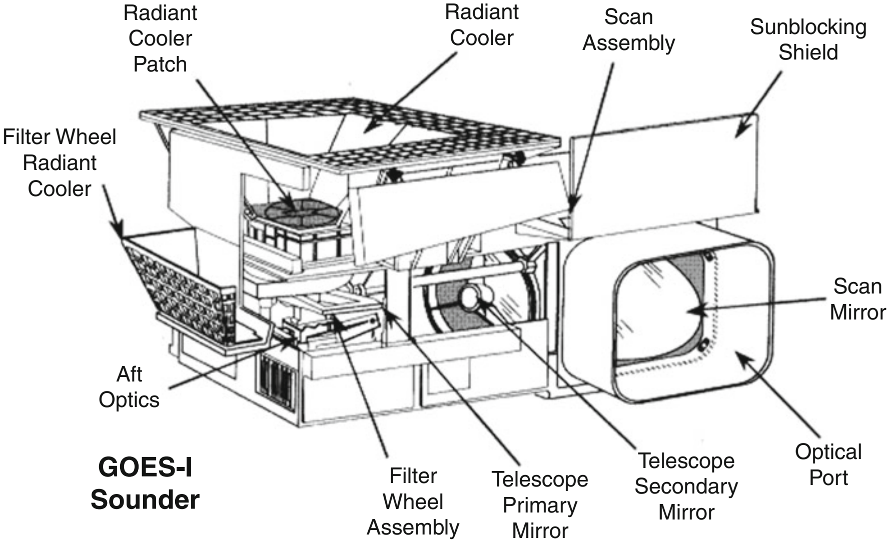


Fig. 3 Engineering sketch of the GOES I series sounder (Graphic courtesy of NASA-GSF and NOAA GOES)

Table 3 GOES I/M sounder environmental monitoring capabilities

THE GOES I/M sounder capabilities				
	Resolution (km)		Accuracy	
	Vertical	Horizontal	Absolute	Relative
Product				
Temperature				
Profile	3–5	50	2–3 °K	1 °K
Land	–	10	2 °K	1 °K
Sea	–	10	1 °K	0.5 °K
Moisture				
Profile	2–4	50	30 %	20 %
Total	–	10	20 %	10 %
Motion	3 layers	50	6 m/s	3 m/s
Cloud				
Height	2 layers	10	50 mb	25 mb
Amount	Total	10	15 %	5 %
Ozone				
Total	–	50	30 %	15 %

these satellites also carry an upgraded Space Environmental Monitor and a Solar X-Ray Imager. They also kept a remote data collection capability from land and ocean-based DCPs. Finally, they kept the communications capability to support search and rescue operations. The network operates as a two satellite configuration

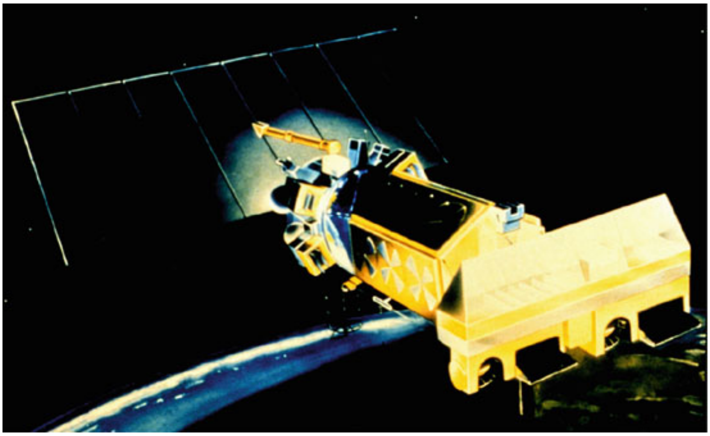


Fig. 4 The GOES 13 satellite (known as GOES N and shown prior to launch) (Graphic courtesy of NASA-GSF and NOAA GOES)

from over the Pacific (GOES West) and Atlantic Ocean (GOES East) with the capability to continuously observe about 60 % of the Earth's Surface. The third spacecraft is available as a spare.

A summary of improvements of the current GOES N/O/P spacecraft in comparison to the spacecraft they are replacing (i.e., the GOES I/M series) is as follows:

- The satellite power system (i.e., solar array and battery systems) has been upgraded in reliability and performance.
- The design lifetime has been upgraded from 7 to 10 years, and sufficient fuel has been provided to support 13.5 years of life which should be fully achievable.
- The command data rate for each satellite has been increased from 250 bits/s to 2,000 bits/s to allow more complex commands at faster data rates. Telemetry data rates have also been improved for faster readouts of up to 4,000 bits/s if required.
- There is a star tracking system that will allow more precise pointing of the spacecraft. This will allow for more accurate registration and navigational capabilities.
- Solar data can now be collected by inclusion of a Solar X-Ray Imager.
- The Space Environmental Monitoring (SEM) subsystem has been upgraded in a number of ways to include an Extreme Ultra Violet (EUV) sensor, an Energetic Proton, Electron, and Alpha particle Detector (EPEAD), and a Magnetosphere Electron Detector (MAGED). This will greatly increase the ability to monitor space weather conditions.
- An additional transponder has been added to support an Emergency Manager's Weather Information Network (EMWIN).
- The data from the spacecraft will be distributed via a digital Low Rate Information Transmission (LRIT) system that will replace the analog WEFAX data distribution system that operated at lower speeds and greater potential for disruption.

- The data collection system will allow the collection of more data from more remote land and ocean-based platforms at higher speeds (i.e., up to 1,200 bits/s) through the use of an eight-phase shift keyed system.

A full description of each of the four major sensing systems on the GOES 13–15 spacecraft, their technical characteristics, and their associated ground system capabilities is available on the NOAA and NASA website (NOAA History).

It is important to know that this fully functional system can provide NOAA meteorologists and experts on the ground with a wide and rich range of key data and actionable information. The readout systems are organized within the NOAA information network to include Dust Storms, Fire Events, Flood Events, Iceberg Events, Ocean Events, Severe Weather Events, Snow Cover, Storm Systems, Tropical Cyclones, Unique Imagery, and Volcano Events.

The design and manufacture of the GOES spacecraft, from GOES 1–15, are in many ways similar to the design of the spacecraft used for telecommunications, satellite navigation, or other types of remote-sensing activities. This can be clearly seen by undertaking a review of the power systems (i.e., solar arrays and batteries), the stabilization and tracking system, fuel used for stabilization, station-keeping and repositioning, thermal control and heat transfer, and tracking, telemetry, and command systems as noted in chapter "▶ Common Elements Versus Unique Requirements in Various Types of Satellite Application Systems." That is, the satellite platforms are all quite similar in design and manufacture. This is really not surprising as most application satellites are manufactured by the same companies and the research scientists and engineers who work on the various subsystems develop technology for a "platform" that can be used on any application satellite. Just as a manufacturer of vehicles might use the same engine or chassis on various automobile models, the same is true for manufacturers of spacecraft. The many innovations that were applied to the design of the GOES N-P spacecraft such as star tracking, additional fuel for extended lifetime, faster data rates for command and telemetry, and improved power subsystems are found in other types of applications spacecraft being deployed today. The innovations that are unique to these spacecraft are those associated with the four sensors. The imager and sounder are improved versions of those found on GOES I-M series, and the space environmental sensors are essentially new capabilities.

GOES-R Series

The current GOES System is expected to continue operating at least for another 5–7 years. The additional fuel and design upgrades to extend the life of GEOS 13, 14, and 15 should actually allow them to operate beyond this time. The GOES-R is to be launched in 2016, and GOES S, T, and U are designed to provide service through 2036. Each of these new class of GOES satellites that are designed will operate with a mission design life of at least 7–10 years and will notably improve current GOES capabilities. These new capabilities of the GOES-R series are described below.

Despite various discussions and initiatives in the past that considered the idea of trying to “privatize” weather and meteorological satellite operations within the USA, it is currently intended and broadly anticipated that the GOES-R, S, T, and U series will continue as NOAA and NASA programs. The acquisition of the end-to-end GOES system includes spacecraft, sensors, launch services, and ground system elements consisting of mission management, product generation, product distribution, archive and access interface, and user interface.

New instrumentation that will be deployed within the GOES-R spacecraft includes that Advance Baseline Imager and the Geostationary Lightning Mapper. Both of these new instruments are expected to provide important new capability that will enhance public and transportation safety and more effective severe storm warning and lead to economic savings and benefits (GOES-R):

Advanced Baseline Imager (ABI): The ABI is the primary instrument on GOES-R for imaging Earth’s weather, climate, and environment. The ABI will be able to view the Earth within a very broad range of 16 different spectral bands, including two visible channels, four near-infrared channels, and ten thermal infrared channels. The ABI is almost revolutionary in its capabilities in that it provides three times more spectral information, four times better spatial resolution, and more than five times as many updates of its sensed data than the GOES-13 to 15 satellites in the current system.

The ABI is designed to observe essentially the entire western hemisphere in various time intervals and to do so at 0.5, 1, and 2 km spatial resolutions in 16 spectral bands as indicated above. The ABI has two main scan modes. The “flex” mode will provide full disk imagery every 15 min, while covering the continental USA every 5 min. It also has the ability to provide data on demand as frequently as every 30 s. It is expected that two mesoscale regions will be continuously scanned, resulting in a 1 min update for those sectors.

The ABI will be calibrated to an accuracy of 3 % (1 σ) radiance for visible and near-infrared wavelengths. For infrared channels, the ABI will be accurate to 1 °K (1 σ) at 300 °K. The ABI on the GOES-R satellite will thus actually improve every product compared to the current GOES Imager and introduce new products. Two new products to be produced by the ABI include the capability to indicate the probability of fog and its density and the accurate detection of precise vectors of atmospheric motion.

The GOES-R series satellites will also provide increased time-response capabilities with respect to fires, volcanoes, as well as hurricanes, tornadoes, and thunderstorms. The projected cost benefits of the GOES-R ABI instrument (and the GLM instrument described below) over the lifetime of the series are estimated by NOAA to be close to $5 billion (US). These projected benefits are expected to come from improved tropical cyclone forecasts, fewer weather-related flight delays, and airline incidences with volcanic smoke and ash plumes, improved production and distribution of electricity and natural gas, increased efficiency in irrigated water usage in agriculture, and higher protection rates and more efficient rerouting for airplanes and ships in the event of a tropical storm or hurricane.

Geostationary Lightning Mapper: The GLM is an optical transient detector and imager operating in the near-IR that maps total lightning (in-cloud and cloud-to-ground) activity with near uniform spatial resolution of approximately 10 km continuously day and night over the Americas and adjacent ocean regions. The GLM will provide early indication of storm intensification and severe weather events, improved tornado warning lead time of up to 20 min or more, and data for long-term climate variability studies. It is anticipated that GLM data will have immediate applications to aviation weather services, climatological studies, and severe thunderstorm forecasts and warnings. The GLM will provide information to identify growing, active, and potentially destructive thunderstorms over land as well as ocean areas.

GLM measurements can provide vital information to help the operational weather, aviation, disaster preparedness, and fire monitoring communities in a number of different and quite significant ways:

- An increased ability to develop short range forecasts of heavy rainfall and flash flooding.
- New capability to provide near-real-time detection of enhanced lightning activity that with associated improved models can predict changes in the intensity change of tropical storms, hurricanes, and cyclones.
- Related improved warning capabilities for tornado and severe thunderstorm in terms of increased lead times as well as a corresponding reduction in false alarms and spurious information. (This is a particularly valuable new capability of importance for storm warning and transportation routing for oceanic regions, mountain areas, and areas where there might be radar outages).
- Improved routing of commercial, military, and private aircraft over oceanic regions, mountain areas, and sparsely populated and remote areas during severe storm conditions.
- More accurate and timely warning of lightning ground strike hazards.
- Development of improved and more accurate numerical weather prediction models increased identification of deep atmospheric convection patterns.
- Increased capability to develop what might be called “lightning climatology” and models of lightning intensity within storms.
- Improved ability to monitor and create mathematical models of a wide range of storm and lightning intensity patterns.

One will note that many of the projected benefits from the ABI and GLM are common and that in many instances the combined analysis of the ABI and GLM products will provide the optimum result (Additional GOES-R Information).

Polar-Orbiting Operational Environmental Satellite (POES) System

The Polar-orbiting Operational Environmental Satellite (POES) system supplements the GOES system. The POES offers the advantage of daily global coverage, by making nearly polar orbits roughly 14.1 times daily. Since the number of orbits per day is not an integer, the suborbital tracks do not repeat on a daily basis, although the local solar time of each satellite's passage is essentially unchanged for any latitude. Currently in orbit there are two satellites, one of which passes over a given point on Earth at the same local time in the morning and the other in the afternoon. These spacecraft, referred to as AM and PM satellites, provide global coverage four times daily. The POES system includes the Advanced Very High Resolution Radiometer (AVHRR) and the Tiros Operational Vertical Sounder (TOVS).

Because of the polar-orbiting nature of the POES series satellites, these satellites are able to collect global data on a daily basis for a variety of land, ocean, and atmospheric applications. Data from the POES series support a broad range of environmental monitoring applications including weather analysis and forecasting, climate research and prediction, global sea-surface temperature measurements, atmospheric soundings of temperature and humidity, ocean dynamics research, volcanic eruption monitoring, forest fire detection, global vegetation analysis, search and rescue, and many other applications. These images and data supplement the information that GOES satellites provide.

In 1998, a new series of NOAA Polar Operational Environmental Satellites (POES) commenced with the launch of NOAA K. The NOAA K and its immediate successors, NOAA-L and NOAA-M, represent an improvement over the previous series of satellites that began with TIROS-N (1978) and continued with NOAA-6 through NOAA-14 (1994).

The NOAA K/L/M POES satellites begin a new era of improved environmental monitoring. The NOAA K/L/M satellites, NOAA-15 to NOAA-17 after launch, include improvements to instruments that are evolutionary. The initial concept was to add more passive microwave instruments and channels in place of the four-channel Microwave Sounding Unit (MSU) and the three channel Stratospheric Sounding Unit (SSU). Combined with command system security and frequency changes, NOAA K/L/M satellites look very much like previous satellites, but have significant changes to essentially every subsystem. A description of the instrumentation of the NOAA K/L/M satellites is presented below.

- The Advanced Microwave Sounding Units (AMSU-A1, AMSU-A2, AMSU-B) are state-of-the-art passive microwave sounders that significantly enhance NOAA's atmospheric sounding and nonsounding products suite. The AMSU instruments have better spatial resolution and upper atmospheric sounding capabilities than the previous MSU instrument flown on the TIROS-N series. The HRPT broadcasts at the old data rate of 665.5 kbps with the new AMSU data replacing what were previously spare words.
- The Advanced Very High Resolution Radiometer (AVHRR/3) provides spectral and gain changes to the visible channels that will allow improved low energy/light detection and adds a sixth channel, called 3A, at 1.6 μm for improved snow and ice discrimination. Channel 3A will be time shared with the previous 3.7 μm

channel, now called channel 3B. The Automatic Picture Transmission (APT) user sees channel 3B as channel 6 using the wedge six grayscale modulation index.
- The High Resolution Infrared Radiation Sounder (HIRS/3) has spectral channel changes that were made primarily to improve soundings and to be congruent with the specifications developed for the GOES-I through M Sounders. The HIRS/3 cooler set point was decreased to approximately 100° K, which will improve the two infrared detectors' performance.
- The Space Environment Monitor (SEM-2) has improved calibration and particle detection capabilities. The Total Energy Detector (TED) measures to a lower energy of 0.05 KeV, and the TED integral F (ALPHA) has two ranges of 0.05–1 and 1–20 KeV. The Medium Energy Proton and Electron Detector (MEPED) has a fourth omnidirectional proton measure at 140 MeV.
- The Solar Backscatter Ultra Violet Spectral Radiometer (SBUV/2) has undergone relatively modest improvements. Its Programmable Read Only Memory (PROM) will be changed to a Random Access Memory (RAM) due to parts obsolescence and to provide more operational flexibility.
- The Data Collection System (DCS) data rate increased from 1,200 to 2,560 bps, and the number of Data Recovery Units (DRUs) doubled from 4 to 8. DCS-2 bandwidth increased from 24 kHz to 80 kHz.
- The Search and Rescue Processor (SARP-2) Data Recovery Units increased from 2 to 3 to handle more global distress messages and to better detect interfering signals.

NOAA-19, designated NOAA-N (NOAA-N Prime) prior to launch, was launched on February 6, 2009, and is the last of NOAA's POES series of weather satellites. NOAA-19 carries a suite of instruments that provides data for weather and climate predictions. Like its predecessors, NOAA-19 provides global images of clouds and surface features and vertical profiles of atmospheric temperature and humidity for use in numerical weather and ocean forecast models, as well as data on ozone distribution in the upper part of the atmosphere, and near-Earth space environments – information important for the marine, aviation, power generation, agriculture, and other communities. The NOAA-19 primary instruments – the Advanced Very High Resolution Radiometer (AVHRR/3), High Resolution Infrared Radiation Sounder (HIRS/4), and the Advanced Microwave Sounding Unit (AMSU-A) – were all designed for a 3-year mission. The Solar Backscatter Ultraviolet Spectral Radiometer (SBUV/2) was designed for a 2-year mission, and the Microwave Humidity Sounder (MHS) was designed for a 5-year mission.

The POES series of satellites will be followed by a series of Earth observation satellites with greatly improved instrumentation, the National Polar-orbiting Operational Environmental Satellite System (NPOESS). The NPOESS Preparatory Project (NPP) satellite was launched on October 28, 2011. NPP's instruments are described later in this chapter.

Consideration of Private Initiatives Involving Polar-Orbiting Satellites

The planning for the Geosynchronous R through U GOES satellites is quite advanced, but the spacecraft options concerning the polar-orbiting satellites for the USA are now being explored, including possible commercial provisioning. One option involves the idea of hosted payloads. Currently the Iridium next constellation satellites are projected to provide a significant cost savings to the FAA by carrying a full set of hosted payloads for position determination of aircraft.

The US Space Policy Statement of June 28, 2010, indicated that when commercial alternatives were available and such options would provide cost savings for governmental space projects then such alternatives should be explored. NOAA has thus sent out an official Request for Information (RFI) seeking to hear of options by October 1, 2015, as to possible commercial offering that might involve: (i) buying of data, (ii) hosted payloads, (iii) rideshares, and (iv) launch services. This RFI is particularly in the context of the possible gap in NOAA's service provision by polar-orbiting satellites that may occur in the next few years. This RFI was clearly labeled as "predecisional," and thus, it is not clear whether going forward some element of commercial program involvement may be possible in the above-mentioned four areas of potential cooperative commercial programs.

Initial Joint Polar-Orbiting Operational Satellite (IJPS) System

Building upon the Polar-orbiting Operational Environmental Satellite (POES) program, an agreement is in place between NOAA and EUMETSAT on what is called the Initial Joint Polar-orbiting operational Satellite (IJPS) System. This program includes two series of independent but fully coordinated NOAA and EUMETSAT satellites. This program involves the exchange of instruments and global data, cooperation in algorithm development, and near real-time direct broadcasting. Under terms of the IJPS agreement, NOAA provides NOAA-18 and NOAA-19 satellites for flight in the afternoon (PM) orbit and EUMETSAT provides MetOp-A and MetOp-2 (B) satellites for flight in the mid-morning orbit (AM). These satellites carry a common core of instruments that includes (Initial Joint Polar-Orbiting Operational Satellite System [IJPS]) (IJPOS) (NOAA History, 2011):

- Third Generation Advanced Very High Resolution Radiometer (AVHRR/3): A six-channel imaging radiometer to detect energy in the visible and IR portions of the electromagnetic spectrum
- High Resolution Infrared Radiation Sounders (HIRS/4): A multispectral atmospheric sounding instrument to measure scene radiance in the IR spectrum

- Advanced Microwave Sounding Unit (AMSU-A): A cross-track scanning total power radiometer to measure scene radiance in the microwave spectrum
- Data Collection System (DCS): To collect and store environmental study data from multiple platforms for transmission to the ground once per orbit to NOAA Command and Data Acquisition stations
- Search and Rescue Satellite-Aided Tracking (SARSAT) Instruments: These are part of the international COSPAS-SARSAT system designed to detect and locate Emergency Locator Transmitters (LET), Emergency Position-Indicating Radio Beacons (EPIRB), and Personal Locator Beacons (PLB) operating at 121.5 MHz, 243 MHz, and 406 MHz to subsequently downlink to a Local User Terminal
- Space Environmental Monitor (SEM): Provides measurements to determine the intensity of the Earth's radiation belts and the flux of charged particles at satellite altitude
- Microwave Humidity Sounder (MHS): A five-channel microwave instrument to measure profiles of atmospheric humidity

In addition, NOAA satellites fly a Solar Backscatter Ultraviolet (SBUV) Radiometer instrument (a nadir pointing, nonspatial, spectrally scanning, ultraviolet radiometer), while EUMETSAT's additional payloads include an infrared interferometer sounder, a scatterometer, an ozone instrument, and a Global Positioning System (GPS) occultation sounder.

Coordination on associated ground segments included in this agreement ensures the sharing of all mission data, blind-orbit data capture support, and telecommunications paths through each other's ground stations for backup command and control functions. The first MetOp satellite was launched on October 19, 2006, from Baikonur Cosmodrome, Kazakhstan. See chapter "▶ Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing" for a more detailed description of the meteorological satellite program of EUMETSAT.

The Defense Meteorological Satellite Program (DMSP): Another Asset

The spacecraft of the Defense Meteorological Satellite Program (DMSP) can "see" the best among all weather satellites with its ability to detect objects almost as "small" as a large oil tanker. Some of the most spectacular photos have been recorded by the night visible sensor; city lights, volcanoes, fires, lightning, meteors, oil field burn offs, as well as the Aurora Borealis and Aurora Australis have been captured by this 830-km-high space vehicle's low moonlight sensor (DMSP).

At the same time, energy monitoring as well as city growth can be accomplished since major and even minor cities, as well as highway lights, are conspicuous. This also informs astronomers of light pollution. In addition to monitoring city lights, these photos are a life-saving asset in the detection and monitoring of fires. Not only do the satellites see the fires visually day and night, but the thermal and infrared scanners on board these weather satellites detect potential fire sources below the surface of the Earth where smoldering occurs. Once the fire is detected, the same weather satellites provide vital information about wind that could fan or spread the fires.

NOAA also currently operates the Defense Meteorological Satellite Program near-polar-orbiting series of satellites. The satellites were initiated by the Defense Department in the mid-1960s and were initially the responsibility of the US Air Force. Each DMSP satellite, orbiting at approximately 516 miles (830 km) above the Earth, crosses any point on the Earth up to twice a day. These satellites see such environmental features as clouds, bodies of water, snow, fire, and pollution in the visible and infrared spectra. Scanning radiometers record information that can help determine cloud type and height, land and surface water temperatures, water currents, ocean surface features, ice, and snow. Communicated to terminals on the ground, the data are processed, interpreted by meteorologists, and used in planning and conducting US military operations worldwide. On May 5, 1994, however, President Bill Clinton decided to merge America's military and civil polar-orbiting operational meteorological satellite systems into a single, national system that could satisfy both civil and national security requirements for space-based environmental data. Called the National Polar-orbiting Operational Environmental Satellite System (NPOESS), NOAA is responsible for the now integrated network (Meteorological satellites).

New and Future NOAA Satellites: The Joint Polar Satellite System (JPSS)

On February 1, 2010, the Executive Office of the President restructured the National Polar-orbiting Operational Environmental Satellite System (NPOESS) into two separate development programs: one aimed at the civilian community, the Joint Polar Satellite System (JPSS), and the Defense Weather Satellite System (DWSS) to satisfy Defense Department requirements. The civilian and scientific community program is led by NOAA who sets the requirements and NASA who is directing the acquisition. JPSS will provide operational continuity of satellite-based polar missions in the afternoon orbit that support its civil regional and global weather and climate requirements. In addition, JPSS will provide oceanographic, environmental, and space environmental information. The system's instrumentation is described below. The Soumi NPP (The Soumi National Polar-orbiting Partnership) Satellite that is named after University of Wisconsin professor Soumi has been in orbit some 5 years. The full network is described in the chart below. The entire three satellite networks plus the experimental test satellite are several years away from full deployment, and thus, this is why NOAA is exploring the commercial options noted above.

Full JPSS Network		
Satellite	**Satellite Full Name**	**Launch Date**
Suomi NPP	**Suomi National Polar-orbiting Partnership**	October 2011
JPSS-1	**Joint Polar Satellite System-1**	2017
JPSS-2	**Joint Polar Satellite System-2**	To be decided
TCTE	**TSI Calibration Transfer Experiment**	To be decided

Planned JPSS Satellite Technical Characteristics

The currently planned capabilities for the JPSS satellites include six key subsystems as follows:

- *Visible/Infrared Imager/Radiometer Suite (VIIRS)*: VIIRS is an electro-optical imager having multiband imaging capabilities which collects calibrated visible/infrared radiances to produce data products for cloud and aerosol properties, land surface type, vegetation index, ocean color, land and sea-surface temperature, and low light visible imagery. The 22-channel VIIRS will fly on the NPOESS Preparatory Project (NPP) and on all JPSS platforms to provide complete daily global coverage over the visible, short/medium-infrared, and long-wave infrared spectrum at horizontal spatial resolutions of 370 and 740 m at nadir. VIIRS is the primary instrument for 21 different types of environmental data records.
- *Cross-track Infrared Sounder (CrIS)*: CrIS is a Fourier Transform Spectrometer that uses a Michelson interferometric sounder capable of sensing upwelling infrared radiances from 3 to 16 μm at very high spectral resolution (∼1,300 spectral channels) to determine the vertical atmospheric distribution of temperature, moisture, and pressure from the surface to the top of the atmosphere across a swath width of 2,200 km.
- *Advanced Technology Microwave Sounder (ATMS)*: ATMS is a cross-track high-spatial-resolution microwave sounder. ATMS data will support temperature and humidity sounding generation in cloud-covered conditions. ATMS has 22 microwave channels to provide temperature and moisture sounding capabilities in the 23/31, 50, 89,150, and 183 GHz spectral range.
- *Ozone Mapping and Profiler Suite (OMPS)*: The OMPS monitors ozone from space. OMPS will collect total column and vertical profile ozone data and continue the daily global data produced by the current ozone monitoring systems, the Solar Backscatter Ultraviolet radiometer (SBUV)/2, and Total Ozone Mapping Spectrometer (TOMS), but with higher fidelity. The nadir sensor uses a wide field-of-view push-broom telescope to feed two separate spectrometers. The nadir total column spectrometer (mapper) measures the scene radiance between 300 and 380 nm with a resolution of 1 nm sampled at 0.42 nm and a 24-h ground revisit time. The limb sensor measures the along-track limb scattered solar radiance with 1 km vertical sampling in the spectral range of 290–1,000 nm. Three vertical slits sample the limb at 250 km cross-track intervals to provide for

better than 7-day ground revisit times to improve the precision of the ozone profiles. The three slits are imaged onto a single charge-coupled device (CCD) (identical to both nadir CCDs). Due to limitations with flight hardware transferred from NPOESS to JPSS, the OMPS on JPSS J1will consist of a nadir sensor only.

- *Cloud and Earth Radiant Energy System (CERES*): The CERES instrument seeks to develop and improve weather forecast and climate models prediction, to provide measurements of the space and time distribution of the Earth's Radiation Budget (ERB) components, and to develop a quantitative understanding of the links between the ERB and the properties of the atmosphere and surface that define that budget. CERES consists of three broadband radiometers that scan the earth from limb to limb. Data from CERES will be used in conjunction with VIIRS to study changes in the Earth's energy balance and key changes in clouds and aerosols to determine the effect of changing clouds on the Earth's energy balance.
- *Total Solar Irradiance Sensor (TSIS*): The Total Solar Irradiance Sensor (TSIS) will measure variability in the sun's solar output, including total solar irradiance. TSIS consists of two instruments: the Total Irradiance Monitor (TIM) that measures the total light coming from the sun at all wavelengths and the Spectral Irradiance Monitor (SIM) that will measure how the light from the sun is distributed by wavelength. These measurements are needed to understand how solar radiation interacts with the Earth's surface and atmosphere. TSIS is an important climate sensor that will help maintain continuity of the climate data record for space-based solar irradiance measurements that now spans over three decades.

JPSS is designed to ensure continuity of crucial climate observations and weather data in the future. Data and imagery obtained from the JPSS will increase timeliness and accuracy of public warnings and forecasts of climate and weather events reducing the potential loss of human life and property damage. The data collected by JPSS will contribute to the unified and coherent long-term environmental observations and products that are critical to climate modelers and decision makers concerned with advancing climate change understanding, prediction, mitigation, and adaptation strategies, policies, and science. JPSS, with its global view, will play a vital role in continuing these climate data records for the USA and the international community.

The NPP Satellite

As mentioned above, the NPOESS Preparatory Project (NPP) satellite was launched on October 28, 2011. NPP will serve as an important link between the current generation of Earth-observing satellites and the next generation of climate and weather satellites of the JPSS. NPP observes the Earth's surface twice every 24-h day, once in daylight and once at night. In its orbit, NPP flies 512 miles (824 km) above the surface in a polar orbit, circling the planet about 14 times a day. NPP sends its data once an orbit to the ground station in Svalbard, Norway, and continuously to local direct broadcast users.

Of the six JPSS subsystems described above, NPP carries five instruments, VIIRS, CrIS, ATMS, OMPS, and CERES to monitor the environment on Earth and the planet's climate (NPP building a bridge).

NPP measurements will be used to map land cover and monitor changes in vegetation productivity. NPP tracks atmospheric ozone and aerosols as well as takes sea and land surface temperatures. NPP monitors sea ice, land ice, and glaciers around the world. In addition to continuing these data records, NPP is also able to monitor natural disasters such as volcanic eruptions, wildfires, droughts, floods, dust storms, and hurricanes/typhoons. In all, NPP monitors the health of Earth from space – providing continuity to decades-long records and setting the stage for future Earth science missions.

An important design feature of NPP's VIIRS instrument is that it tracks land cover changes and vegetation productivity, extending the successful and widely used data records of NASA's Moderate Resolution Imaging Spectroradiometer (MODIS), a similar instrument launched aboard NASA's Terra and Aqua spacecraft in 1999 and 2002.

NPP's CERES instrument continues the Earth radiation budget data record started by the Earth Radiation Budget (ERB) instrument on the Nimbus-7 satellite in 1978 and continued through a series of NASA satellites since then, including CERES instruments on the satellites Terra and Aqua.

The CrIS and the ATMS instruments on board NPP work together, providing global high-resolution profiles of temperature and moisture. These advanced atmospheric sensors create cross sections of storms and other weather conditions, helping with both short-term "nowcasting" and long-term forecasting. CrIS measures continuous channels in the infrared region and has the ability to measure temperature profiles with improved accuracy over its predecessor instruments on operational satellites and comparable accuracy to the Atmospheric Infrared Sounder (AIRS) on Aqua. NOAA will be using CrIS for numerical weather prediction, and because it is a brand new instrument, its use on NPP provides a real-world test of the equipment before NOAA's upcoming Joint Polar Satellite System (JPSS) missions. The ATMS instrument works in both clear and cloudy conditions, providing high-spatial-resolution microwave measurements of temperature and moisture. ATMS has better sampling and two more channels than previous instruments like the Advanced Microwave Sounding Units (AMSU), and it combines all of their abilities into one instrument. Working in concert, CrIS and ATMS together comprise the Cross-track Infrared Microwave Sounding Suite (CrIMSS).

The OMPS instrument measures the ozone layer in our upper atmosphere, tracking the status of global ozone distributions, including the in "ozone hole" region. It also monitors ozone levels in the troposphere, the lowest layer of our atmosphere. OMPS will extend a 40-year-long record of ozone layer measurements while also providing improved vertical resolution compared to previous operational instruments. Closer to the ground, OMPS's measurements of harmful ozone will improve air quality monitoring and, when combined with cloud predictions, help to

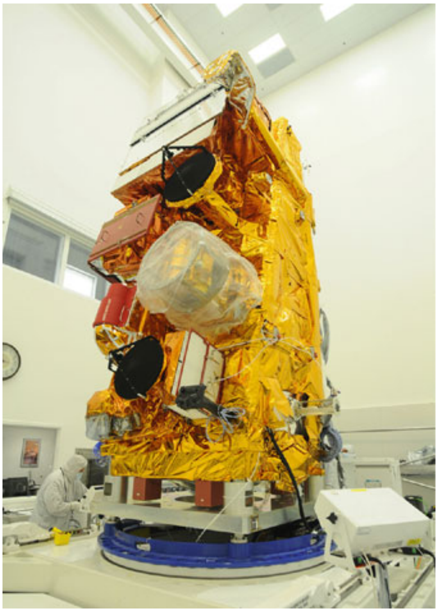


Fig. 5 The NPP satellite prior to its Oct 2011 launch (Photograph courtesy of NASA)

create the Ultraviolet Index, a guide to safe levels of sunlight exposure for people. The complexity of the NPP satellite can be appreciated in the photograph shown as Fig. 5 when the NPP satellite was in its testing phase prior to its launch.

The GOES and POES Ground Systems

The GOES Ground System is, in fact, a “System-of-Systems” that comprises the end-to-end framework for collecting, processing, and disseminating critical environmental data and information from the satellites. It supports the launch, activation, and evaluation of new satellites and the in-depth assessments of satellite data quality. Data from the satellites are received at ground facilities, where the data are processed to monitor and control the satellite and to generate products that are used by NOAA, its users, and the world meteorological community. The GOES ground system consists of components at the Satellite Operations Control Center (SOCC) at Suitland, Maryland; Command and Data Acquisition (CDA) facilities at Wallops, Virginia, and Fairbanks, Alaska; and Wallops Backup (WBU) facility at NASA Goddard Space Flight Center (GSFC) in Greenbelt, Maryland.

The POES mission operates with a NOAA-provided constellation of two operational satellites in circular, near-polar, sun-synchronous orbits that provide scheduled downloads of environmental data collected from space to the POES Ground System for satellite monitoring and control and mission processing, analysis, and distribution. The POES Ground System is also a “System-of-Systems” that includes collecting, processing, and disseminating critical environmental data and information from the POES satellites. Operational elements are located at Fairbanks, Alaska; Wallops, Virginia; and Suitland, Maryland. It contains subsystems located in the following NESDIS Offices: Office of Satellite Operations (OSO), Office of Research and Applications (ORA), and the NOAA National Data Centers (NNDC). See chapter "▶ Ground Systems for Satellite Application Systems for Navigation, Remote Sensing, and Meteorology" for a more detailed description of the ground system of the meteorological satellite program of the USA.

Deep Space Climate Observatory

On February 11, 2015, the so-called Deep Space Climate Observatory (DSCOVR) was launched on a Falcon 9 launcher into a deep space location in the L-1 Lagrange Point. This is a relatively stable gravitational point between the Sun and Earth that is 1,600,000 km or 1 million miles from Earth or four times further than lunar orbit. The purpose of this satellite is to monitor the Earth's atmosphere and variations in the ozone layer on one hand and to also monitor solar activity including the more violent storms that occur, especially during solar max. This satellite was first proposed by Vice President Al Gore and was initially named "Triana" which was the name of the person in the crow's nest that first saw land on Columbus' first voyage to the New World.

DSCOVR will give NOAA's Space Weather Prediction Center (SWPC) forecasters higher-quality measurements of solar wind conditions, improving their ability to monitor and warn of severe and potentially dangerous space weather events. This will be key in being able to provide alerts with regard to coronal mass ejections that can wipe out electrical power grids and in-orbit satellites. The DSCOVR satellite is to replace NASA's 17-year-old ACE research satellite as America's primary warning system for solar magnetic storms and solar wind data. (ACE will continue its role in space weather research).

"DSCOVR will be our eyes on the sun and give us early warning when it detects a surge of energy that could trigger a geomagnetic storm destined for Earth," said Stephen Volz, Ph.D., assistant administrator for NOAA's Satellite and Information Service (NOAA now).

Conclusion

The US Meteorological Satellite System is now essentially integrated under the operation of the National Oceanic and Atmospheric Administration with NASA assisting with the design and launch of these satellites and the US Department of Defense, and particularly the National Reconnaissance Office and US Air Force, assisting with regard to the design and operation of the Defense Meteorological Satellite Program (DMSP). The US meteorological system consists of a combination of geosynchronous imaging and sounder satellites (GEOS system) that provide a near-real-time image of the Earth disk on a 24 h a day (i.e., day and night) basis, while the POES system provides a sun-synchronous image of the entire Earth with some 14 passes over the Earth for each satellite. The combination of two US satellites and two European satellites provide very rapid updating of information. The new JPSS system will provide expanded capability. NOAA is currently exploring possible commercial program options with regard to polar-orbiting, sun-synchronous satellite services pending the full deployment of the JPSS network.

The coordination of various international programs is accomplished via the World Meteorological Organization. Chapter "▶ International Meteorological Satellite Systems" presents the status of various national and regional meteorological satellite systems and the global coordination of these systems.

Cross-References

▶ International Meteorological Satellite Systems
▶ Introduction to Space Systems for Meteorology

References

DMSP-Defense Meteorological Satellite Program, http://www.ngdc.noaa.gov/dmsp/index.html. Last viewed on 2 Jan 2016

GOES I/M Brochure, http://goes.gsfc.nasa.gov/text/goesimbroch.html. Last viewed on 2 Jan 2016

GOES I/M Imager, http://goes.gsfc.nasa.gov/text/imager.html. Last viewed on 2 Jan 2016

GOES I/M Sounder, http://goes.gsfc.nasa.gov/text/sounder.html. Last viewed on 2 Jan 2016

Meteorological satellites, http://www.centennialofflight.gov/essay/SPACEFLIGHT/metsats/SP35.htm. Last viewed on 2 Jan 2016

NOAA History, 2011 by Gary Davis, History of the NOAA Satellites –Initial Joint Polar Orbit Satellite System), http://goes.gsfc.nasa.gov/text/history/History_NOAA_Satellites.pdf (Last viewed July 14, 2016

NOAA now in DSCOVR's "Driver Seat" as NASA Officially Hands over Command, http://www.nesdis.noaa.gov/DSCOVR/. Last viewed on 2 Jan 2016

NPP building a bridge to a new era of earth observations, http://www.nasa.gov/mission_pages/NPP/mission_overview/index.html

International Meteorological Satellite Systems

Sergio Camacho-Lara, Scott Madry, and Joseph N. Pelton

Contents

S. Camacho-Lara (✉)
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Santa María Tonantzintla, Puebla, Mexico
e-mail: sergio.camacho@inaoep.mx

S. Madry
Global Space Institute, Chapel Hill, NC, USA
e-mail: Scottmadry@mindspring.com

J.N. Pelton
International Space University, Arlington, VA, USA
e-mail: peltonjoe@gmail.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_56

Abstract

The oldest and most extensive meteorological satellite systems are those of the USA and of Europe, as operated by the Eumetsat system. These are addressed in detail in the preceding two chapters. This chapter describes the meteorological satellite systems of China, India, Japan, Russia, and South Korea. These meteorological satellite systems are extensive and provide a number of sophisticated meteorological satellite sensing capabilities both from geostationary and polar-orbiting satellite systems. Today all of these various satellite systems – those of China, Europe, India, Japan, Russia, South Korea, and the USA – are in various manners linked together and share data. This international coordination of meteorological data is accomplished through the World Weather Watch (WWW) programme of the World Meteorological Organization and the Coordination Group for Meteorological Satellites (CGMS).

These international cooperative efforts – supplemented by bilateral or regional agreements – allow a degree of standardization with regard to the formatting and display of meteorological data and a systematic process for sharing of vital weather data. This sharing of meteorological data is important on an ongoing basis – but this can be particularly important – when there is a failure of a meteorological satellite, a launch failure, or a delay in the deployment of a replacement satellite. In some cases, countries such as the USA have even "loaned" meteorological satellites to other countries when failures or launch delays have created gaps in critical coverage areas.

The various international satellites around the world that are deployed in different orbital locations and with varying periodicity provide a very useful redundancy of coverage that is particularly important in tracking major storms and obtaining the most up-to-date information of atmospheric, oceanic, and of arctic conditions.

This chapter provides a description of the meteorological satellite systems of China, India, Japan, South Korea, and Russia and their current status. Researchers can also consult the various universal reference locations (i.e., URLs) for these various meteorological satellite systems which can be useful in obtaining the more recent information about the deployment and operation of these systems.

Keywords

China Meteorological Administration (CMA) • Geostationary Operational Meteorological Satellite (GOMS) • Elektro Satellites of Russia • Fengyun Meteorological Satellite System of China • INSAT System of India • Himawari System of Japan • Japanese Geostationary Meteorological Satellite (GMS) Systems • Communications, Ocean, and Meteorological Satellite (COMS) of South Korea • Meteor Satellites of Russia • MTSAT of Japan • World Meteorological Organization (WMO) • World Weather Watch (WWW)

Introduction

Weather is of vital interest to people, as it affects agriculture, industry, transportation, and many of our daily life activities. Violent weather sometimes threatens our safety and even our lives when extreme meteorological events come upon us with little warning, such as hurricanes, typhoons, ice storms, tornados, and tropical and winter storms, and affect the areas where we live. There are also weather events that are longer time in the making. These are flooding and drought events that can be due to periodic phenomena like El Niño or La Niña or may be due to weather patterns caused by climate change.

For thousands of years, people have been observing weather patterns to determine when to plant, travel, store food, and even how to use the acquired knowledge for strategic military advantage. These observations, rudimentary at the beginning, gave rise to the discipline called meteorology, which at first was based on records of in situ obtained data such as temperature, precipitation, wind speed, direction, and solar radiation.

To predict the weather, modern meteorology depends upon the acquisition of in situ and space-obtained data and on near instantaneous exchange of weather information across the entire globe. To better understand the global climate system and to anticipate its future evolution, not only do we need global data sets, we must also coordinate climate analysis, modeling, and predictions on an equally global basis in order to establish the correct climate state and to create powerful modeling tools for climate prediction.

The World Weather Watch Programme

The World Weather Watch (WWW) was established in 1963, and this activity represents the core of the WMO's combined observing system. The WWW includes the Global Observing System (GOS), the Global Telecommunication System (GTS) and the Global Data-processing and Forecasting System (GDPFS) that are operated by its members. The WWW, thus, serves to make available meteorological and related environmental information to all countries. The WWW is a unique achievement in international cooperation. In few other fields of human endeavor has there ever been such a truly worldwide operational system to which virtually every country in the world contributes for the common benefit of humankind and does so every day of the year for decades on end. These arrangements, as well as the operation of the WWW facilities, are coordinated and monitored by WMO with a view to ensuring that every country has available all of the information it needs to provide weather services on a day-to-day basis as well as for long-term planning and research (World Weather Watch).

An increasingly important part of the WWW Programme provides support for developing international cooperation related to global climate and other environmental issues and to sustainable development.

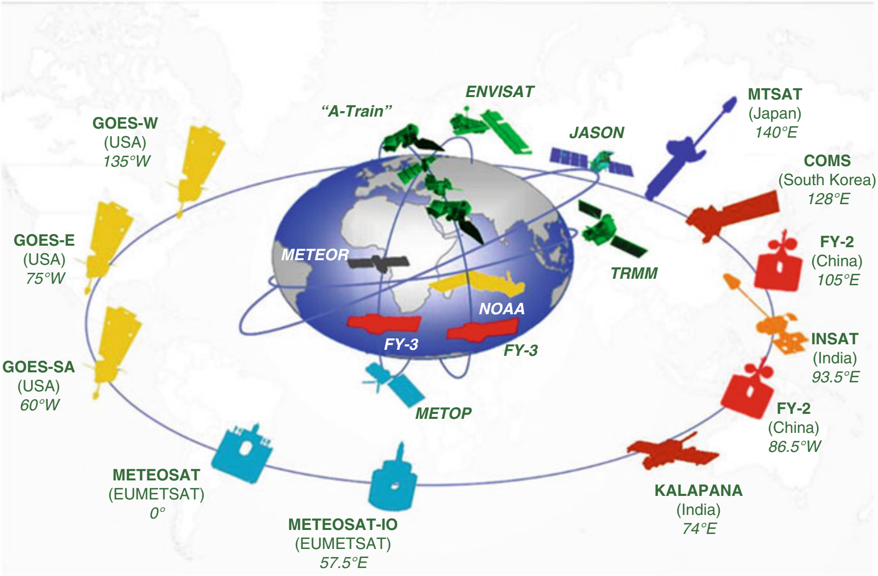


Fig. 1 The space-based component of the global observing system (Courtesy WMO)

As a result of the implementation of the WWW, a significant number of projects and activities have evolved from the need to efficiently coordinate efforts and to make the data acquired and information derived from satellite data available to those who need them or can use them to further knowledge of weather-related phenomena.

The global observing system (GOS) is made up of land, ocean, air and space observation systems that are owned by the Member countries of WMO and provides observations of the state of the atmosphere and ocean surface for the preparation of weather analyses, forecasts, advisories, and warnings for climate monitoring and environmental activities. It is operated by national meteorological services and national or international satellite agencies. Developing the space-based part of the GOS is one of the main components of the WMO Space Programme (Global Observing System).

The space-based global observing system includes three operational near-polar-orbiting satellites and six operational geostationary environmental observation satellites as well as several research and development satellites. These systems are shown in Fig. 1.

Polar-orbiting and geostationary satellites are normally equipped with visible and infrared imagers and sounders, from which many meteorological parameters can be derived. Several of the polar-orbiting satellites are equipped with sounder instruments that can provide vertical profiles of temperature and humidity in cloud-free areas. Geostationary satellites can be used to measure wind velocity in the tropics by tracking clouds and water vapor. Satellite sensors, communications, and data assimilation techniques are evolving steadily so that better use is being made of the vast amount of satellite data. Improvements in numerical modeling, in particular, have made it possible to develop increasingly sophisticated methods of deriving the temperature and humidity information directly from the satellite radiances. Research and development (R&D) satellites comprise the newest constellation in the space-based component of the GOS. R&D missions provide valuable data for operational use as well as for many WMO supported programmes. Instruments on R&D missions either provide data not normally observed from operational meteorological satellites or improvements to current operational systems.

GOS also includes solar radiation observations, lightning detection, and tide-gauge measurements. In addition, wind-profiling and Doppler radars are proving to be extremely valuable in providing data of high resolution in both space and time, especially in the lower layers of the atmosphere. Wind profilers are especially useful in making observations at times between balloon-borne soundings and have great potential as a part of integrated networks. Doppler radars are used extensively as part of national and increasingly of regional networks, mainly for short range forecasting of severe weather phenomena. Particularly useful is the Doppler radar capability of making wind measurements and estimates of rainfall amounts.

While programmatic coordination is done through the WWW, coordination regarding compatibility and complementarity among polar-orbiting and geostationary meteorological satellites is done through the Coordination Group for Meteorological Satellites.

The Coordination Group for Meteorological Satellites (CGMS) came into being on September 19, 1972, when representatives of the European Space Research Organization (since 1975 called the European Space Agency, ESA), Japan, the USA, observers from the World Meteorological Organization (WMO), and the Joint Planning Staff for the Global Atmosphere Research Programme met in Washington to discuss questions of compatibility among geostationary meteorological satellites.

CGMS provides an international forum for the exchange of technical information on geostationary and polar-orbiting meteorological satellite systems. It consists of 15 member organizations and two observers. The members of the CGMS are China Meteorological Administration (CMA), Centre National d'EtudesSpatiales (CNES), China National Space Administration (CNSA), European Space Agency (ESA), EUMETSAT, India Meteorological Department (IMD), Intergovernmental Oceanographic Commission/UNESCO (IOC/UNESCO), Japan Aerospace Exploration Agency (JAXA), Japan Meteorological Agency (JMA), Korea Meteorological Administration (KMA), National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), Russian Federal Space Agency (ROSCOSMOS), Russian Federal Service for Hydrometeorology and Environmental Monitoring (ROSHYDROMET), and World Meteorological Organization (WMO). EUMETSAT has run the Secretariat since 1987.

The CGMS Secretariat represents CGMS Members in a number of other international coordination bodies such as the Committee on Earth Observation Satellites (CEOS) and its related Earth Observation International Coordination Working Group (EO-ICWG), the Group on Earth Observation (GEO), and the Space Frequency Coordination Group (SFCG).

The meteorological polar-orbiting and geostationary satellites are provided by States through their national and international organizations. The meteorological systems of the USA and Europe (through EUMETSAT) are presented in previous chapters. The sections that follow present the meteorological polar-orbiting and geostationary satellite systems of China, India, Japan, South Korea, and Russia.

China: The Fengyun Meteorological Satellite System

The China Meteorological Administration (CMA) has responsibility for weather forecasting and monitoring of meteorological conditions including those related to climate change. The National Satellite Meteorological Center (NSMC), which is affiliated to the CMA, is responsible for the operation of China's meteorological satellite network.

China has a number of international agreements to obtain meteorological information from other countries' satellite systems that are coordinated through the World Meteorological Organization. Nevertheless, for strategic reasons, China has been for some years implementing a fully functional national satellite system of its own that is known as the Fengyun Meteorological Satellite System. Fengyun means "wind" and "cloud." The Fengyun I series has been fully deployed since September of 1988 with the launch of FY-1A followed by FY-1B in September of 1990. This satellite network is a polar-orbiting meteorological system. This was then followed by the Fengyun 2 series which was fully deployed as of early 2012. The Fengyun 2 is a geostationary-orbiting satellite system. The odd number series is the polar-orbiting satellite series, the even number series is the geostationary. Each satellite is followed by a letter indicating, in alphabetical order, its launching sequence. For instance, 'FY-2B' identifies the second satellite that has been launched in the FY-2 geostationary series. The last in the Fengyun II series of geostationary satellites the FYII-7, renamed as FY-2 F, was successfully launched on a Long March 3 launch vehicle on January 13, 2012 and subsequently placed at 112°E above the equator (National Satellite Meteorological Center of CMA).

Figure 2 shows the launch of FYII-7.

The CMA is now in the process of implementing the Fengyun III meteorological satellite series. This is an upgraded polar-orbiting meteorological satellite network that constitutes the replacement for the initial Fengyun I series (Fenyun 3).

According to the China Meteorological Administration, the Fengyun III satellite series will have a number of expanded capabilities over the Fengyun I series. The defined objectives for the Fengyun network, once it is deployed in orbit, will be as follows:

- To provide global measurements of temperature gradients in three dimensions
- To collect moisture soundings of the atmosphere and thereby to measure cloud and precipitation parameters in support of numerical weather prediction (NWP)

Fig. 2 The launch of the geostationary Fengyun II-7 satellite in January 2012 (Graphics courtesy of Chinese National Space Agency)

- To provide global imagery of large-scale meteorological and/or hydrological events and biosphere environment anomalies (by integrating imaging from FY-III and FY-II satellites)
- To provide geophysical parameters in support of global meteorological change and climate monitoring
- To provide global and local meteorological information for specialized meteorological users working in such areas as aviation, marine transportation, and fishing
- To collect and relay environmental data from the ground segment to national and international users and scientific analysts

The FY-III operational network will initially consist of two polar-orbiting satellites. These satellites will be deployed in phased orbits so that one of the satellites provides coverage in the daytime (AM) and the other in the evening (PM). The payload will be different for AM/PM satellites with sensors optimized for operation in the sunlight and for the one designed for nighttime operations. The appropriate time slots for the AM and PM satellite operations have been coordinated through the World Meteorological Organization (WMO) (FY-3 (Fengyun 3)).

Table 1 Overview of Fengyun-3 spacecraft series of CMA/NSMC

Spacecraft	Launch (projected launch)	LTDN (local time on descending node) (h)	Mission service type
FY-3A	*May 27, 2008*	10:00	R&D (experimental)
FY-3B	*Nov. 04, 2010*	14:00	R&D (experimental)
FY-3C	*Sept. 23, 2013*	10:00	Operational
FY-3D	2017	14:00	Operational
FY-3E	2018	10:00	Operational
FY-3F	2019	14:00	Operational

The FY-III series began with a developmental phase. In this developmental phase, the FY-IIIA satellite (launched on May 27, 2008) and the FY-IIIB satellite (launched November 4, 2010) gathered experience but with sounders that had limited capabilities. Table 1 shows the dates of launch and projected launch and the type of mission of the FY-III series of satellites.

The second generation of China's polar-orbiting meteorological satellite (FY-3), with a three-axis stabilization mode, carries 11 observation sensors and provides the functions of global, all-weather, multispectral, three-dimensional, and quantitative Earth observations. A description of the sensors, as well as access to the data, is provided by the Fengyun Satellite Data Center (Fenyun data center aspx).

India: The INSAT Satellite System

The Indian Space Research Organization (ISRO) has deployed INSAT meteorological satellites for three decades. The first INSAT 1-A satellite was launched on April 10, 1982. This spacecraft was a hybrid geostationary satellite that was capable of providing telecommunications services, but it also contained a meteorological package. Following a misfunction in INSAT 1-A, an identical INSAT 1-B was launched on August 30, 1983. On April 3, 1999 the INSAT 2E, an upgraded hybrid communications and meteorological satellite, was also launched. Up to this point, all of the satellites deployed by India for meteorological sensing had been hybrid telecommunications satellites that also included a meteorological package. Further, all of these satellites had been supplied by overseas suppliers.

On September 12, 2002 the Indian Space Research Organization launched a dedicated meteorological satellite that had been designed and manufactured within ISRO. This satellite was initially named Metsat but was renamed Kalpana-1 in 2003 in the honor of the Indian-born American astronaut Dr. Kalpana Chawla who perished in the Columbia shuttle accident. This satellite – like all of the spacecraft in the INSAT satellite series – was launched into geostationary orbit. This Indian designed and manufactured spacecraft was also unique in that it was launched on the Indian polar satellite launch vehicle in its first launch and successful mission. This dedicated Kalpana-1 satellite contains a very high resolution radiometer (VHRR)

Fig. 3 Kalpana-1 meteorological satellite designed and launched by ISRO (Graphics courtesy of ISRO)

that operates in the visible bands as well as the thermal and water-vapor infrared bands. It also contains a data relay transponder to transmit data to a number of ground locations (Fig. 3).

On April 28, 2003 the INSAT 3A, placed at 93.5°E longitude, another hybrid satellite capable of providing telecommunications, broadcasting, and meteorological services was launched to complete India's current meteorological satellite configuration. The INSAT 3A, as the latest of these satellites, includes very high resolution radiometers (VHRRs) operating in multibands and charge coupled device (CCD) multispectral cameras as well as a package to obtain search and rescue signals by downed pilots, ships in distress, or other emergency beacon signals. This satellite's meteorological package included the following instruments:

- A very high resolution radiometer (VHRR) with imaging capacity in the visible (0.55–0.75 μm), thermal infrared (10.5–12.5 μm), and water-vapor infrared (5.7–7.1 μm) channels. This radiometer can provide 2 × 2 km and 8 × 8 km ground resolutions, respectively.
- A CCD camera that provides 1 × 1 km ground resolution in the visible (0.63–0.69 μm), near-infrared (0.77–0.86 μm), and shortwave infrared (1.55–1.70 μm) bands.

- A data relay transponder (DRT) having global receive coverage with a 400 MHz uplink and 4,500 MHz downlink for relay of meteorological, hydrological, and oceanographic data from unattended land and ocean-based automatic data collection-and-transmission platforms.

The combined network of the INSAT 2E, the Kalpana 1, and the INSAT 3A, thus, provides comprehensive geosynchronous meteorological data for India and surrounding areas.

The entirety of the Indian meteorological satellite capabilities – unlike most other meteorological satellite networks of other countries – relies exclusively on geostationary satellite sensing and thus do not include polar-orbiting meteorological satellites to obtain data from much lower orbits. For most national applications, the 1 × 1 km resolution is considered adequate for interpreting major weather formations. The ability to collect data from ocean buoys and land-band sensors also allows more precise interpretation of meteorological data via the data relay transponders on the Kalpana-1 and INSAT 3A satellites.

The meteorological data derived from this meteorological satellite network is processed and disseminated by the INSAT Meteorological Data Processing System (IMDPS) operated by the India Meteorological Department (IMD). The above described satellites are able to provide up-to-date information on upper atmosphere winds, sea surface temperature, and precipitation index data. The products derived from the combined network include cloud motion vectors, sea surface temperature, outgoing long-wave radiation, and quantitative precipitation indices. These products are used for weather forecasting that employs both synoptic and numerical weather prediction.

INSAT-VHRR imageries are used extensively by Indian news agencies to provide localized weather forecasts. At present, the most detailed and synoptic weather system observations over the Indian Ocean from geostationary orbit are provided by the INSAT system. INSAT's very high resolution radiometer (VHRR) data in visible and other spectral bands is currently available in near real time at 90 Meteorological Data Dissemination Centers (MDDC) in various parts of the country. With the commissioning of direct satellite service for processed VHRR data, MDDC type of data can be provided at any location in the country. A low cost and very low data rate (300 bits/s) reception unit has been developed for national users wishing to receive data directly from this and other Indian meteorological satellite packages. A cooperative agreement has been signed with EUMETSAT for using meteorological data from Meteosat-5 at 63° East in exchange for weather images collected by INSAT (Listing of ISRO Satellites).

IMD has installed 100 meteorological data collection platforms (DCPs), and other agencies have installed about 200 DCPs all over the country and even on the Indian base station in Antarctica. DCP services are provided using the data relay transponders of Kalpana-1 and INSAT-3A.

For quick dissemination of warnings against impending disaster from approaching cyclones, specially designed receivers have been installed at the vulnerable coastal areas in Andhra Pradesh, Tamil Nadu, Orissa, West Bengal, and

Gujarat for direct transmission of warnings to the officials and public in general using the broadcast capability of INSAT. IMD's area cyclone warning centers generate special warning bulletins and transmit them every hour in local languages to the affected areas. Three hundred and fifty such receiver stations have been installed by IMD.

The Japanese Geostationary Meteorological Satellite System (Himawari) and the QZSS Network

The Japan Meteorological Agency (JMA) carries out a number of missions under the Japanese Meteorological Service Act as well as the broader Act for Establishment of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT). In addition to collection of meteorological and weather-related data, JMA is charged with an active role in several specific areas. These areas include the following:

- Prevention and mitigation of natural disasters
- Safety of transportation
- Development and prosperity of industry
- Improvement of public welfare

To meet these goals, JMA focuses its efforts on monitoring the Earth's environment and forecasting natural phenomena related to the atmosphere, the oceans, and indeed the entire Earth. It is also charged with conducting research and technical development in related fields and to this end the JMA has an active partnership with the Japanese Aerospace eXploration Agency (JAXA). JMA also engages in international cooperation activities regarding both meteorology and seismology to meet Japan's international obligations and to promote partnerships with various national meteorological and hydrological services as well as with various related international agencies – particularly the World Meteorological Organization (WMO) and the United Nations Environmental Programme (UNEP).

Particular emphasis is placed on the prevention and mitigation of natural disasters, as Japan is prone to a variety of natural hazards such as typhoons, heavy rains, tsunamis, and earthquakes. JMA, as the sole national authority responsible for issuing weather/tsunami warnings and advisories, is required to provide reliable and timely information to governmental agencies and residents for the purposes of natural disaster prevention and mitigation.

In this way, JMA plays a vital role in natural disaster mitigation and prevention activities in the country through cooperation and coordination with relevant authorities, including the central government and individual local governments. Thus, in addition to the collection of meteorological data via meteorological agencies, JMA also seeks to use satellite remote sensing data to investigate earthquake, volcano, and other disaster phenomena.

JMA collects meteorological data from an extensive number of earth and ocean-based sensors as well as via upper atmosphere sensing devices. Meteorological satellites are a very important part of its overall observation and data collection process (The Mission of the Japanese Meteorological Agency).

In 1977, Japan launched its first geostationary meteorological satellite (GMS) into geostationary orbit with an orbital location to cover the Western Pacific and East Asia as part of a space-based component of the global observation system (GOS) under the auspices of the WMO World Weather Watch (WWW) programme. Since then Japan's meteorological satellite sensing capabilities have continued to expand and to date there have been five satellites in orbit. The Japanese meteorological satellites are known by a number of different names and thus, it is important to note that these satellites are variously known as the Japanese geostationary meteorological satellites (GMS), the Himawari (meaning Sunflower in Japanese) system, and multifunctional meteorological satellites (MTSATs). From July 7, 2015, the MTSAT 2 (Himawari 7) has been replaced by the Himawari 8 as the primary meteorological satellite system for Japan. During the period around 2003 and 2004, the service coverage for Japan was quite disordered for over an 18-month period. This was due to the delays in the manufacture of the MTSAT 1R satellite that was being constructed by Space Systems Loral while this corporation was going through bankruptcy. There were additional delays due to problems with the performance of the Japanese IIA launch vehicle. During this period, the USA loaned Japan the GOES 9 satellite to temporarily serve in the stead of MTSAT 1R until this satellite was successfully launched in 2005.

The current Japanese meteorological satellite network provides a wealth of information, including data on cloud height and distribution, upper-air wind, and sea surface temperature distribution. The observational data received from the spacecraft allows JMA and other national meteorological and hydrological services to continuously monitor significant meteorological phenomena such as typhoons, storm fronts, and low-pressure systems. The data collected by Japan's meteorological satellites are also directly assimilated into the numerical weather prediction system, which in turn contributes to the timely issuance of disaster prevention information and weather forecasts from JMA and related weather agencies.

The Multifunctional Meteorological Satellite (MTSAT-1R) was launched in 2005 after 2 years of delay. This allowed the GOES 9 to be returned to US operation. The MTSAT 1R was capable of performing observations every 30 min with imaging channels consisting of a visible band and four infrared bands. MTSAT-2, launched in 2006, took over many of the imaging functions of MTSAT-1R in 2010 and is now on standby. MTSAT-2 imagery distribution services for L-band frequency High-Rate Information Transmission (HRIT) and Low-Rate Information Transmission (LRIT) via MTSAT-1R was discontinued in December 2015. As a replacement of those services, JMA started the HimawariCast service which disseminates primary sets of imagery via a communication satellite from January 2015. The interval between full-disk observations by Himawari-8 is 10 minutes, which is much shorter than the 30/60 minutes of the MTSAT series. Additional data is obtained by JMA from various polar-orbiting satellites, such as the NOAA and POES series, operated by the USA as well as the MetOps satellite operated by EUMETSAT of Europe.

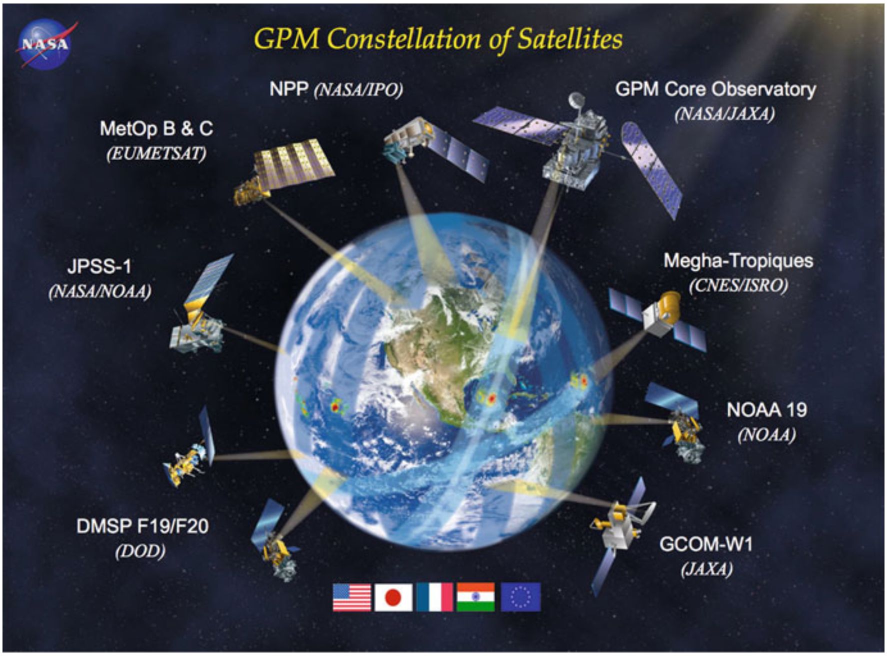


Fig. 4 The GPM Constellation (Graphic courtesy of NASA)

Further, JMA obtains key information from two other "experimental" satellites. From 1995 to 2015, the Tropical Rainfall Measuring Mission (TRMM), a JAXA/NASA project, provided information on tropical and subtropical rainfall. TRMM has shown the importance of taking measurements at different times of day to improve observations of weather systems and real-time monitoring of hurricanes. The Global Precipitation Measurement (GPM) is an international satellite mission to provide next-generation observations of rain and snow worldwide every three hours. NASA and JAXA launched the GPM Core Observatory on February 27, 2014 which extends the observations to higher latitudes, covering the globe from the Antarctic Circle to the Arctic Circle. GPM is composed of one core satellite and approximately eight constellation satellites. Led by JAXA and NASA, the GPM program will be conducted in cooperation with NOAA, CNES, ISRO, China, and the European Union (Fig. 4).

The design of the GPM Core Observatory is an advancement of TRMM's highly successful rain-sensing package, which used an active radar capable of providing information on precipitation particles, layer-by-layer, within clouds, and a passive microwave imager capable of sensing the total precipitation within all cloud layers. Since light rain and falling snow account for a significant fraction of precipitation occurrence in middle and high latitudes, the GPM instruments extend the capabilities of the TRMM sensors to detect falling snow, measure light rain, and provide, for the first time, quantitative estimates of microphysical properties of precipitation particless.

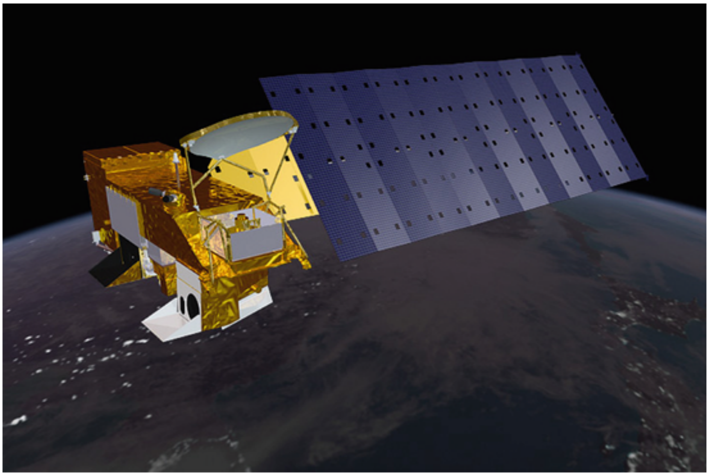


Fig. 5 The AQUA satellite over the Pacific Ocean at night (Graphic courtesy of NASA)

Another NASA research satellite named AQUA also provides detailed oceanographic and tropical rainfall data as well as atmospheric images to JMA utilizing the onboard moderate resolution imaging spectroradiometer(MODIS) (Fig. 5) (Aqua Satellite).

Data from these satellites are indispensable in observing typhoons, monitoring the global and marine environment, and producing initial fields for numerical weather prediction.

Currently, the Himawira 8 and 9 meteorological satellites were manufactured by the Mitsubishi Electric Company with the Boeing Corporation serving as subcontractor. These advanced satellites are now launched with these launches occurring on the Japanese IIA launch vehicle. The Himawira 8 entered service in July 2015 and is expected to be in service from 2015 to 2022. The Himawira 9 is scheduled for launch in 2016 and expected to operate from 2022 to 2029 (JMA/MSC: Himawari-8/9).

The Russian Geostationary Operational Meteorological Satellite (GOMS) and Polar-Orbiting Meteorological (Meteor) Systems

The Russian meteorological satellite network was once one of the most robust in the world but budgetary constraints led to the decrease in the number of satellites in orbit. This has led to the Russian Federation relying on meteorological data from Europe, the USA, and other satellite networks. The Russian government has now strongly committed to restoring the network of weather satellites that existed during the time of Soviet Union. This is a large challenge in that the Russia needs to monitor weather and climate conditions across the country's 11 time zones which is by far the greatest challenge in terms of meteorological forecasting that any nation in world must face.

In the last few years, Russia has concentrated on developing and launching a full array of new meteorological satellites in both geostationary and polar orbit. This program is now underway but will not be fully completed for another decade.

Russia announced in September 2010 that it plans to fully restore its weather satellite network by 2030 under a State-sponsored program. As a first priority, the Russian Federation State has undertaken to deploy a near-polar orbit constellation of Meteor-M satellites, beginning with the Meteor-M No 1 working on an 830-km circular sun-synchronous orbit, in order to bolster meteorological service for the vast expanse of the country and also to deploy and operate the latest Geostationary Operational Meteorological Satellite (GOMS) system. The geostationary satellites are to be comprised of the Elektro-L1 (launched in 2011) and Elektro-L2 (launched in December of 2015).

Elektro-L2 will enable local and global weather forecasting, analysis of oceanic conditions through images acquired in 10 spectral ranges, including three optical (1 km resolution) and seven infrared channels (4 km resolution). In addition, Elektro-L2 carries a suit of instruments for heliophysics, which will monitor critical “space weather” phenomena, such as solar flares, radiation levels and the condition of the Earth’s magnetosphere. The spacecraft will be able to see the entire disk of our planet and transmit resulting images every 30 minutes under most circumstances or every 10–15 minutes when needed urgently, for example, to monitor natural disasters. The data from the satellite will be primarily used by Russia’s civilian weather agency Roshydromet. The spacecraft will also be a part of the WMO international constellation of weather satellites, providing data to users around the world (http://www.russianspaceweb.com/elektro-l2.html. Last accessed: June 24, 2016).

The Meteor-3 Polar-Orbiting Meteorological Satellite System

The polar-orbiting meteorological satellites, Meteor-3 presently operating in Russia, provide possibilities for acquiring data for hydrometeorological and helio/geophysical support as well as global environmental monitoring. The system’s spacecraft are located on near-polar circular orbits (height of approximately 1,200 km, with an inclination of 82.5°). The characteristics of the instruments onboard the Meteor-3 series of satellites are shown in Table 2.

The scientific instrument package onboard the Meteor-3 spacecraft enables regular instant acquisition of images of cloudiness and of the Earth’s surface in visible and infrared bands, data on air temperature and humidity, and sea surface temperature and cloud temperature. Acquired corpuscular and X-ray irradiance and total emitted radiation energy data are used for geophysical studies.

Beyond regular scientific hardware, the Meteor-3 spacecraft are often equipped with experimental and research instruments. The Meteor-3 satellite No.5, which was launched on August 15, 1991, carried the scanning spectrometer for global ozone distribution mapping TOMS instrument, developed by NASA. The TOMS instrument failed to provide operational service after December 27, 1994.

Table 2 Weather satellite Meteor-3 permanent onboard equipment

Instrument	Spectral band, um	Ground resolution, km	Swath width, km	Operating schedule
Scanning TV-sensor with onboard data recording system for global coverage mode	0.5–0.8	0.7 × 1.4	3,100	Recording, direct transmission
Scanning TV-sensor for automatic data transmission mode	0.5–0.8	1 × 2	2,600	Direct transmission
IR-radiometer for global coverage and direct data transmission modes	10.5–12.5	3 × 3	3,100	Recording, direct transmission
Scanning 10-channel IR-radiometer	9.65–18.7	35 × 35	400	Recording, direct transmission
Radiation measuring system	0.17–600 MeV	–	–	Recording, direct transmission
Radiochannel	466.5 MHz – data transmission to control centers			
	137.850 MHz – data transmission to local acquisition stations			

Meteor-3-7, the last of the Meteor 3 series, was launched on January 25, 1994 and had a complement of instruments that included TV camera systems observe daytime Earth cloud cover in the visible spectrum (MR-2000M, MR-900B), an infrared radiometer (Klimat) to produce global photomosaics of the Northern and Southern Hemispheres, a Radiation Measurement Complex (RMK2) to register flux densities of protons in the 5–90 MeV and electrons in the 0.15–3.0 MeV energy regions, a scanning 10-channel IR radiometer for atmospheric thermal sounding (SM) and sensors and instruments from other space agencies through international cooperation. The latter included the instrument ScaRaB, a Scanner for Radiation Budget developed by France, Germany and Russia.

Meteor-3M

A next-generation satellite series, Meteor-3m was conceived in the late 1990's aimed at rebuilding Russia's meteorological satellite infrastructure. The first Meteor-3M1 spacecraft was launched from Baikonur on Dec. 10, 2001; technical issues with the US-built SAGE-III instrument had postponed the launch from December 2000. Meteor-3M1 functioned until March 2006. On July 8, 2014, Russia orbited its latest version of a weather-forecasting and remote-sensing satellite, known as Meteor-M No. 2 (Meteor-M2). The 2,778-kilogram Meteor-M2 was designed to watch global weather, the ozone layer, the ocean surface temperature and ice conditions to facilitate shipping in polar regions and to monitor radiation environment in the near-Earth space. The payload package onboard Meteor-M No. 2 includes:

Multi-channel imaging scanner, MSU-MR
Multi-channel imaging complex, KMSS
Ultra-high frequency temperature and humidity radiometer, MTVZA-GYa
Infrared Fourier spectrometer, IKFS-2
Radar complex, BRLK Severyanin
Heliophysics instrument complex, GGAK-M
Radio relay complex, BRK SSPD

The satellite was designed to operate in orbit for five years. It will become the second spacecraft in the Meteor-3M network, complementing the Meteor-M1 satellite, which was launched on Sept. 17, 2009. In addition, the Russian space program funds the development of the Meteor-M3 satellite, which is designed to carry a phased-array radar for high resolution observations of the ocean surface. Russia's Hydrometeorological Center, with the help of the Russian Federal Space Agency, Roscosmos, plans to deploy a total of six Meteor-M weather satellites operating in a low-earth orbit constellation. These satellites are being launched utilizing the new Soyuz-2 high-performance Soyuz booster. These satellites have service lifetime of 5–7 years. Subsequent generations of these polar orbit satellites will have a longer life of perhaps 12–15 years (Russia to have five weather satellites).

The Meteor-M satellites are being manufactured by the Moscow-based VNIIEM, NPP (Science and Production Enterprise "All-Russian Scientific and Research Institute of Elektro-mechanics") under contract to the Russian Federal Space Agency, Roscosmos. VNIIEM/NPP was also responsible for manufacturing the earlier Meteor 1, Meteor-2, Meteor-3 series of satellites and the Geostationary Operational Meteorological Satellite (GOMS) weather satellite also known as Elektro-1.

The Russian Geostationary Operational Meteorological Satellite

The latest GOMS satellites, the Elektro-L series, have been designed by Roscosmos scientists and engineers in conjunction with climatologists and meteorologists to provide a wide variety of data, including weather analysis and forecasting on a global and regional scale. These satellites will be able to monitor changes in the climate as well as day-to-day weather patterns plus data from the Sun and information on cosmic radiation (Meteorological System).

These satellites are designed to provide synoptic images of the entire visible hemisphere of Earth at a resolution of 1 × 1 km per pixel (in the visible light band) and 4 × 4 km (in the Infrared band) and do so every 30 min. The weight of these spacecraft in operational mode is about 1,500 kg, and their service lifetime is projected to be about 10 years (Meteor-M).

Elektro-L, also known as GOMS-2, was developed by ROSHYDROMET/PLANETA/Roscosmos. The Elektro-L is a successor spacecraft to the Geostationary Operational Meteorological Satellite (GOMS) that is also referred to as Elektro-GOMS. Elektro-GOMS was launched on October 31, 1994 but was never brought to full operational service due to technical problems.

The overall mission objectives of Elektro-L including satellites that are still to be launched are as follows:

- To provide an operational basis multispectral imagery (hydrometeorological data) of the atmosphere (including the cloud-covered sky)
- To provide complete updated images of the Earth's surface within the hemispheric coverage region (visible disk) of the spacecraft
- To provide information on high-energy cosmic radiation
- To collect heliospheric, ionospheric, and magnetospheric data
- To provide the required communication services for the transmission/exchange of all data with the ground segment
- To provide the services of data collection for the data collection platforms (DCPs) in the ground segment as well as to provide the services of the COSPAS/SARSAT program which are to pick up the emergency search and rescue signals of pilots, distressed people at sea, or other isolated travelers in distress (ERS European Remote Sensing Satellite).

The Elektro-L spacecraft have been built by NPO Lavochkin Research and Production Association of Moscow in association with Roshydromet/Plantera/Roscosmos. The spacecraft employs the so-called navigator platform. This is a general purpose bus, which is three-axis stabilized, and provides a pointing accuracy of better than 0.05°. The angular drift is on the order of 5 × 10−4 o/s. A deployable solar array provides a power of 1.7 kW at end of life, while the spacecraft's mean power consumption is estimated to be about 700 W. The total mass of the spacecraft is about 1,620 kg with a payload mass of 435 kg. The Elektro-L spacecraft design lifetime is projected to be 10 years. Figure 6 shows the Elektro-L spacecraft.

The Elektro-L1 spacecraft was successfully launched on January 20, 2011 on a Zenit-2 launch vehicle with a Fregat-SB booster from the Baikonur Cosmodrome, Kazakhstan and began operations in geostationary orbit over the Indian Ocean in March 2011. The Elektro-L2 was succesfully launched in December of 2015.

The Elektro-L spacecraft have several key component subsystems. These are the onboard radio engineering complex (OREC), the multispectral scanning unit – geostationary scanner(MSU-GS), the Helio/geophysical Instrument complex (GGAK-E), the onboard data sampling system (ODSS), and the geostationary search and rescue system (GS&RS). The first three of these components, that are most critical to the meteorological and climatological mission, are briefly described below.

OREC (onboard radio engineering complex). The objectives of the RF communication system are to provide all data transmission, relay, and retransmission services with the ground segment. These RF relay functions include the following:

- The sensor data downlink to the ground acquisition and distribution center is in X-band (7.5 GHz) at a data rate of 2.56–15.36 Mbit/s.
- Data reception from ground segment data collection platforms (DCPs) at 400 MHz (UHF) or DCP data relayed via LEO satellites at a frequency of

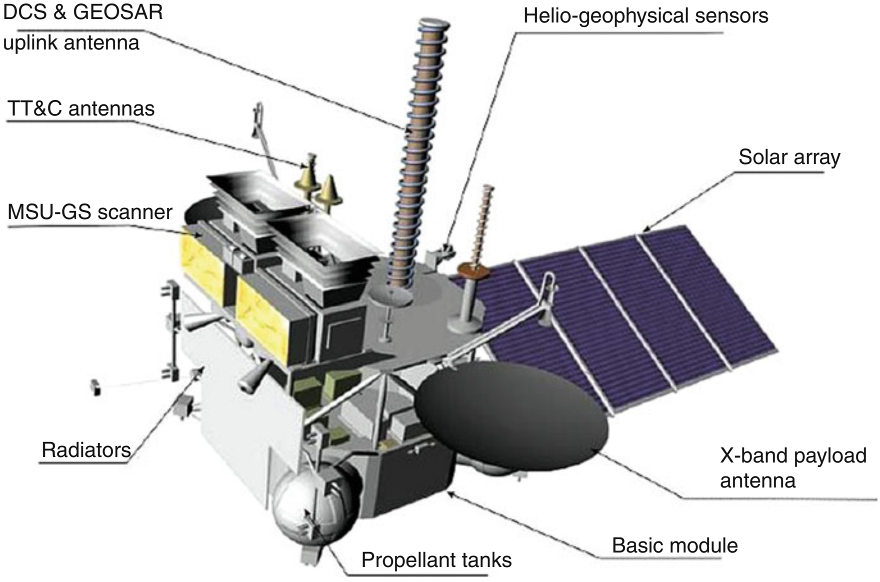


Fig. 6 The Elektro-L spacecraft and some of its components (Image credit: Roshydromet/Planeta)

470 MHz. This data is transmitted from Elektro-L to the ground acquisition and distribution center in S-band at 1.7 GHz.

- Onboard reception of processed hydrometeorological data products in X-band (8.2 GHz) and relay of this data (in S-band at 1.7 GHz) to all customers.
- Exchange of hydrometeorological data and remote sensing data between regional centers in X-band (at 8.2 and 7.5 GHz) with data rates of up to 15.36 Mbit/s.
- Data reception of COSPAS-SARSAT messages at 406 MHz and retransmission of these messages at 1.54 GHz.

The multispectral scanner unit (MSU-GS) is a 10-channel radiometer. The objectives of this radiometer unit are to obtain solar-reflected imagery and brightness temperature measurements from the top of Earth's atmosphere and from the Earth's surface (ocean and land). In addition, the tropospheric moisture content is also determined.

The MSU-GS instrument is a multispectral scanner to take imagery in three visible and seven infrared bands. These measurement bands are closely parallel to the instruments now operating onboard the MeteoSat-8/MSG-1 spacecraft of EUMETSAT.

Helio/Geophysical Spectrometry Instrument Complex GGAK-E

The system is designed to monitor the penetrating radiation's spectra and density in the near-Earth environment and the magnetic field state. The system records the following helio/geophysical information (HGI):

- The density of electron fluxes with energies in four bands from 0.04 to 1.7 MeV
- The density of proton fluxes with energies in four bands from 0.5 to 90.0 MeV
- The density of alpha particles fluxes with energies from 2 to 12.0 MeV
- Intensity of galactic cosmic radiation with energies greater than 600 MeV
- Solar X-ray radiation intensity with energies from 2 to 10 Å°
- Intensity of solar ultraviolet radiation in four wave bands up to 1,300 Å°

These very sophisticated subsystems allow the Russian meteorologist to have access in near real time to a wealth of data. The Russian meteorological network will continue to improve as the Elektro-L series GOMS satellites and the full Meteor-M and advanced Meteor-M satellite constellation are deployed (Russia Launches).

South Korea's Communication, Ocean, and Meteorology Satellite (COMS)

South Korea's first Communication, Ocean, and Meteorological Satellite (COMS-1), dubbed Cheollian, was launched successfully by Arianespace using an Ariane 5 ECA rocket on June 27, 2010 from the Guiana Space Center in Kourou, French Guiana. The COMS-1, or Cheollian, is a multipurpose geostationary satellite capable of performing communication, ocean, and meteorological functions. The COMS satellite has been placed in geostationary orbit at 128° East. Its mission is scheduled to last 7 years; however, the satellite has a design life of 10 years. The COMS satellite is operated by the Korea Aerospace Research Institute (KARI) which serves a quasi-space agency for South Korea.

The COMS satellite has three payloads: one for meteorology, one for ocean observation, and one for communications. COMS will provide meteorology data to end users around the globe and oceanography data for the Korean Peninsula. This multipurpose satellite will also carry out experimental satellite communications services in Ka-band. As prime contractor for COMS, EADS Astrium was responsible for the design and building of the satellite including both the meteorology and ocean imagers. The communications payload was provided by KARI in Korea, as a customer furnished equipment.

COMS provides continuous image monitoring with the extraction of high resolution meteorological data from its multispectral imager. It will give early warning of hazardous weather conditions including storms, floods, sandstorms, etc., and provide data on the long-term changes in sea surface temperatures and cloud patterns. Earth observation data from COMS will be relayed to a processing station. Once processed, the data will be resent via the COMS satellite to weather forecasters and Earth observation centers around the world.

COMS will carry an Ocean Imager to monitor marine environments around the Korean Peninsula and provide data (on chlorophyll, etc.) to assist the fishing industry in the region. It will also monitor both long- and short-term changes to the marine ecosystem.

The communications payload onboard COMS will allow “in-orbit verification” of advanced Ka-band communication technologies and will support experiments covering wide-band multimedia communication services.

COMS carries the following payloads:

- Meteorological Imager (MI): The Imager is a multispectral channel two-axis scanning radiometer and is capable of providing imagery and radiometric information of the Earth’s surface and cloud cover over five channels – one visible channel (of 1 km ground resolution) and four Infrared channels (of 4 km resolution).
- Geostationary Ocean Color Imager (GOCI): The advanced ocean imager has a sophisticated focal plane providing for ocean data acquisition from geostationary orbit. The ocean imager will provide data over eight imaging bands in the visible spectrum. Ground resolution over Korea is 500 m.
- Meteorology data dissemination function: Using an S-band-receiving antenna and an L-band-transmitting antenna, this function will allow dissemination in HRIT and LRIT format of weather data.
- Ka-band Communications Payload (COPS): The Ka-band payload will provide three regional beams simultaneously. The Ka-band payload will provide the beam switching function for high-speed multimedia services including the Internet via satellite in the public communications network for all coverage.

COMS is KARI’s first geostationary satellite and will provide Korea with its own meteorology and ocean data, thus giving increased independence. COMS is part of a 15-year Korean space plan begun in the 1990s and followed systematically ever since (Introduction to COMS).

After a period of early operation, satellite communication and meteorological/ ocean data services will be offered for public use. According to the national long-term plan for space development, a second geostationary multifunction satellite will be launched sometime after 2014.

Conclusion

The in-orbit global meteorological satellite resources represented by the USA, Europe, China, India, Japan, Russia, and South Korea are today quite considerable, and in the coming years, these satellite systems will continue to grow in scope and capability. The addition of new sensor capabilities to monitor heliographic and cosmic radiation and the deployment of more satellites in various relevant orbits will help to chart various elements of climate change, the melting of the arctic ice caps, the dimensions of the holes in the ozone layer, and increasing temperatures on land and in the ocean around the world. These expanded space facilities will play a critical role in providing better and longer-term weather forecasts but also in developing new strategies to adapt to the many effects of climate change itself.

The collaborative efforts that come from the United Nations World Meteorological Organization (WMO), the World Weather Watch (WWW), the United Nations Environment Programme (UNEP), and similar international and regional organizations can be of enormous value. Already most meteorological satellite data from around the world is freely shared in the common cause of dealing with the effects of tropical storms, monsoons, typhoons, hurricanes, and tornadoes. As new tools such as hyper-spectral sensing, lightning strike monitoring, and cosmic radiation monitoring, the opportunities for even more international collaboration will continue to evolve. This progress will come in many ways such as improved instrumentation, more sophisticated data analysis and formatting, and better ways to monitor not only weather patterns but longer-term trends in climate change.

Today all of the countries involved in satellite meteorology as discussed in this and previous chapters can design, build, and launch their own satellites – a significant change from early in the space age.

In the future, additional countries will deploy sophisticated meteorological satellites which will thereby enrich tomorrow's space capabilities. In 20 years, longer-term weather forecasts and much more sophisticated sensing of climate change will surely emerge from all of today's efforts to strengthen international cooperation in this vital area.

Cross-References

- ▶ Ground Systems for Satellite Application Systems for Navigation, Remote Sensing, and Meteorology
- ▶ Introduction to Space Systems for Meteorology
- ▶ United States Meteorological Satellite Program

References

Aqua Satellite by Nasa (2016), http://aqua.nasa.gov/Last. Accessed 31 Mar, 2016

Elektro-L – RussianSpaceWeb.com (2011), www.russianspaceweb.com/elektro.html. Last Accessed 31 Mar 2016

ERS – European Remote Sensing satellites (2006), http://www.geoportal.org/geonetwork/srv/en/metadata.show?id=604. Last accessed 31 Mar 2016

Fenyun 3, *2nd Generation Polar-Orbiting Meteorological Satellite Series* (2010), https://directory.eoportal.org/get_announce.php?an_id=12759. Last accessed 32 Mar 2016

Fenyun Data Center (translated from Chinese) (2016), http://satellite.cma.gov.cn/portalsite/default.aspx. Last accessed 31 Mar 2016

Global Observing System (GOS), World Meteorological Organization (2016), https://www.wmo.int/pages/prog/www/OSY/GOS.html. Last accessed 31 Mar 2016

Notes

Introduction of COMS (Communications, Ocean and Meteorological Satellite) of South Korea (2010), http://nmsc.kma.go.kr/html/homepage/en/chollian/choll_info.do. Last accessed 31 Mar 2016

JMA/MSC: Himawari-8/9 http://www.jma.go.jp/jma/jma-eng/satellite/. Last accessed June 24, 2016

Listing of ISRO Satellites (2015), http://www.isro.org/satellites/allsatellites.aspx. Last accessed Dec 2015

Meteor-M1 – GlobalSecurity.org (2011), http://www.globalsecurity.org/space/world/russia/meteor-m1.htm. Last accessed 31 Mar 2016

Meteorological System with the Geostationary Operational Meteorological Satellite. ELECTRO GOMS (1994), http://sputnik.infospace.ru/goms/engl/goms_1.htm. Last accessed 31 Mar 2016

National Satellite Meteorological Center of CMA FENGYUN, www.nsmc.cma.gov.cn/NSMC_EN/Channels/100090.html

Russia launches new weather watcher (2014), http://www.russianspaceweb.com/meteor_m2.html. Last accessed 31 Mar 2016

Russia to have five weather satellites by 2013, Moscow, RIA Novosti, 20 Oct 2008, pp. 1–2, http://wwwspacedaily.com and http://www.globalsecurity.org/space/world/russia/meteor-m1.htm. Last accessed 31 Mar 2016

The Mission of the Japanese Meteorological Agency (2016), http://www.jma.go.jp/jma/en/Background/mission.html. Last accessed 31 Mar 2016

World Weather Watch, World Meteorological Organization (2015), http://www.wmo.int/pages/prog/www/index_en.html. Last accessed 31 Mar 2016

Part V

On-Orbit Robotic Servicing, Hosted Payloads and Active Debris Removal

Innovations in Hosted Payload Satellite Services

Joseph N. Pelton and Scott Madry

Contents

Abstract

One of the important new developments in commercial and governmental satellite systems is the active deployment of hosted payloads. The prime reason for the use of hosted payloads is to save costs and avoid the expense of a more costly dedicated mission. The hosted payload approach may involve the deployment of experimental packages that are typically only a one-of-a-kind project, or it can involve many operational packages that are "piggybacked" on a large low earth

---

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

S. Madry
Global Space Institute, Chapel Hill, NC, USA
e-mail: madrys@email.unc.edu

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_103

orbit constellation with many satellites so equipped. This is an approach that has been particularly promoted within US space programs in response to the US 2010 official space policy. This White House policy emphasized the use of hosted payloads, where cost savings and operational efficiency so allowed. This approach to the use of hosted payload is also being employed around the world by many different entities for a variety of purposes.

Examples provided here include the IRIS experiment that was included on an Intelsat satellite, the Anik G1 with an X-band package, the WAAS package that flew on the Galaxy 15 satellite and the UHF package that is flying on the Intelsat 22 satellite. Another example of the specialized package flying as a hosted payload is the case of those experiments that are currently flying on the large Inmarsat Alphasat. The above examples typically involve very specific individual hosted payload packages.

There can be much different type programs where the hosted payload approach involves the deployment of a small operational package on each of a number of satellites within a large-scale satellite network. In this case the example provided is with regard to the Aireon packages that are being deployed with the Iridium NEXT Satellite System.

A decade ago, the hosted payload approach was a very occasional and unusual approach and most often involved a one-of-a-kind experimental package, but today "hosted payloads" have become a much more common practice with large companies such as Intelsat General and SES even having dedicated units that focus exclusively on hosted payload activities. Annual conferences on the topic of hosted payload now draw many hundreds of attendees. This growing interest in hosted payload flying on satellite networks has also led to the formation of the Hosted Payload Alliance with a quite large and growing global membership. In short, hosted payload activities in the course of the past decade have become a big business involving a large number of satellites and significant spacecraft and ground system investment.

This chapter addresses the various types of hosted payload activities that are now in progress or planned and provides some analysis of the reasoning behind various hosted payloads and the pros and cons of such undertakings. This analysis considers not only the impact on capital investment, speed of implementation, launch costs, operational costs, and advantages and risks that are associated with various host payload projects that have become a part of the application satellite industry. In many instances the use of hosted payload strategies has been employed in governmental, military, and commercial programs to test new capabilities. Also governmental and military programs have flown on commercial satellite systems.

Somewhat akin to the concept of hosted payloads is the concept of incremental or supplemental payloads that are secondary or even tertiary payloads that are launched as add-on to primary launch operations as part of a single launch deployment into outer space. This "piggybacked" launch operation can lead to cost savings, but this proliferation of smaller satellites in orbit can add to the growing problem of orbital debris.

Consolidation of smaller payloads such as student experimental packages by placing them on a larger satellite as hosted payload or flying them to the

International Space Station and returning them after the experiment is finished is now a common practice. This use of NanoRacks type experiments that fly on the International Space Station in particular can be highly cost effective, allows astronaut oversight of experiments, and eliminates orbital debris issues.

Keywords
Aireon package • Alphasat • Anik G • Ariane launch vehicles • Arianespace • Geostationary Satellite Communications Control Segment (GCCS) • Harris Corporation • Hosted payloads • Hosted Payload Alliance • Inmarsat • Intelsat • IRIS (Internet Routing in Space) • Iridium generation next • Marisat • NanoRacks • National Aeronautics and Space Administration (NASA) • SES • Soyuz launch vehicle • TEMPO project • Vega launch vehicle • WAAS package

Introduction

The development of fixed-satellite systems (FSS) developed very quickly from its earliest days. Each generation of these satellites more than doubled in capacity during the first decade of development from 1965 to 1975. The advent of maritime mobile satellite services in the mid-1970s, however, led to two new approaches to deployment of space segment. Three Intelsat V satellites were equipment with maritime mobile subsystems to demonstrate the feasibility of hybrid systems. The Marisat satellite was built under contract to the US Navy, but half of the system capability was left for the Comsat General Corporation to sell maritime mobile services to commercial customers. The success of the Marisat project that involved an active partnership between the US Navy and Comsat General led to other joint military and commercial projects where military hosted payloads have flown on commercial satellites.

These two early examples, namely, the maritime packages on Intelsat satellites and the Marisat program, demonstrated the concept that more than one type of service or operating system might ride on a single satellite. Although dual operational missions on one satellite – if not even more – can add complications and perhaps additional risks, it can also provide economies in launch, power systems, and operational costs.

When the Ariane 1, 2, and 3 rockets were developed, the so-called SYLDA allowed for dual launches of payloads. When the Ariane 4 launcher was first developed in 1988, a special new unit was developed that allowed a number of additional small payloads to be launched in addition to the primary payload. This was known as the Ariane Structure for Auxiliary Payloads (ASAP). On flight 35 this configuration was used to launch not only the Spot 2 satellite but six other auxiliary satellites that were of the 200 kg class or smaller. Today there are many options available to launch small satellites effectively and at low costs (Mowry and Chartoire).

The two key trends in satellite applications over the years have been to build and launch larger and more capable satellites and more recently alternative approaches of designing and building simpler and less costly satellites that are launched in constellations. The objective, however, has been the same and that is to deploy more cost-effective satellite networks that are more capable and responsive to new market demand. This trend has been supported by flexible and more cost-effective launch arrangements. This evolution of the satellite application market has now given rise to the new option of putting hosted payloads on large-scale geosynchronous (GEO) satellites as well as placing smaller hosted payloads on low earth orbit constellations.

This chapter, in the sections that follow, will discuss examples of hosted payloads on GEO satellites, hosted payloads on non-GEO satellites, the Hosted Payload Alliance, regulatory issues presented by hosted payloads, and launch options associated with small satellites versus putting small packages on large satellites as hosted payloads.

GEO-Based Satellites with Hosted Payloads

IRIS (Internet Routing in Space)

In 2006, the US Strategic Command (STRATCOM) urged the commercial satellite industry to create and fly a geostationary payload that could generate and process signals in space and also be optimized as an Internet router. CISCO, one of the leading developers of Internet routers, responded by developing by 2009 what was named the IRIS space payload with IRIS standing for Internet Routing in Space. CISCO working in partnership with IGC developed the IRIS package and an arrangement was negotiated with Intelsat to launch the IRIS as a hosted payload on the Intelsat 14 satellite. Thus the first Internet router with signal processing capability was launched in November 2009. Space Systems Loral, the manufacturer of the spacecraft, integrated the hosted payload on to the satellite before launch. During in-orbit tests, IRIS allowed the US Strategic Command (STRATCOM) to integrate terrestrial and space communication nodes by means of a common network layer protocol developed for this purpose. This was one of the defined objectives of the US military's so-called "netcentric warfare" strategy.

Satellites with onboard processing and regenerative capabilities have been successfully deployed in recent years, but an IP router in space with dynamic response capabilities was for the first time achieved with the IRIS payload. During the tests integrated net-based communications among various nodes operated by the US Army, the Air Force, the Marines, the Navy, and the Coast Guard were able to operate using the TCP/IP protocol using a common open interface standard.

Telesat Canada and the X-Band Hosted Payload on the ANIK G1 Satellite

Anik G1, built by SSL for Telesat Canada, is a fixed-satellite services (FSS) multimission C-band and Ku-band GEO spacecraft designed to support a variety of applications, including direct-to-home television broadcasting in Canada and broadband, voice, data, and video services in South America.

Anik G1 carries an X-band government communication payload with coverage over the Americas and the Pacific. This is the first commercial satellite to ever have substantial government X-band coverage over the Pacific, making it ideal for

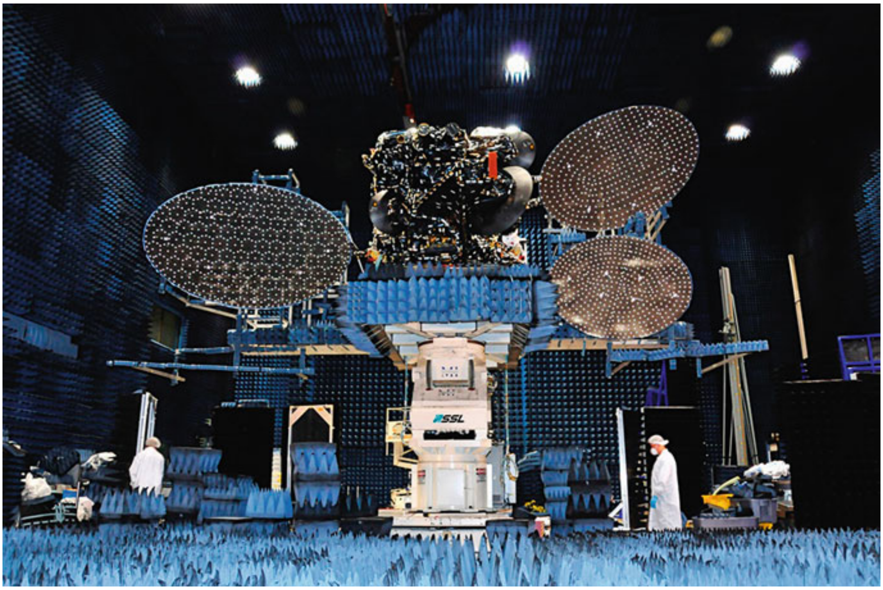


Fig. 1 Anik G1 communication satellite with X-band military band hosted payload (Graphic provided by Telesat Canada)

naval platforms. The payload, which is compatible with NATO standards, is leased to Astrium Services and supports various government applications such as maritime operations, integrated transit, and deployment operations (ANIK G1) (See Fig. 1).

SSL was responsible for all aspects of the GEO spacecraft design, integration, and test as well as the hosted payload mission integration. SSL provided launch site support, as well as support of orbit raising and in-orbit testing (IOT) of Anik G1 before handing off satellite operations to Telesat. SSL continues to provide on-orbit customer support for the duration of the 15-year mission.

Anik G1 was completed, accepted by the customer, and delivered on schedule, whereupon the satellite was placed into storage until the launch date was confirmed. Anik G1 was successfully launched in April 2013 and is currently operational on orbit and is providing coverage of the US Canada, Alaska, and the Pacific Ocean (ANIK G1).

Governmental Hosted Payloads on GEO Satellites

The TEMPO Hosted Payload Project by NASA

The NASA TEMPO project, which stands for Tropospheric Emissions: Monitoring of Pollution, is unusual in several regards: (i) it is a US governmental research project with a capped budget of $90 million; (ii) the TEMPO payload is being deliberately designed as a hosted payload to be launched on a commercial GEO satellite; (iii) the design, engineering, and manufacture of the TEMPO payload are being accomplished by a sophisticated partnership of NASA employees, academic professionals drawn from various universities and research institutes, and commercial aerospace companies; and (iv) the GEO satellite that will host the payload has not yet been selected even though the TEMPO payload is to be finished in 2017 and the launch is planned for 2018.

The team for TEMPO that is managed from NASA Langley has won the right to implement this project on a competitive Earth Venture Instrument mission, out of 14 proposals. It is intended to the first space-based instrument designed to monitor major air pollutants across the North American continent on a real-time basis during daytime hours. This TEMPO project is part of the Earth System Science Pathfinder program and if successful could lead to other governmental payloads be designed to be hosted on other commercial satellites.

TEMPO project manager, Wendy Pennington, who is based at NASA Langley has said: “Many of us in NASA have been talking about using commercial geostationary space for climate research for a long time and now we have an opportunity to do so. It is a very exciting time for us.” If this NASA experimental package accomplishes its proposed objectives, it will, for the first time, make accurate observations of tropospheric pollution concentrations of ozone, nitrogen dioxide, sulfur dioxide, formaldehyde, and aerosols with high resolution and high frequency during daytime hours for the entire US, Canada, and Mexico land masses (NASA Tempo Project).

GCCS-WAAS Package to Augment GPS System on Telesat and Intelsat GEO Satellites

The Wide Area Augmentation System (WAAS) on the Galaxy 15 satellite operates with a system of dozens of ground stations in various ground locations in North America in order to provide necessary augmentations to the standard GPS positioning navigation signal. A network of precisely surveyed ground WAAS reference stations is strategically positioned in continental locations in the United States as well as at selected sites in Alaska, Hawaii, and Puerto Rico. These sites thus collect GPS satellite data and thus make aircraft in very near to “real time” aware of their location with even greater accuracy. Using this information, a WAAS message is developed at the master station to correct signal errors. These correction messages are then broadcast from ground uplink stations through commercial GEO communication satellites to link to receivers onboard aircraft using the same frequency as GPS.

In short the WAAS as discussed in greater detail in later chapters is specifically designed to provide “additional accuracy, availability, and integrity” which US FAA officially considers to be necessary in order to enable airline company users to rely on GPS for all phases of flight, from en route through the GNSS Landing System (GLS) approach for all qualified airports within the WAAS coverage area. With the Wide Area Augmentation System (WAAS), accuracies of 1–2 m in horizontal and 2–3 m in vertical directions are consistently achieved (FAA 2008).

This key satellite package is able to link to many dozens of sites and provide so-called “ground truth” information as to the exact location of each of the WAAS reference sites and provide this precise information to all aircraft in the skies.

In order to adequately cover the United States and to provide on-orbit redundancy in case of satellite failure, it is desirable to have at least two operational WAAS space payloads on separate platforms. Lockheed Martin Transportation and Security Solutions was placed under contract by the FAA to be the prime contractor for providing GCCS-WAAS services. The payload provides users with satellite-based augmentation signal (SBAS) navigation waveforms at the GPS L1 and L5 frequencies. The navigational payload operates in bent-pipe mode and simultaneously translates two C-band uplink signals into two L-band downlink channels.

In 2003, Lockheed Martin contracted with Intelsat (then PanAmSat) and Telesat for hosting of L1/L5 GCCS-WAAS navigation payloads on the Galaxy 15 (PRN 125 at 133W) and on the Anik-F1R (PRN 138 at 107.3W), respectively. Lockheed Martin, which was at the time the owner of the payload, was responsible for FCC frequency licenses. This was the basis of the subsequent ITU coordination. These space payloads thus now provide the FAA’s Geostationary Satellite Communications Control Segment (GCCS) services as specified under contract with Lockheed Martin for WAAS geostationary satellite leased services.

The Lockheed Martin contract with Intelsat includes two elements. The first element is hosting of a redundant L-band WAAS transponder system on Galaxy 15 (including integration, testing, program oversight, etc.) followed by ten years of operations from service commencement. The second element was placement of a ground uplink station (GUS) located at Intelsat’s Napa, California teleport. Lockheed Martin also owns the hardware installed at the ground control site. Separate lease contracts and ground station arrangements are in place with Telesat Canada that operated the Anik F1 satellite (FAA 2008).

A simplified graphic (See Fig. 2) shows how the package on GEO satellites such as the Galaxy 15 or Anik F1 connect the WAAS reference stations to in-flight aircraft.

Hosted Payloads on LEO Constellations

The largest commercial arrangement for carrying hosted payload involves the Iridium NEXT constellation that is currently being deployed. These hosted payloads are to support the Aireon electronic service to support air traffic management and route control. Aireon is a joint venture between Iridium and Nav Canada. This service is designed to provide commercial airlines and other customers with exact position-location data. The concept is to provide airliners with additional security but also achieve a reduction in fuel costs since this data will allow airlines to fly more precise and fuel-saving routes. The designer and manufacturer of the antennas and electronics for these hosted payloads is the Harris Corporation (See Fig. 2).

The data will be relayed by Automatic Dependent Surveillance Broadcast or ADS-B antennas plugged into the Harris-supplied AppStar payload box. Instead of

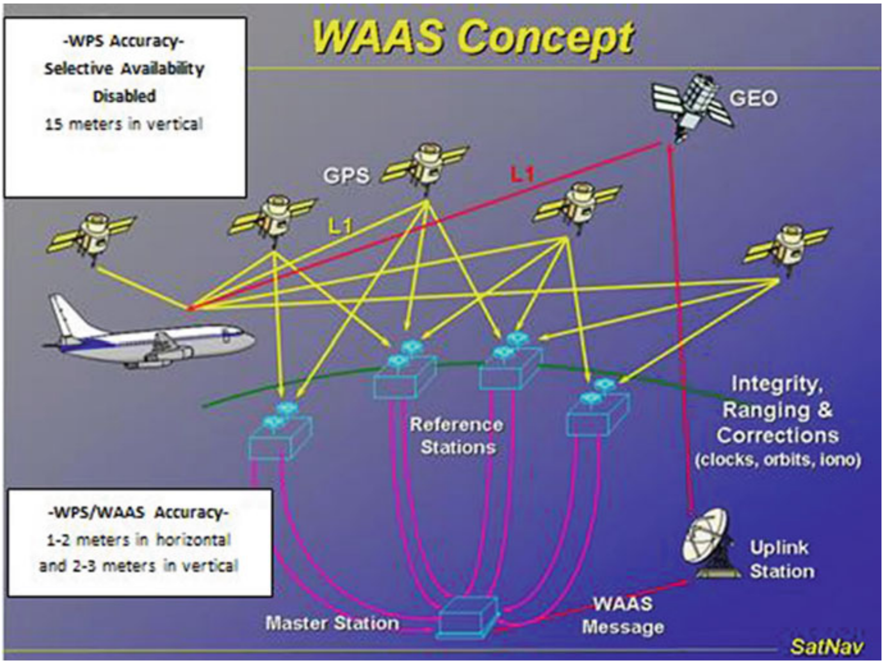


Fig. 2 The WAAS concept: showing how reference stations and GEO satellite packages relay information to aircraft (Graphic courtesy of FAA)

using radar to establish an aircraft's position, aircraft can now be equipped with an ADB-B antenna. This will allow an aircraft to determine its position using GPS. This allows for the location of the plane and other information to be transmitted to a network of ground stations. This change in route determination is a key part of the FAA's NextGen – the Next Generation Air Transportation System, which is scheduled to be in full operation by 2020. Other air traffic control entities around the world are also shifting over to similar systems.

Iridium has estimated that this hosted payload system for traffic management and control could add on the order of $45 million annually in additional revenues. These revenues are, however, contingent on actual sales by the Harris Corporation of this capability, and thus as far as Iridium NEXT is concerned, it is a speculative and essentially entrepreneurial activity.

The Harris Corporation has not only designed the Aireon hosted payload packages but is responsible for marketing these packages for air traffic management and control but also for other purposes. Ultimately the actual revenues will depend on what types of arrangements Harris is able to negotiate with client airlines and other customers. The final determination of the amount of hosting fees, to be paid by Harris to Iridium over period through 2021, will not be finalized until the entire Iridium NEXT constellation is in orbit, which currently is slated to happen in 2018. Deployment of this system has been slowed due to the launch failure of the Falcon

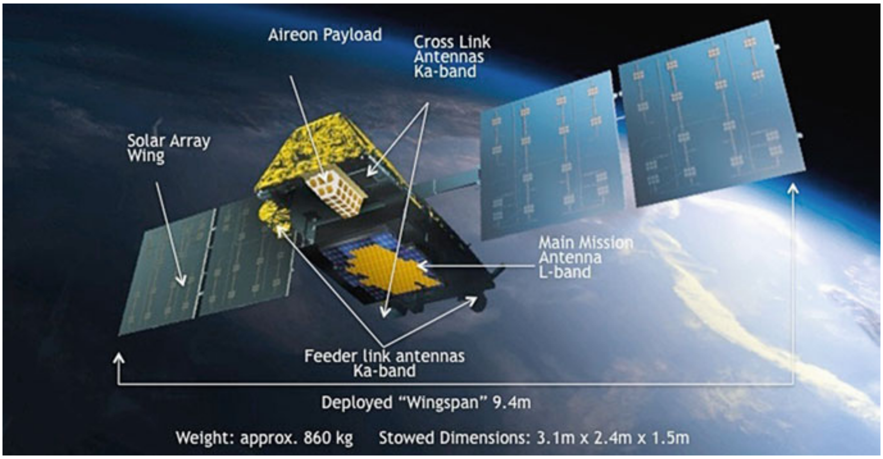


Fig. 3 The one of the Iridium NEXT satellites showing location of Aireon hosted payload (Graphic courtesy of Iridium)

9 launch vehicle and the many scheduled flights on this now heavily used launcher system.

Iridium is paying Harris $114.7 million to design and install the electronic boxes for the Aireon venture. These systems, however, are expected to generate up to $200 million in hosting fees during the first four to five years of operation. The boxes are based on Harris' AppStar reconfigurable payload (Ferster 2013) (See Fig. 3).

The AppStar box as designed by Harris is optimized to accommodate a variety of different payloads. The unique idea is to create an integrated system where you can slide a card into the AppStar chassis to update the software so that the actual physical augmentations are minimal and thus the costs of collecting data from multiple small hosted payload are greatly reduced.

Thus while the Aireon packages are the prime hosted payloads for the Iridium NEXT constellation, other smaller packages such as radiation monitors and dosimeters can easily and cost effectively be included. In this case the customer would be the US Air Force, and the prime purpose would be able to distinguish between satellite-jamming signals and natural space radiation.

Both Iridium and Harris contend that all of the services that the hosted payload capabilities included in the Aireon package will allow the aircraft traffic management and control, radiation monitoring, or other services to be provided at a fraction of the cost that would have been incurred if a dedicated system had been deployed separately for this purpose (Ferster 2013) (See Fig. 3).

What is of particular concern to all of those that agree to rely on hosted payloads for various types of services is what happens if satellites are lost or power impairments restrict or cut off the ability to provide these add-on services.

The large-scale nature of the Iridium NEXT system provides a good deal of redundancy and thus insurance against failures or reliability problems that lead to loss of performance capabilities. This large-scale deployment means that if one or even a few satellites are lost, this does not in a significant way impair the entire global service. Further, system spares will presumably be able to restore complete global service without extended delay. In the particular case of air traffic management and control, reliability and continuity of service are matters of some significant concern since this is such a vital service that could impact the safety of passengers on airliners.

Launch Arrangements for Multiple Deployment of Small Satellites

The obvious alternative to flying hosted payloads on low earth orbit constellations is finding lower cost and reliable ways to deploy spacecraft packages than have previously been the case. Thus many of the launch operators while moving to accommodate large-scale spacecraft such as the Ariane 6 have also been seeking ways to accommodate those wishing to deploy constellations or at least multiple satellites in the 50 kg to 200 kg class of satellites.

This has become a significant issue since a number of entities such as One Web and Space X have recently indicated plans to deploy hundreds if not thousands of small satellites of this approximate size. These systems are currently envisioned as free-flying spacecraft rather than being designed as hosted payloads. Currently the largest hosted payload system is the Aerion payload on Iridium NEXT, but if this particular deployment is successful, even larger networks using host payloads can be expected in the future. The importance of this growing new market potential is certainly being considered by launch providers.

The graphic below shows the proposed configuration that has been designed to launch three supplemental payloads of 200 kg small satellites on the Soyuz launcher as well as how five 200 kg payloads could be launched on a Vega (See Fig. 4). The question at this time is whether the predominant path forward will be the separate

Fig. 4 Soyuz and Vega launcher cross sections showing how three small (200 kg) satellites plus a 3000 kg main mission satellite could be launched on Soyuz Launcher or 5–200 Kg satellites could be launch on a Vega launcher (Graphic provided by the Ariane Guyana Launch Center)

launch of these small payloads in the 50 kg–200 kg class as independent flyers, or will the prime option be a predominant pattern of hosted payloads. Currently the answer seems to be a combination of both options. This is not only an area of major concern and interest to launcher organizations, but it is certainly a concern for those addressing the issue of space orbital debris. Hosted payload, as part of larger spacecraft are today, much more likely to be deorbited or successfully placed in a safe parking orbit after end of life.

Hosted Payload Alliance

The Hosted Payload Alliance was formed in 2011 in order to increase awareness of the benefits of hosted government payloads on commercial satellites with membership open to satellite operators, satellite manufacturers, system integrators, and other interested parties. The formal goals of the organization include the following:

Goals

1. Serve as a bridge between government and private industry to foster open communication between potential users and providers of hosted payload capabilities
2. Build awareness of the benefits to be realized from hosted payloads on commercial satellites
3. Provide a forum for discussions, ranging from policy to specific missions, related to acquisition and operation of hosted payloads
4. Act as a source of subject-matter expertise to educate stakeholders in industry and government

The Hosted Payload Alliance has suggested that there are at least five key reasons why a hosted payload approach might be taken. Although cost savings were behind the US Government 2010 directive to examine hosted payload options, the Alliance has also emphasized shorter times to orbit, more reliable design, more options to access favorable orbital locations, and more operational options on the ground and in space (Hosted Payload Alliance).

Shorter time to space. Because the development of an entire satellite system is not required, a hosted payload on a commercial satellite can reach space in a fraction of the time that it would take to develop a free flyer program. Roughly 20 commercial satellites are launched to GEO orbit each year, and each one presents an opportunity to add on additional capability.

- **Lower cost.** Placing a hosted payload on a commercial satellite costs a fraction of the amount of building, launching, and operating an entire satellite. Cost reductions can result from shared integration, launch, and operations with the host satellite.

- **A more resilient architecture.** Hosted payloads enable a more resilient space architecture by distributing assets over multiple platforms and locations. Rather than creating a single platform with multiple capabilities that could be a target for adversaries, spreading capabilities over multiple locations has the potential to contribute to a more resilient space architecture.
- **Increased access to space.** Roughly 20 commercial launches each year provides multiple opportunities for access to multiple orbit locations during the year.
- **Operational options.** Hosted payloads have multiple options to use existing satellite operation facilities with shared command and control of the hosted payload through the host satellite or a completely dedicated and separate system operated by the hosted payload owner (Hosted Payload Alliance).

Beyond these factors, one might also note that hosted payloads – starting with the Marisat program – can also be a useful mechanism for sharing of on-orbit capability between multiple users and between commercial and governmental user in particular. On the other side of the coin, there can be problems with regard to shared spacecraft facilities in the case of launch failures, power or component failures, and other breakdowns that lead conflicting priorities among partners in joint programs.

Regulatory and Frequency Allocation and Coordination Issues

The advent of hosted payloads can lead to several complications in the area of satellite regulation and especially with regard to meeting national and international regulations with regard to radio frequency coordination. As each new satellite is first envisioned, there is typically a filing process within a country regulatory process as to the radio frequencies that are to be used. If these filings are approved at the national level, the entity that serves as the official “national administration to the International Telecommunication Union (ITU)” results in a filing to the Radio Regulation Board (RRB) of the ITU. This allows this proposed use of these frequencies for space communications to be posted so that any other ITU administration can indicate that there might be interference and this problem of potential frequency interference could be coordinated. If a satellite and its proposed frequency use are sent to the ITU without the hosted payload’s frequencies included, this can lead to the process having to be conducted twice. Further in the case of hosted payloads that are included within a large number of satellites in a low earth orbit constellation, the number of organizations that have other LEO constellations as well as a significant number of satellites in GEO orbit could potentially be affected.

The timely filing of information with the national administration that makes the official input to the ITU concerning any hosted payload is thus quite important. Last minute additions to a satellite of one or more hosted payloads could not only complicate the technical and operational design of a satellite but also could trigger an entire new round of frequency coordination activities that are time consuming and potentially expensive to conduct.

In addition to frequency coordination, there are other issues involving the national regulation of spacecraft launches. There can be national regulations requiring a review process and due diligence as to the inclusion of toxic gases for station-keeping, review with regard to possible orbital debris that could be associated with any launch, and other regulatory requirements. Thus it is prudent and economically and technically efficient for any hosted payload that is associated with any launch to be identified well and advance and all required information that is associated with a hosted payload to be identified at the time that a filing is made with the national regulatory agency. This would avoid the need for supplementary information to be posted with the ITU and especially avoid the need for additional frequency coordination meetings with other nations should the hosted payload and its radio frequency usage require frequency coordination.

Conclusion

The rise in the interest and actual deployment of hosted payloads for commercial, governmental, and military programs is one of the more interesting and important developments in the commercial and governmental spacecraft world in the past decade. There have been a number of drivers that have motivated this trend. The pursuit of cost savings has undoubtedly been a prime motivator behind many of these projects.

Today there are more and more launches into earth orbit, more competition for radio frequency assignments and orbital locations, more and more concern about orbital debris and interference between and among satellites, as well as increased competition to deliver more cost effective capacity to orbit. All of these factors could serve to motivate decision makers to pursue hosted payload options for a wide range of different types of programs.

The range of opportunities associated with hosted payloads now includes at least the following possible option involving commercial, governmental, or military satellites and supplementary packages:

- “One-off” experiments to test new technology or demonstrate new system capability that require only a modest amount of power, reasonably small antenna, and essentially a small amount of mass and volume to test new capabilities.
- Operational capabilities such as the GCCS-WAAS relay service that requires only a very few small packages to complete a mission.
- Military programs that opt to pursue a dual-use approach to future needs. This would include programs where military system requirements are defined with the ability for there to be commercial capacity to be added to the mission. This allows the commercial capacity to be sold for supplementary revenues (Thales and UK Skynet).
- LEO constellations that can accommodate smaller packages that are designed to perform another function. This could be a LEO constellation that adds a commercial package for another service such as is the case with the Aireon package

that is added to the Iridium NEXT LEO network, or it could be a commercial network that carries a governmental or military package for defined purposes or even a governmental satellite with a supplemental package (or packages) to meet some additional service need.

Cross-References

- ▶ Economics and Financing of Communications Satellites
- ▶ Overview of the Spacecraft Bus

References

Anik G1 X-Band Military. http://www.hostedpayloadalliance.org/getattachment/5488cec6-9c60-4095-95b7-01edc97c282e/Anik-G1-X-Band-Military-Payload.aspx. Last referenced 14 Jan 2016

FAA, Air Traffic Bulletin: What Air Traffic Controllers Need to Know about WAAS June, 2008. http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/gnss/library/documents/#q2. Last referenced 14 Jan 2016

Hosted Payload Alliance, Benefits of hosted payloads. http://www.hostedpayloadalliance.org/Hosted-Payloads/Benefits.aspx#.VhF3zM5dHcs. Last referenced 14 Jan 2016

C. Mowry, S. Chartoire, Experience launching Smallsats with Soyuz & Vega from the Guiana Space Center. http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=2947&context=smallsat. Last referenced 14 Jan 2016

NASA TEMPO Project, http://www.nasa.gov/langley/TEMPO/. Last accessed 2 Oct 2015

Warren Ferster, “Harris Corporation to Market Iridium Next Hosted Payload Capacity” Space News April 29, 2013. http://spacenews.com/35106harris-corp-to-market-iridium-next-hosted-payload-capacity/#sthash.Xavv03Fr.dpuf. Last referenced 14 Jan 2016

On-Orbit Servicing and Retrofitting

Joseph N. Pelton

Contents

Abstract

The first half century of satellite applications has entailed rapid technological growth with the deployment of bigger, better, and more sophisticated satellites that have responded to rapidly expanding space application markets. The development of more reliable and higher-capacity rockets with increased lift and higher

---

J.N. Pelton (✉)
International Space University Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_104

cost efficiencies has generally reinforced the trend to always seek new economies of scope and scale.

Currently, however, there are widely diverging thoughts about "What Next?" Some feel that even larger high-throughput satellites and corporate consolidation and mergers are the way forward. Others are promoting growth through large-scale low earth orbit constellations with networks that might contain as many as thousands of application satellites in so-called mega-LEO systems. Others are increasingly concerned about orbital debris and the need for active debris removal and improved deorbiting systems. Yet others are being to think that on-orbit repair and servicing of application satellites to extend their usable life may represent yet another important new development.

New techniques associated with on-orbit servicing and repair have begun to emerge in the last few years. There have been many proposed new ways forward. These proposals include refueling of satellites with depleted maneuvering systems, redeployment of satellites from low earth orbits that failed to reach GEO, and even repurposing of components on derelict satellites such as large aperture antennas or solar power systems to create new and cost-effective satellites in space rather than deorbiting them as space debris. These redeployments, repair, or augmentation of defective satellites, and even repurposing of parts from derelict satellites to create new spacecraft, could offer new economies of scale to make satellite applications more cost-effective and extend usable lifetimes. This capability might be critical to coping with orbital space debris problems.

It is noteworthy to understand that some of the techniques and capabilities needed to undertake on-orbit servicing, repair, or satellite upgrades are quite parallel to the capabilities needed to undertake active orbital debris removal or mitigation. This chapter examines some of the new capabilities that are being developed to carry out on-orbit servicing, repair, or repurposing. This chapter also include some brief discussion of how these technologies might be commercially applied to space debris mitigation and active removal techniques – and in the relatively near future.

Keywords

Autonomous Space Transport Robotic Operations (ASTRO) • CleanSpace One • ConeXpress • Defense Advanced Research Projects Agency (DARPA) of the USA • DART mission of NASA • DEOS mission of German Space Agency • DEXTRE robotic manipulator of NASA • DLR of Germany • Enhanced Orbital Replacement Unit Temporary Platform (EOTP) • European Space Agency (ESA) • Japanese Space Agency (NASDA and JAXA) • MacDonald Dettwiler and Associates (MDA) Space Infrastructure Servicing (SIS) • Orbital Express program of DARPA • PRISMA of Sweden • Raven mission • Rendezvous and proximity operations (RPOs) • Robotic Refueling Mission (RRM) • US Air Force XSS-11 mission • ViviSat

Introduction

A great deal of progress has been made in what are often called “on-orbit operations” in recent years. These technical developments include the ability to locate and mate with other orbiting satellites, on-orbit refueling, retrofit and repair of satellites, and artificially intelligent and ground-controlled robotics to carry out a variety of missions. Space-faring nations that have been most active in this area include the United States (NASA and DARPA), the European Space Agency (ESA), Germany (DLR – the German space agency), Japan (NASDA-JAXA), and China (Chinese National Space Agency). In addition there have been a number of commercial space ventures that have sought to develop new technology related to active space debris removal and on-orbit servicing. Some of these commercial efforts are described later in this chapter. This is a highly sensitive area in that many of the technologies that might be developed for these purposes might also be consider a space weapon in that such a capability might be used for deactivating or attacking a satellite or space vehicle.

Currently the future of satellite applications seems to be at a crossroads. Some see the future involving the deployment of a large number of small satellites deployed in low earth orbit constellations with such satellites being manufactured like television sets in automated production lines. Others see larger and larger satellites of very high power that would typically be deployed in geosynchronous satellites. These are known as high-throughput satellites. There are yet others that see the potential of refueling satellites and providing them with enhanced solar cell arrays and new battery systems and even new or augmented payload systems so that existing networks can have their practical lifetimes extended by many years and performance upgraded. Some of these concepts envision such future concepts as using 3-D printers in space to fabricate replacement components, the use of specialized robots to engage in harvesting of solar arrays or antenna systems from defunct satellites to reuse them on retrofitted spacecraft.

It is, of course, possible that all three of these alternative futures (i.e., high-throughput satellites, MegaLEO constellations, or retrofit and repair of existing satellites) could be implemented by different space application operators. The factors that could shape the future are manyfold. They include (i) further cost reductions in commercial launch services; (ii) the extent to which increased orbital debris build-up continues to occur and impinges on space safety and/or results in accepted debris mitigation practices; (iii) the development of new and highly competitive terrestrial or “protozone-based” high altitude platform systems; and (iv) the successful development of many critical new technologies in areas such as space robotics, space tugs to support spacecraft refueling, enhanced digital encoding and processing capabilities, and improved space power systems (such as quantum dot solar power systems, enhanced batteries, 3-D printers designed for space-based operations, etc.). Even entirely new factors such as the decreasing strength of the Earth’s magnetic field and its reduced ability to protect satellites from destructive solar storm events could come into play.

This chapter concentrates on just one thing. This is the potential to create significant new abilities to extend the lifetime of application satellites and/or retrofitting and repairing them so that their usable lifetime might be significantly extended. Thus, the focus of this chapter is on robotically enabled missions. This starts, in the first instance, with the ability to accurately locate and dock with existing spacecraft and then once linked to a spacecraft to carry out multiple tasks. These tasks could then include the repair, retrofit, or upgrade of satellite power systems and replacing or augmenting payload systems related to telecommunications and broadcasting, navigation and timing, remote sensing, or meteorological-related services. This analysis will also note that some of the systems designed to accomplish such missions might be designed with the dual purpose of achieving the active deorbit of orbital space debris or steering a satellite into an altered orbit where atmospheric drag could accelerate deorbit over time so as to meet the UN recommended rule of achieving deorbit within 25 years of end of life.

Maneuvering and Mating in Space with Servicing Vehicles

There has been a great deal of experience acquired in using optical sensors and docking systems in space for several decades now. The following listing provides just some of the experimental activity that has been carried out around the world. Japan has carried out early experiments in deep space to undertake satellite location and robotic docking operations under its Experimental Test Satellite-VII program. NASA and participants in the International Space Station have acquired a good deal of experience with docking and capturing spacecraft using the Canadarm and the so-called RRM and DEXTRE mechanism for simulated repair and refueling operations. In the past two decades, the following efforts to accomplish on-orbit activities in space, commonly known as rendezvous and proximity operations (RPO), have been carried out with the results briefly noted below:

- The Japanese space agency in 1997 (then named NASDA) carried out several docking missions with the ETS-VII mission with a chaser and target system.
- The US Air Force XSS-11 mission in 2005 accomplished a close proximity inspection of several satellites with general success, but this test flight did not attempt a docking (U.S. Air Force).
- The NASA DART spacecraft (Demonstration for Autonomous Rendezvous Technology) in 2005 attempted an autonomous rendezvous with a spacecraft known as MUBLCOM satellite. This disabled spacecraft was no longer functional or capable of any maneuverability. This mission was only partially successful since there was a slight collision during this test of a RPO activity (NASA Dart).
- The US Defense Advanced Research Project Agency (DARPA) undertook a test with spacecraft named the Orbital Express during 2007. This Orbital Express spacecraft demonstrated the ability to carry out on-orbit refueling and servicing of another spacecraft (Orbital Express 2009).

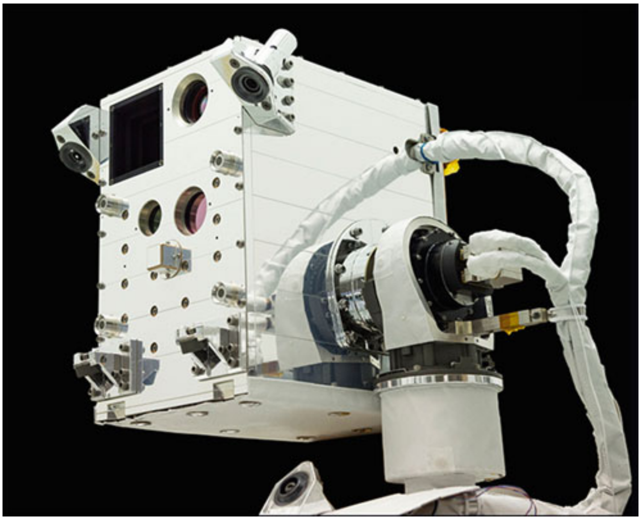


Fig. 1 The Raven 2 robotic serving and “relative time navigation” system

- The Swedish Space Corporation PRISMA in 2010 successfully demonstrated the ability of two microsatellites to fly in close proximity formation (Successful Launch 2010).
- The Chinese SJ-12 in 2010 maneuvered close to the SJ-06 F spacecraft for reasons thought to be close proximity inspection. A year later in October 2011 after successful liftoff, Shenzhou 8 docked with China’s Tiangong 1 spacecraft, which had been launched in September 2011. This was China first in-space docking (China’s 1st Space Docking Mission 2011).
- The Robotic Refueling Mission together with DEXTRE has demonstrated the ability to carry out refueling and repair at simulations conducted at the International Space Station.
- The German DEOS program will shortly carry out a number of tests with two small satellites – a chaser and target – that is similar in nature to the Orbital Express experiments.
- After it launches to the International Space Station in 2016, Raven will demonstrate a real-time relative navigation system that would enable future spacecraft to autonomously rendezvous with both prepared vehicles and those not designed for servicing (Introducing RAVEN 2016) (See Fig. 1).

The technology that allows spacecraft to dock and undertake refueling and retrofit operations is also useful in another context other than spacecraft operations. The new trend to deploy small satellites in very large low earth orbit constellation requires very precise system management and control of thrusters to maintain spacecraft in their proper orbits. When the Iridium satellite constellation was first deployed, there were a number of instances of operator error (sometimes called cockpit error) that led to the loss of some satellites during the constellation’s early deployment. In the new era where LEO constellations might include thousands of satellites in a single network as well as thousands of satellites in competitive networks, precision control becomes ever more important. The precise control of operational and spare satellites and the avoidance of collision with other satellites are important capabilities. These techniques are not only key in terms of avoiding the loss of satellites, but this is important to eliminating on-orbit crashes that could generate thousands of new elements of orbital debris.

The two major orbital collisions that generated thousands of new elements of orbital debris involved, in the first instance, the shooting down of the defunct Chinese weather satellite Fengyun IC by Chinese military in January 2007 (Keck 2014). This was followed by the crash between Iridium 33 and a Russian Kosmos 2251 satellite in February 2009 (Keck 2014). The other case was the collision between a Chinese missile and a defunct Chinese Funyun weather satellite. These two collisions generated almost 6000 trackable orbital debris elements or about 25 % of the debris population. Further such collisions in low earth orbit could jeopardize the future of all types of satellite applications. In short, software and operational control developed to engage in docking of satellites can also help with the precise control and maneuvering of satellites in large constellations of satellite and the avoidance of collisions.

The deployment of such constellations with a very large number of satellites is considered a major risk factor and is thus one of the latest concerns that might jeopardize all future satellite operations. Dr. Donald Kessler, the NASA scientist that first warned of the possible dangers of cascading orbital debris that is today known as the "Kessler syndrome," has projected that a major collision will occur about once in 10 years going forward. This prediction by Dr. Kessler, however, was made prior to current plans to deploy MegaLEO constellations that undoubtedly will greatly increase the projected risk factors (Interview).

On-Orbit Servicing, Retrofit, and Repair of Communications Satellites

The efforts to develop the technology to carry out on-orbit servicing, retrofit, and repair of defective or broken elements of an applications have increased significantly in recent years. The follow section reports in more detail on some of the most important on-orbit test that have been carried out.

Orbital Express Space Operations Mission

This joint program of the US Defense Advanced Research Projects Agency and the NASA Marshall Spaceflight Center was designed to test both spacecraft retrofit and active deorbit of debris. The Orbital Express program experiment was launched on March 8, 2007, using an Atlas V launcher. This project was designed to test the on-orbit interaction of two different especially designed satellites. The larger spacecraft of the two was the ASTRO "servicing spacecraft." The smaller NEXTSat served as the "client" spacecraft to be captured. The NEXTSat spacecraft was envisioned as a prototype design for future spacecraft that could be designed for in-orbit servicing, retrofit, and refueling.

The two satellites were designed for proximity maneuvering in space and thus could duplicate the activities necessary to capture a debris element for active debris removal. Secondly, if it is possible to service on-orbit satellites to resupply them with fuel, batteries, and new electronics and antenna systems, then the population of satellites launched into orbit can be reduced. This means fewer satellites and upper stage rockets that would need to be disposed of and thus would lead to the creation of less space debris. As noted above, however, plans to deploy MegaLEO constellations with thousands of satellites represents a new trend that could lead to a major increase in orbital debris risks.

The acronym ASTRO for the server satellite stood for Autonomous Space Transport Robotic Operations. This ASTRO servicing satellite was almost 1000 kg in total mass and was fueled with nearly 140 kg of hydrazine propellant. Its height and diameter were nearly 2 m. Its robotic arm allowed for capture and manual docking. During docking, it was possible to transfer fuel or retrofit or augment elements of the NEXTSat target vehicle. This DARPA mission indicated that retrofit or refueling could allow having space capabilities on-orbit much faster than building and launching a new satellite (see Fig. 2).

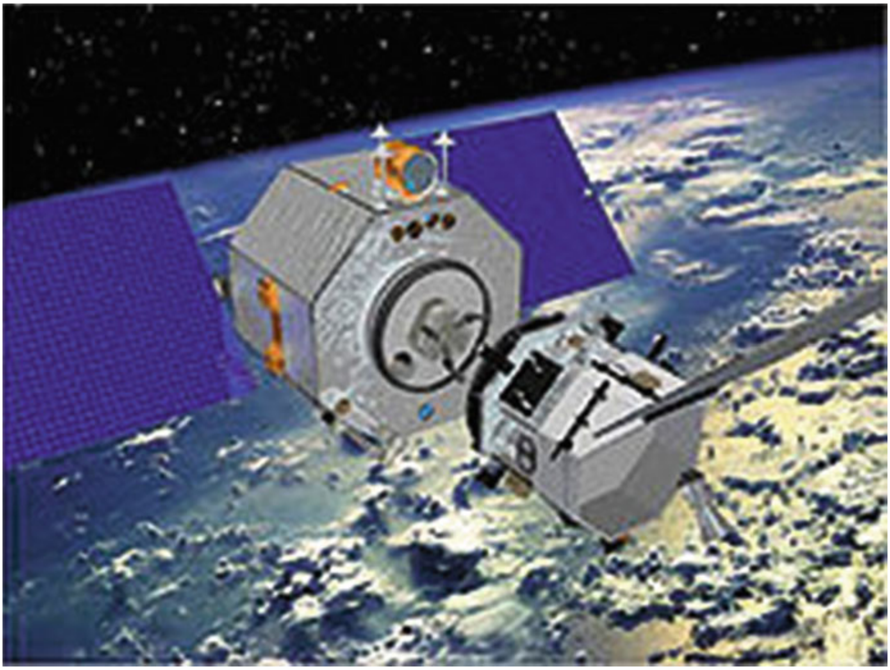


Fig. 2 The ASTRO "servicing" spacecraft and smaller NextSat pictured in orbit (Graphic Courtesy of NASA)

The NEXTSat target spacecraft had a mass of only 225 kg and was only about 1 m in diameter. Both of these spacecraft are depicted as flying in orbit in the figure below (Orbital Express Space Operations).

This joint program cost about $300 million for the design and fabrication of the two spacecraft and the Atlas V launch. This was the first such space experimental program for on-orbit servicing, although Japan in 1997 (i.e., then NASDA and now JAXA) was able to carry out the first robotic rendezvous docking between two spacecrafts in orbit under its experimental test satellite (ETS-VII) program (ETS-VII).

NASA Robotic Refueling Mission (RRM)

The NASA Robotic Refueling Module (RRM) was installed on the International Space Station by the Atlantis Space Shuttle in 2011. This activity represented the last official mission for the shuttle launch system before it was retired. This specially designed module has a mass of approximately 250 kg and was configured in a shape essentially like a 1 m cube. The RRM contained a wide range of multiuse tools that were used to conduct a number of experiments involving the repair, retrofit, and augmentation of a hypothetical spacecraft in orbit. The RRM contained many different types of tools available on demand. The most significant experiment was demonstrating how the equivalent of nearly 2 l of ethanol could be transferred to a satellite lacking fuel without the liquid escaping into space. The RRM experiments confirmed that not only spacecraft designed for on-orbit retrofit or refueling could be serviced in space but also satellites that had not been designed for this purpose could also be repaired, updated, or given new fuel. Part of the RRM experiments relied on the highly flexible robotic system available on the ISS known as the DEXTRE. This is the dexterous robotic system that is also called by NASA as the Special Purpose Dexterous Manipulator (SPDM).

The key to the RRM project experiments was the ability to use the specially designed DEXTRE or dexterous robotic extension of the Canadarm 2 system that is installed on the International Space Station (ISS). Both the Canadarm 2 and DEXTRE were designed and manufactured by the Canadian Space Agency (CSA) that has made space robotics its area of special competence.

The DEXTRE robotic manipulator is capable of many complex operations that can be executed through ground commands and has been used for many repairs on the ISS quite separate from the satellite repair and refueling experiments associated with the RRM tests (DEXTRE).

The key to the RRM experiments was the utilization of DEXTRE's Enhanced Orbital Replacement Unit Temporary Platform (EOTP). After the Atlantis Shuttle departure, the RRM unit was installed at its permanent location on the ISS known as the ExPRESS Logistics Carrier 4 (ELC-4). This location was necessary in that it allowed the RRM toolkit to establish telecommunications links to NASA's ground command so that the DEXTRE system could carry out many simulated repair experiments.

This new configuration allowed the DEXTRE robot to retrieve RRM tools from a multi-tool module. Ultimately the RRM experiments included manipulating, cutting, and repositioning wiring and uncovering protective blankets that would otherwise obstruct repair operations. It also allowed the unscrewing of a variety of caps and access valves in order to transfer fluid and simulate refueling. At the end of this operation, DEXTRE was able to put a new fuel cap on the fuel tank that had been opened. Specifically RRM tools were used to open up a fuel valve and transfer its stored liquid ethanol from one tank to another using a robotic fueling hose.

The NASA RRM mission, since it was able to use the DEXTRE robotic system installed on the International Space Station, was able to accomplish these retrofit, repair, and refueling experiments at much lower cost than the Orbital Express mission. Further these operations were much more detailed and intricate than those conducted on the Orbital Express mission (NASA Robotic Refueling Missio).

Deutsche Orbitale Servicing (DEOS) Mission

A very similar on-orbit servicing mission is currently being carried out by the German Space Agency (DLR) in 2016. The spacecraft manufacturer and prime contractor for this mission is the German firm known as SpaceTech Gmbh Immenstaad. In this case also there is a “servicer” spacecraft (known as the Phase A program) as well as a “client” spacecraft (known as the Phase B program) that is captured and then services provided to it. The specific objective of the DEOS program is to demonstrate how a defective spacecraft that is tumbling in an uncontrolled manner could be captured and suitably retrofitted so that it could resume operations rather than becoming a defunct spacecraft and thus worthless space debris. Further, this mission is designed so that if the on-orbit servicing program to restore operational capability to the “client” satellite is for some reason not successful, then the “servicer” (or capturing spacecraft) can link together with the “client” and successful deorbit both spacecraft in tandem so as to eliminate ongoing orbital debris (DEOS).

The figure below shows the DEOS robotic “servicer” (phase A) spacecraft and the “client” (phase B) spacecraft about to be caught by a grappling arm in space. In addition to capturing and stabilizing the tumbling “client” satellite, the “servicer” will seek to undertake difficult refueling operations as well. Finally, it will seek to retrofit electrical and other equipment on the “client” spacecraft. The techniques involved in this mission would, of course, be quite parallel to efforts to capture a defunct satellite for subsequent deorbiting (Fig. 3).

The DEOS project is designed so that both spacecraft will be directly in communications with the ground at all times. During the special Low Earth Orbit Proximity (LEOP) experiments that DLR will conduct, the use will be made of a geosynchronous data relay satellite to maintain communications. In addition, a backup supplementary ground station network will be available. This ground network will act as a fail-safe tracking, telemetry and command, and control capability (DEOS).

Fig. 3 The DEOS experiment shows phase A (servicing satellite) and phase B (client) (Image courtesy of DLR, the Germany Space Agency)

In addition to the directly relevant DEOS program, the German space agency DEOS has another development program involving synthetic aperture radar (SAR) imaging by two satellites flying in close formation just a few 100 m apart. This involves the TerraSAR-X (TSX) launched in 2007 and TanDEM-X (TDX) launched in 2010. The techniques involved with keeping these two satellites in very close proximity without colliding are, of course, quite useful to understanding how to mate satellites in orbit without crashing into one another (Maurer et al. 2012).

Phoenix Program by DARPA

The Phoenix Program by the US Defense Advanced Research Projects Agency (DARPA) represents the continuing engineering and design programs of this agency in the area of on-orbit servicing and robotic construction in space. This is clearly an extension of the earlier Orbital Express project. This program, rather than being in low earth orbit (a few 100 km above the Earth's surface), is planned for the much more difficult and demanding regime of geosynchronous orbit (GEO). Ground-based telecommands are much more difficult because commands and responses

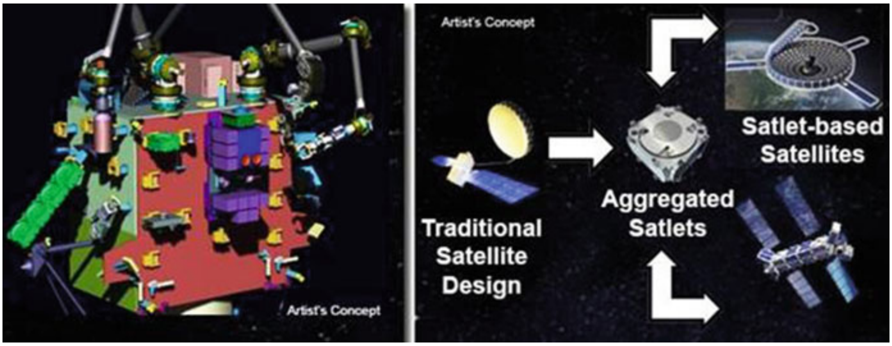


Fig. 4 DARPA architectural concept of robotic mechanism capable of integrating Satlet Modules to create new functional spacecraft (Graphics courtesy of DARPA)

involve a quarter to a half second delay since in the GEO orbit is almost a tenth of the way to the Moon. Mating a spacecraft with another or establishing it in a precise orbit in LEO is much, much easier than in GEO Orbit. Coping with the movements of a satellite 300–400 km away is much different than when the spacecraft is at least 36,000 km if not 50,000 km away.

And the objectives are much more ambitious than seeking to move a satellite at end of life to a safe "parking orbit," with a range of much more difficult tasks. The end result is that the Phoenix program is undertaking the creation of a whole new architecture for satellite design and reconstruction via on-orbit robotic systems. These concepts involve assembly and disassembly and efforts not only to achieve extended capability in orbit but taking components or even antennas and solar arrays from one satellite and installing them on another on-orbit . Some of the concepts involve the use of modular units that could be assembled to create larger and more capable spacecraft, perhaps over several years or longer. Over time, part of the idea in the new lexicon of DARPA might be to redeploy so-called satlets in order to renew their use as totally reconstituted satellite systems. (See graphic below)

This would take the concept of active debris removal to a whole new level, not just going out and "collecting" space junk to deorbit but rather to go out in order to "harvest" it and then integrate at least parts of the old spacecraft into a new functional spacecraft. This approach might make coping with orbital debris more cost-effective on one hand, but on the other, it would undoubtedly require some innovation with regard to applicable space law since such salvage and reuse is not part of the currently understood rule of law with regard to outer space (Phoenix) (see Fig. 4).

This Project Phoenix effort certainly raises new aspects of international space law. Does outer space salvaging translate as an exact parallel to the law of the sea? Do such concepts conjure up a vision that this would be a sort of space weapon that could act not only to deorbit or harvest parts on a defunct space object but also could disable or attack the spacecraft of other nations?

Raven: The Autonomous Rendezvous Experiment

Raven is yet another joint DARPA and NASA development project that in this instance represents a follow-on to the Robotics Refueling Mission. It is a part of the DARPA Space Test Program-Houston 5 (STP-H5) payload. The Raven objective is to optimize real-time relative navigation system for proximity navigation. It is designed for on-orbit servicing including those spacecraft not originally designed for servicing. This experiment flies on the International Space Station (ISS) and was launched in early 2016.

For this experiment, the STP-H5 payload is mounted on an exterior platform (ELC-1). Mission operators can use this platform to transmit data related the instantaneous tracking of arriving and departing spacecraft to the ISS. The goal is to improve Raven's performance in preparation for space flight on an independent robotically controlled autonomous spacecraft.

Using the cost efficiency of the International Space Station as a test bed, the DARPA-NASA team will examine how Raven's sensors, avionics, and analytic algorithms document the ability of future spacecraft to effectively mate with other satellites (Raven Advancing Autonomous Rendezvous Technologies).

Commercially Backed Orbital Remediation Programs and Initiatives

There are a number of private companies and institutions that are intent on seeking to address the space debris problem.

CleanSpace One

This is a project of the Swiss Space Center and the Federal Polytechnical School of Lausanne – or the Ecole Polytechnique Federale de Lausanne (EPFL). It began with students designing a CubeSat for scientific measurements with the mission to observe and map airglow – a light phenomenon found in the upper atmosphere. This project was launched in 2009 and completed its mission after several years in orbit. In February 2012, Professor Volker Gass, Director of Swiss Space Center (SSC), decided it would be desirable to try and design a small satellite capability that could track and retrieve the original CubeSat.

With the support of the Swiss Space Center and EPFL, the CleanSpace One project was thus born. Professor Volker Gass stated: "Our work is guided by the principle that the person responsible for the mess is also responsible for cleaning it up. If everyone were to put their own house in order, then outer space would be neat and tidy" (Swiss Create).

Claude Nicollier, the first Swiss astronaut and currently Professor of Spatial Technology at the EPFL, is likewise a strong proponent of this project as well and has said: “It has become essential to be aware of the existence of this debris and the risks that are run by its proliferation.”

Currently, the CleanSpace One that is a small three-unit cube satellite (30 × 10 × 10 cm) is planned for launch in 2016 or 2017. Figure 6 shows a simulation of the CleanSpace One spacecraft overtaking the original CubeSat launched in 2009.

The tracking and rendezvous for the CleanSpace One are quite complicated as shown in the attached graphics. The concept is for it to clamp on to the first CubeSat, and then they would deorbit in tandem. The graphic below is current, but the indication of 16,000 tracked space debris elements as indicated in the graphic window is no longer the latest count. As noted earlier, there are 22,000 objects of 10 cm or larger now being tracked.

This project is clearly more an act of principle and public commentary than a full-scale program that will make a major contribution to the orbital debris cleanup effort. It is the removal of the largest debris elements in low earth orbit that is most critical, and this effort would remove only one element out of over 22,000. The publicity that this program has generated, however, is in itself helpful. The Swiss effort to clean up their debris may well inspire other countries to follow suit. Public opinion is a key part of the effort to “cleanup space” (Orbital Cleanup).

Dutch Space and the ConeXpress Orbital Vehicle for Life Extension

This is a life extension project to provide on-orbit servicing to geosynchronous satellites. The project known as ConeXpress is designed to exploit the spare capacity of Ariane-5 so as to make the launch of this system as cost-effective as possible. Thus, the ConeXpress uses the conical section under the primary satellite payload fairing structure. A launch of the ConeXpress would utilize the standard Ariane-5 conical payload adapter as its main structure. This approach allows for a launch to GEO orbit for an estimated cost of only about 35 million euros. The proposed approach for lifting a failed payload launch from a lower or medium earth orbit to GEO orbit would be accomplished by using electric propulsion. This would use the technology developed for the SMART-1 mission to the Moon that the European Space Agency has successfully demonstrated. To a certain extent, this concept of a mission to extend the life of a GEO orbiting satellite derives from the ESA’s Robotic GEO Orbit Restorer (ConeXpress Orbital).

According to the analysis provided by Orbital Recovery Limited, the ConeXpress Orbital Life Extension Vehicle could be used to extend the life of a GEO satellite by up to 12 years. Such an approach to satellite lifetime extension is no long just a theoretical concept. It has now been reported in detail by Intelsat that its reboosting of the failed launch of the Intelsat 19 to GEO orbit resulted in up to $800 million in added revenues to be generated from this reclaimed satellite. A further bonus was that this failed satellite launch would otherwise have become a space debris element in lower earth orbit.

The ConeXpress platform is currently being developed by Dutch Space in Leiden in cooperation with the European Space Agency in the Netherlands. The anticipated weight of the ConeXpress at launch would be 1400 kg, and it would be stowed on the Ariane 5 within a 2.6 m diameter and 1.35 m height conical shape, and its solar power array would generate about 4 kW. Ariane-5 launch schedules currently offer several opportunities per year to make use of its otherwise-unused capacity in the cone shaped part of its launch configuration. The ConeXpress stack comprises the following components. These are (i) the payload adapter, (ii) an extension cylinder incorporating a separation mechanism, and (iii) mountings for the inner structure. The inner structure accommodates equipment such as avionics and the rendezvous and the docking payload. The ConeXpress deploys its antennas, solar wings, and thruster-steering mechanisms after release from the Ariane 5. It is then ready to steadily fire its electric ion thrusters that will take it on a slow spiraling orbit during what could be up to a 6-month journey to GEO. During this long transfer operation, and while preparing for rendezvous and docking with a GEO satellite, ConeXpress looks like a small conventional geostationary communication satellite with its solar panels pointing north–south.

To date, there are no confirmed customers for the ConeXpress Orbital Lifetime Extension Vehicle, but it could clearly be used not only to extend the lifetime of GEO satellites but also to elevate them to the end-of-life parking orbit operating as a space tug. It could also serve to reduce orbital space debris by disposing of satellites at end of life (ConeXpress OLEV 2014).

MacDonald Dettwiler Associates' Space Infrastructure Servicing (SIS) Vehicle

The MDA Space Infrastructure Servicing (SIS) vehicle is advertised as one of the first operational capabilities to provide a robotics and docking system for a number of possible on-orbit operations. This system will be based on work that MDA has previously performed for NASA and the Canadian Space Agency with regard to the Canadarm 2 and DEXTRIX robotic systems as well as for various Department of Defense agencies. The SIS vehicle's robotic arm is being designed to be used for refueling but can also be used for many other tasks as well. This vehicle could be used to support on-orbit repairs, maintenance, or other tasks such coping with antennas or solar arrays that are stuck or did not fully deploy. It could also be for towing smaller space objects into alternative orbital locations or removal of space debris from geosynchronous orbit or other tasks (Foust 2013) (see Fig. 5).

An initial arrangement was announced in March 2011, under which Intelsat would utilize the MDA SIS craft for on-orbit servicing of its satellites. Subsequently some 10 months later, however, Intelsat and MDA were not able to conclude specific contractual arrangements, and this agreement was terminated as of January 2012. To date, no other satellite operator has signed up to use on-orbit servicing or mission extension services (Doug Messier).

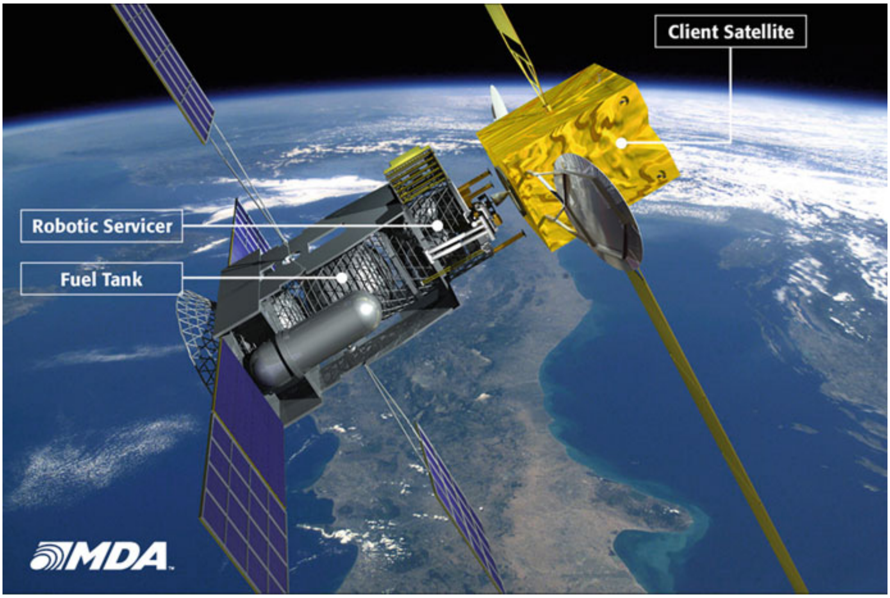


Fig. 5 The MacDonald Dettwiler Associates (*MDA*) space infrastructure servicer attached to a "client" satellite (Graphic Courtesy of MDA of Canada)

The technology for on-orbit services is now proven in a number of governmental and commercial systems, but the market that is supported by commercial operators has yet to develop. It seems likely that the systems to extend the life of satellites via on-orbit servicing is likely to develop first. This means that active on-orbit debris removal (or boosting to a graveyard orbit) would likely evolve subsequently. In some instances, on-orbit servicing vehicles will be used both for mission extension and could subsequently be used to remove spacecraft to graveyard orbits as its final mission.

At this time, ConeXpress, ViviSat, and MDA are all at a stage where they could manufacture and operate on-orbit servicing systems if there were contractually committed customers. These vehicles could assist with the following:

- Extension of the life of operational satellites by refueling and more
- Elevate satellites to GEO orbit in the case of failed launches
- Recycle old satellites to new uses
- Provide transport services to move GEO satellites to graveyard orbits 300 km above GEO
- Assist large satellites in LEO orbit to reenter the Earth's atmosphere and burn up

Despite the positive results achieved by Intelsat to move a satellite to GEO orbit and provide a full lifetime of services, such a market has not yet developed. Currently there are no commercial operators or countries willing to sign up for these services. In short, the lack of commercial or governmental customers has delayed progress in this area.

ViviSat Mission Extension Vehicle

Another commercial approach to mission extension and retrofit space vehicles is known as ViviSat. This Mission Extension Vehicle is being designed as a cost-effective and streamlined spacecraft capable of on-orbit servicing. This simple but versatile spacecraft is being designed so it could be employed by satellite owners to extend mission life and also help to dispose of geosynchronous satellites at end of life. It has been "advertised" as an alternative to the MacDonald Dettwiler and Associates (MDA) and its Space Infrastructure Servicing (SIS) vehicle discussed above. The claim made by ViviSat is that their docking vehicle could mate successfully with a higher percentage of the nearly 500 geosynchronous satellites that are currently in orbit – or scheduled to be launched.

The ViviSat module is being designed to link up with a satellite that has depleted its fuel but is otherwise operational. This module can thus provide additional fuel and possibly provide other upgrades and thus allow continued operations. An alternative application would be to rescue a satellite that had been unsuccessfully launched and not fully achieved geosynchronous orbit such as has now been successfully done by Intelsat. In this case, the ViviSat module would ferry the satellite to its intended GEO orbit location and then release it to operate normally once it had been checked out by on-orbit test of its various capabilities.

ViviSat is a partnership of US Space Inc. with ATK (now Orbital ATK). This mission extension vehicle is designed to use the A700 satellite bus. The design of the ViviSat module was announced as being "finalized" in March 2012 and was thus ready for construction as visualized below. At this time, no satellite operators have been willing to sign on as customers for this on-orbit type servicing, and thus the fabrication of the ViviSat unit is pending such a contract. The problem related to mission extension vehicles, on-orbit services modules, and spacecraft capable of active space debris is that there is currently a lack of an established customer base willing to pay for the construction and operation of such a new type of space vehicle. The figure below illustrates how a ViviSat mission extension vehicle would look in space while mated to a "client" satellite that was being accessed for refueling (Satellite Life Extension Services) (see Fig. 6).

Other Initiatives of Relevance

The various civilian space agencies around the world plus DARPA are investing the most those at research effort and monies to develop space robotics, mission extension vehicles, close proximity maneuvering and mating in space navigational

Fig. 6 Graphic depicting ViviSat mated to a “client” satellite (Graphic Courtesy of ViviSat)

systems and retrofit, and refueling capabilities. The next generation of research may even seek to use 3-D printers to fabricate replacement parts of satellites with failed components. The above discussion highlights some of the most important efforts to develop new robotic systems for refueling and retrofit of satellites as well as systems that could assist with active deorbit of defunct satellites that have become space debris elements. In addition to the governmental research programs, there are smaller efforts that involve nanosatellites, CubeSats, and small satellites that are trying to develop new technology, software, and robotic capability that relates to on-orbit servicing and/or active orbital debris removal. The CleanSpace One project presented above is just one of these efforts and was selected because it is currently in process. Some of these smaller-scale programs include governmental, commercial, and even university research programs such as those at Stanford University, EPFL, and the Technical University of Delft. These efforts include the following:

- “CleanSpace One” undertaken by the Swiss Space Systems (S-3) together with EPFL of Switzerland
- TanDEM-X by Germany’s DRL
- PRISMA with situational pointing and control (Italian Space Agency)
- GRACE to plot Earth’s gravity variation (by NASA)
- Stanford University Space Rendezvous Laboratory (SLAB)

The research at SLAB, for instance, is based on 10 years of experience in the implementation and flight operations of GNC subsystems for formation flying and on-orbit servicing missions (e.g., GRACE, TanDEM-X, PRISMA, DEOS, etc.). Ultimately partnerships at national level (e.g., NASA Ames, JPL, AFRL, etc.) and international level (e.g., DLR, DTU, ESA, TU Delft, etc.) will pave the way for breakthrough demonstrations of new technology (Stanford Space Rendezvous Laboratory).

Conclusion

The evolution of space applications over the decades has allowed satellites to become more and more capable, flexible, cost-efficient, and able to achieve much longer lives. One of the newest capabilities that is just now evolving from concept to practical development is that commonly called mission extension vehicles. These spacecraft can not only extend the lifetime and functionality of satellites but also redeploy them into practical orbits at the beginning of life. At the end of life, these same systems could also be used for disposal of defunct satellites in order that they do not become dangerous orbital space debris. Various space agencies plus the US DARPA are investing significantly in the development of such new types of systems for a variety of practical and strategic reasons. Commercial space ventures such as MacDonald Dettwiler Associates, ViviSat, and ConeXpress have identified potential commercial markets for mission extension vehicles, but the actual award of contracts in this new area has been slow to date.

It is not clear which applications may prove most important in the longer run. At this stage, the options remain severalfold. The rescue of a satellite meant for GEO orbit that is erroneously placed in a low earth orbit due to launcher failure has been demonstrated in a practical sense. The use of such spacecraft for lifetime extension by refueling and retrofit also has considerable future potential. Such systems could assist with active orbital debris removal at end of life. Finally, such technology may prove to have strategic value both in terms of rapid deployment of capacity and an offensive or defensive space capability.

As this technology matures and a full capacity to deploy space tugs and service vehicle in GEO, and provide transport between LEO and GEO, the practicality of these technologies will be better proven. Today such on-orbit servicing, precision navigation and mating, and robotic construction and applications are just starting to demonstrate their capabilities and economic attractiveness. In future years the business case will become much clearer. It is also possible that the launch insurance industry may in future years also see an opportunity in this new emerging space capability to finance or serve as underwriters for on-orbit repairs, retrofit, or even debris removal operations.

Cross-References

▶ Lifetime Testing, Redundancy, Reliability, and Mean Time to Failure
▶ Tracking of Orbital Debris and Avoidance of Satellite Collisions

References

Agreement, (2012), http://www.parabolicarc.com/2012/01/13/mda-terminates-satellite-servicing-agreement-with-intelsat-looks-south-for-business/. Last accessed 31 Jan 2012

China's 1st Space Docking Mission, Space.com (2011), http://www.space.com/13451-china-space-docking-test-shenzhou-8-german-experiment.html. Last accessed 4 Dec 2015

ConeXpress Orbital Life Extension Vehicle – ESA, (EU & Industry *esa* bulletin 127 - august 2006), www.esa.int/esapub/. . ./bul127h_caswell.pdf. Last referenced 4 Dec 2015

ConeXpress-OLEV (CX-OLEV), Gunter's Space Page. EU & Industry *esa* bulletin 127 (2006) (2014), http://space.skyrocket.de/doc_sdat/conexpress-ors.htm. Last referenced 4 Dec 2015

DEOS, (2013), http://www.research-in-germany.org/en/research-areas-a-z/space-technology/Research-Projects/DEOS.html. Last referenced 4 Dec 2015

DEXTRE, (2013), https://www.nasa.gov/mission_pages/station/structure/elements/dextre.html#.VY2kG_lVhHw. Last referenced 4 Dec 2015

ETS-VII (Engineering Test Satellite VII)/Kiku-7, (1999), https://directory.eoportal.org/web/eoportal/satellite-missions/e/ets-vii. Last referenced 4 Dec 2015

J. Foust, Satellite servicing efforts grapple with the business case, (2013), http://spacenews.com/34747satellite-servicing-efforts-grapple-with-the-business-case/#sthash.wnEnB9nB.dpuf. Last referenced 14 Apr 2013

Interview with Dr. Donald Kessler by Joseph N. Pelton, June 2014

Introducing RAVEN that will launch in 2016, http://ssco.gsfc.nasa.gov/Raven_Autonomous_Rendezvous.html. Last referenced 4 Dec 2015

Iridium/Cosmos Satellite Collision, (2009), www.space.com/5542-satellite-destroyed-space-collision.html. Last referenced 4 Dec 2015

Z. Keck, China conducted anti-satellite missile test, (2014), http://thediplomat.com/2014/07/china-conducted-anti-satellite-missile-test/. Last referenced 4 Dec 2015

E. Maurer, S. Zimmermann, F. Mrowka, H. Hofmann, Dual satellite operations in close formation flight, German Space Operations Center (2012), http://www.spaceops2012.org/proceedings/documents/id1275173-Paper-004.pdf. Last referenced 4 Dec 2015

NASA Dart, (2005), https://www.nasa.gov/pdf/148072main_DART_mishap_overview.pdf. Last referenced 4 Dec 2015

NASA Robotic Refueling Mission, (2015), http://ssco.gsfc.nasa.gov/robotic_refueling_mission.html. Last referenced 4 Dec 2015

Orbital Cleanup Satellite to be Launched in Partnership with S3, (2012), http://space.epfl.ch/CleanSpaceOne. Last referenced 4 Dec 2015

Orbital Express: Prototype Satellites Primed for In-Flight Servicing, Space.com (2009), http://www.space.com/3546-orbital-express-prototype-satellites-primed-flight-service.html. Last referenced 4 Dec 2015

Orbital Express Space Operations, (2007), http://archive.darpa.mil/orbitalexpress/index.html. Last referenced 4 Dec 2015

Phoenix, DARPA's satellite-recyling program, Space.com, (2008), http://www.space.com/13342-phoenix-darpa-satellite-recycling-program.html. Last referenced 4 Dec 2015

Raven Advancing Autonomous Rendezvous Technologies, (2015), http://ssco.gsfc.nasa.gov/Raven_Autonomous_Rendezvous.html. Last referenced 4 Dec 2015

Satellite Life Extension Services, (2015), http://www.vivisat.com/?page_id=10. Last referenced 4 Dec 2015

Stanford Space Rendezvous Laboratory, (2015), https://people.stanford.edu/damicos/. Last referenced 4 Dec 2015

Successful Launch of the Swedish PRISMA Satellites, (2010), http://www.sscspace.com/successful-launch-of-the-swedish-prisma-satellites. Last referenced 4 Dec 2015

Swiss Create Janitor Satellite for Space Clean Up. The Guardian, (2012), https://www.theguardian.com/science/2012/feb/15/swiss-create-janitor-satellite-space-cleanup. Last referenced March 30, 2016

U.S. Air Force XSS-11 Micro Satellite Mission, http://www.kirtland.af.mil/shared/media/document/AFD-070404-108.pdf. Last referenced 4 Dec 2015

Advanced Manufacturing Technologies and 3D Printing

Yves Durand, Martine Lutz, and Florence Montredon

Contents

Abstract

This chapter gives an overview of a sector of satellite technology which is rapidly developing and has to be taken into account when planning a new space mission. Additive manufacturing, usually called 3D printing, is extremely well adapted to the constraints of spacecraft development, therefore quickly gaining acceptance in the field of space technology. But 3D printing is not the only innovative technology that may change the way that satellites will look like in the future. In addition to the expected continuous integration of electronics, new materials,

---

Y. Durand (✉) • M. Lutz • F. Montredon
Thales Alenia Space, Cannes, France
e-mail: yves.durand@thalesaleniaspace.com; martine.lutz@thalesaleniaspace.com; florence.montredon@thalesaleniaspace.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_105

and even meta-materials, associated with new manufacturing techniques, will give the designer a renewed freedom to design more powerful and innovative space systems.

Keywords

Advanced manufacturing technologies • Advanced materials • Nanocomposites • Ceramics • Thermoplastics • Friction-stir welding (FSW) • Additive manufacturing (AM) • 3D printing • Micro-lattice structures

Introduction

The space manufacturing industry traditionally implements new technologies with a substantial delay when compared to other industrial sectors, due to the extreme caution imposed by the quasi-impossibility of repair in orbit, associated with the harsh environments of space missions. The necessity of implementing very thorough (and expensive) qualification processes, along with a strong predilection of mission managers for “proven flight heritage” equipment, makes it usually quite difficult to embark breakthrough technologies.

This trend may change with a new flow of “advanced manufacturing technologies” and techniques. Several innovative manufacturing technologies, new materials recently introduced, or that now become within our reach, seem well suited to the specific constraints of space hardware development. These may provide so many benefits in terms of performance, cost, or schedule that the compulsory development logic based on “proven flight heritage” may be challenged for the benefit of more efficient future space missions.

In addition, these emerging technologies may bring such promising capabilities that they may change the way space systems are conceived, designed, and developed.

This is particularly true for 3D printing technology, remarkably well adapted to the combination of high technology, specific requirements, and low volume typical of most space hardware development. 3D printing can also be a fantastic tool for fast prototyping and early validation. It may also prove to be an efficient tool for the necessary reduction of the environmental footprint of future space systems.

This chapter will also introduce some of the new emerging technologies that seem promising for upcoming satellites, explaining how they will contribute to improved spacecraft performance, increased flexibility, and more competitive development schedules.

A particular emphasis will be given to additive manufacturing technology, commonly called 3D printing, and how it can bring multiple advantages to spacecraft design and development.

Advanced Manufacturing Technologies

Introduction

A series of recent and significant innovations in materials, manufacturing, and processes is rapidly pushing its way into the space industry by challenging traditional manufacturing techniques, engineering methods and tools, and even mission design concepts. It is therefore important to know about them, to stay informed of associated latest developments, and, most importantly, to understand how these technology breakthroughs will change the way we think of, and the way we will design, future space systems.

It should be also noted that there is very active research in this field which should be constantly monitored by systems engineers, as new innovative technologies could challenge even further the design of spacecraft and make space missions more ambitious and affordable.

Advanced Materials

The first feature of new manufacturing technologies is the appearance of innovative and high- performance materials, with properties that change traditional structure design and dimensioning. These must now be taken into account when designing a new space system.

Currently, the most promising developments for structural materials are nano-based composites and ceramics. Nano-based composites now have electrical and thermal properties traditionally specific to metallic structures and can therefore be considered as a replacement when low mass or high stability is required.

Concerning inorganic materials, structural ceramics are now increasingly used, mainly because of their insensitivity to moisture, but there are still further improvement studies needed for dimensionally stable applications. Finally, the introduction of thermoplastic technology for space structures, derived from aeronautics, is important to mention as being very promising, even if specific developments are still necessary to cope with space requirements.

Advanced Composite Material

The thermal conductivity of standard composite material (carbon fiber-reinforced polymer) is driven by fiber reinforcement. So far, the use of high-modulus pitch-based carbon fibers is the only solution available to reach an in-plane thermal conductivity close to aluminum. Various products using this technology have been studied and developed worldwide for space structures where lightweight and high-stability structures are needed.

For example, such fibers which comply with the stringent pointing requirements of the Herschel satellite have been used to solve severe thermal stability issues on the Star Trackers support panel. To assure uniform temperature distribution on the structure, and a low thermal expansion coefficient, a sandwich construction with honeycomb core and high-conductivity composite skins was adopted. The K1100 pitch fiber, which exhibits excellent thermal performance (about 1000 W/m/°C along the fiber), was used permitting minimization of temperature gradients on the support while reducing the panel thermal expansion coefficient to nearly zero.

Unfortunately such fibers do not yet provide the same out-of-plane conductivity as aluminum, and they appear to be both expensive and fragile, which affects their potential benefit at satellite level. However, emerging solutions based on nanotechnologies pave the way for alternative solutions based on vertically aligned carbon nanotubes or graphene. Such solutions enable a combination of in-plane and out-of-plane thermal conductivities, leading to more efficient thermal management. Innovative lightweight radiator panels are therefore being developed based on these promising technologies.

Structural Ceramics

For several years, there has been a strong demand for the use of ceramics as structural materials in place of metals and alloys for use in severe environments. Consequently, structures made of new ceramics such as nitrides, carbides, and other covalently bonded materials have emerged.

Historically, silicon carbide (SiC) was the first ceramic material for manufacturing space structures. Significant developments have been made toward procuring SiC powders with the desired properties, through hot pressing, hot isostatic pressing, or standard sintering process. Because of its high thermal conductivity combined with a low coefficient of thermal expansion, this material is a good candidate for space structures requiring a high dimensional stability (intrinsic stability combined with low sensitivity to thermal gradients).

Silicon nitride (Si_3N_4) ceramic material is usually used for high-temperature applications (gas turbines, furnace equipment, etc.) as it shows excellent thermomechanical performances together with refractory property. This unique combination of mechanical strength, low coefficient of thermal expansion, and no moisture expansion (as inorganic material) makes silicon nitride a good candidate for highly dimensionally stable structures for space instruments.

Silicon nitride has now been qualified and successfully used for space applications. The standard grade for gas pressure sintering is made of 90 % Si_3N_4, 6 % Al_2O_3, and 4 % Y_2O_3 and sintered at 1 MPa with nitrogen gas pressure. The resulting material offers a very high Weibull modulus (up to 20) and high strength reliability, which enables to size structures with a very low occurrence of failure, even for highly loaded parts, making this material far less fragile than carbides.

But the most sensitive parameter for large optical components, for example, is the thermal expansion coefficient. At room temperature, silicon nitride has a very low coefficient of thermal expansion (1.4 10–6 m/mK), making it one of the lowest available values. This property adds with low density, high stiffness, strength, and long-term stability. Around 150 K, the coefficient of thermal expansion is near 0 and is 0.9 10–6 m/mK at 235 K, which is quite attractive for space instruments.

Full silicon nitride structures, like beams for truss structures or inserts, have also been developed and qualified. Assembly can be advantageously performed with bolts or through a dedicated brazing process. Future generations of silicon nitride will address the need to improve thermal conductivity, the only weak point of this unique material.

As a multiphase material, with large content of additives, it is also possible to enhance Si_3N_4's properties for specific applications by adjusting the type of additive selected originally for high-temperature application:

- Modification of sintering additives
- Replacement of standard powders with higher-grade Si_3N_4 nanopowders for higher thermal conductivity and lower sintering temperature
- Addition of high-conductivity particles, like silicon carbide or aluminum nitride or carbon nanotubes, cubic boron nitride, or diamond powder

Thermoplastics

Standard composite materials are currently used for satellite structures, principally to save mass. They are made of thermoset material, i.e., epoxy resin and continuous carbon fibers, and are cured in an autoclave or oven. The assembly of composite structures is based on inserts and involves several steps from drilling, insert positioning, and potting with epoxy adhesive up to polymerization.

Thermoplastic resin is different as it softens when heated and hardens again when cooled. There are many types of thermoplastics with varying crystalline organization and density. Some types commonly used in the car industry have been utilized in airplane design, which makes it possible to introduce them on spacecraft structures. The thermoplastic technology offers great qualities for space structure design such as:

- Storage procedure simplification, as compared to thermoset materials which require a storage at −18 °C and have a limited lifetime (no constraint for thermoplastics)
- Processing outside the autoclave, requiring less time and cost
- A welding technique for joining (which results in higher-temperature resistance and less sensitivity to moisture aging in a more cost-efficient way)

The thermoplastic technology is already mature for the manufacturing of aeronautical parts, and huge efforts in R&D have been made. It is also necessary to address the high stiffness requirement specific to space structure applications. When melted, the thermoplastic matrix has a higher viscosity than the uncured thermoset resin, which may yield to difficulties with impregnation of the carbon reinforcement. This is a real challenge for ultrahigh modulus carbon fibers, still being studied, which will bring future innovations.

Innovative Surfaces

Internal and external surfaces of satellites play an important role, essentially for thermal management, through dedicated thermo-optical properties: absorption from external sources, heat dissipation or transfer, and heat rejection into space. There are many developments concerning smart surfaces, which aim to propose high-performance alternatives to paint: coatings with enhanced thermo-optical properties or surface texturing.

Besides innovative black surfaces for thermal and/or optical management, it is worth mentioning that laser technologies also have promising applications for surface treatment. Called surface texturing, these technologies have very interesting applications for space. For example, lasers can be used to clean or etch surfaces, replacing liquid or mechanical surface treatment. More generally they can be used to add new functions to a surface for improved performances. The prevention of outgassing, particle contamination, as well as shorter manufacturing cycles is a great advantage for space hardware production.

More generally, surface texturing can improve bonding performances with less scattering, allowing the reduction of structural inserts, better corrosion resistance, and improved hardness. Industrial flexibility is improved thanks to the increased time allowed between surface preparation and bonding.

In addition, these mechanical surface treatments improve the environmental friendliness of the processes, avoiding the use of dangerous solvents such as hexavalent chromium or hydrofluoric acid.

In the field of optical observation, it is essential to master many parameters of the optical payload in order to guarantee the optical quality throughout the satellite's life. Stray light is one of the major factors to be taken into consideration for the instrument quality. It can affect both the geometric image quality and the radiometric image. For Earth imaging and for astronomy observation, it is very important to reduce the stray light level in the instrument. To do this, black coatings are usually implemented on all mechanical surfaces close to the optical beam. The reflectivity properties of the black coating usually have a substantial impact on the instrument architecture and mass.

For most space missions, black surfaces are based on paints or anodizing.

Paints are generally easy to apply but there are some limitations such as optical baffle design. It is difficult to guarantee a uniform thickness everywhere inside baffles. It is also quasi-impossible to paint certain sharp surfaces (edges of vanes, for instance). The problem is that paint thickness homogeneity is essential for reaching the desired optical properties.

Anodizing is a more expensive type of coating and is more often used for complex aluminum parts. For substrates containing copper (as aluminum allow 6061, often used in space), adhesion of the coating is not fully guaranteed, and some particles may be generated from the black surface, which may then cause particle contamination leading to stray light. Alternative solutions, dry or liquid (through a sole-gel process), are being investigated to combine high optical quality and process robustness.

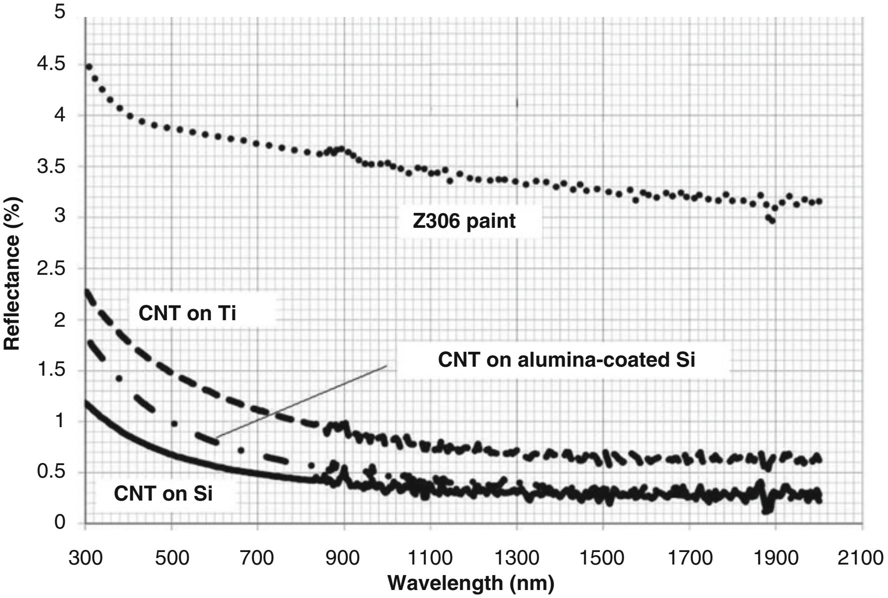


Fig. 1 Bidirectional/hemispherical reflectance of the three MWCNT samples and Z306 paint

Carbon nanotubes can now be used to reach optical performance (absorptivity). Thin films of vertically aligned carbon nanotubes made of CVD (CVD-VA-CNT) yield outstanding absorptivity performances. The bidirectional reflectance distribution function of grown CNT on different substrates was investigated and compared to Aeroglaze Z306TM paint which is widely used in space. The CNTs clearly outperform commercial black paints in terms of integrated absorptance as indicated in Fig. 1 (Butler et al. 2010). An exceptionally low stray light reflection (<10-4) was reported (Debra 2015) and the directional-hemispherical reflectance of this material was evaluated at 0.995–0.999 within the 400–1000 nm spectral range (Pambaguian et al. 2013; Burger et al. 2015).

The challenge now is to develop a robust process without sacrificing optical properties.

Robotics in Satellite Assembly and Test

The aerospace industry, and the satellite industry in particular, requires a highly qualified workforce, with skills that are obtained only after several years of training. To ensure competiveness, highly skilled workers must be focused on the value-creating aspects of the manufacturing process.

Integration of collaborative robots into advanced manufacturing systems enabled by information and communications technology (ICT) can be a key driver for industry growth, contributing to increased productivity, innovation capacity, and flexibility to react to market demands. The paradigm for robot usage has changed from a model where robots work with complete autonomy to a scenario in which robots collaborate with humans. This means utilizing the fortes of both humans and robots: the cognitive and dexterity capabilities of humans and a robot's capacity to do repetitive work and provide assistance.

It is thus foreseen that collaborative robots will support highly skilled operators, allowing the operators to focus on complex tasks with high value added while bringing great flexibility with low/no setup and change over time and better ergonomics.

Additive Manufacturing (3D Printing) for Space Missions

The 3D Printing Process

Additive manufacturing (AM), also known as 3D printing, refers to the various processes used to realize a three-dimensional object directly from a 3D file. These processes offer new freedom in the design of 3D objects. The 3D parts are manufactured layer by layer under computer control from a CAD file producing a raw material almost without any limitations for the shapes.

The main differences between the various processes are in the ways by which the layers are deposited to create parts and in the materials that are used. For example, AM parts can be processed from:

- Liquid light-polymerized resin (stereolithography (SLA))
- Powdered material (selective laser sintering (SLS) for polymers and laser beam melting (LBM) also known as selective laser melting (SLM) for metals, electron beam melting (EBM), direct metal deposition (DMD))
- Solid material (fused deposition modeling (FDM) for polymers and wire deposition for metals)

For spaceflight applications, powder bed processes are clearly the most suitable because they have been proven with aluminum and titanium and also because these processes are the most accurate, therefore suitable for complex geometries. Powder or wire deposition technologies are also fields of development, although less tested, and may be considered when size limitation of powder beds technologies is an issue.

For space applications, the combination of low series (number of parts between 1 and 100) and weight saving objectives makes AM an almost ideal process. AM and powder processes, in particular, allow for complex shapes at nearly no cost! Most of the time, metallic parts were manufactured from machined bulk material with computer numerical control and specific programming to reach a competitive weight (cavities and ribs).

Figure 2 gives an example of a simple equipment support made of a small panel, inserts, and a few aluminum brackets. The mass saving for the equivalent support

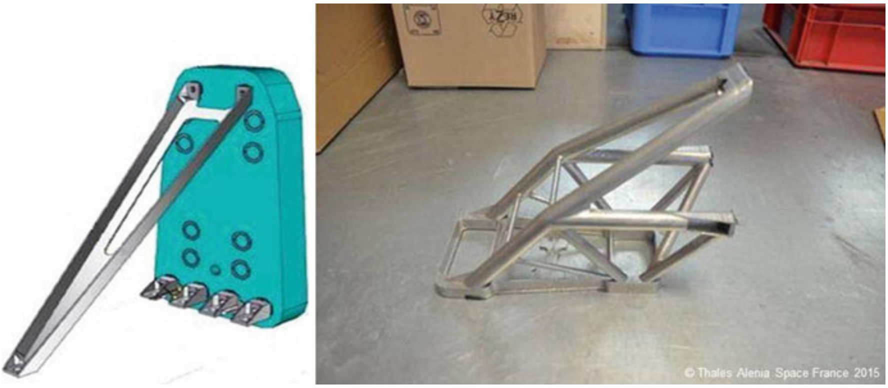


Fig. 2 Comparison of a simple satellite bracket in standard design (700 g) and its monolithic 3D-printed equivalent design (300 g)

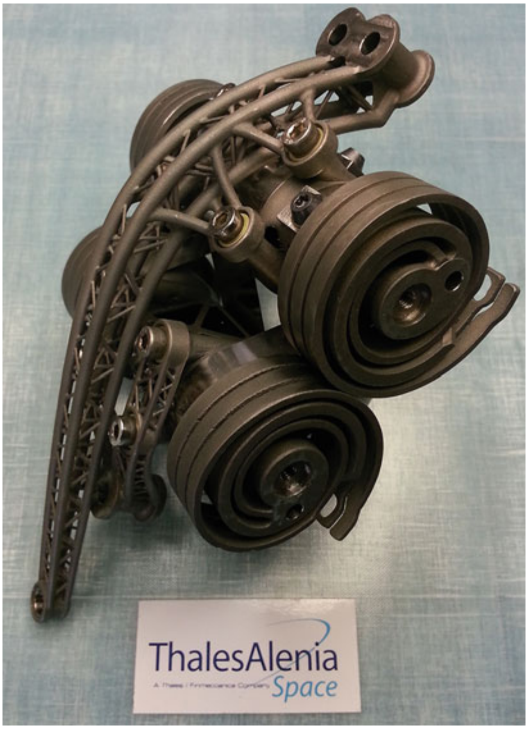


Fig. 3 Example of a complete mechanism realized with 3D printing

made with AM in monolithic aluminum is about 400 g for an original mass of 700 g. This simple equipment support has been flying on a telecom satellite since 2015.

Figure 3 is another example of a more complex solar array deployment mechanism, designed by Thales Alenia Space, which could not have been manufactured with conventional techniques. It uses substantially less mass than if it had been manufactured by any another method.

In some rare cases such as satellite constellations, when the notion of series can be enlarged to a few tens of units to typically up to 1000 parts, casting technologies are still a more competitive solution.

High costs and long schedules were also common because of the complex tools necessary to achieve a complex geometry. Additive technologies offer an alternative to subtractive techniques for producing very complex, low-weight shapes without tool investment.

Mass is obviously a major concern because of launching costs and has become an increasing concern for the atmospheric reentry of fragments large enough to become a potential hazard. The ESA Clean Space program or the French Space Law now imposes severe environmental considerations in order to limit the impact of end-of-life spacecraft atmospheric reentry of large debris on Earth. Metallic parts, especially those with high melting points such as titanium, and with those with high mass are specifically targeted. 3D printing may be particularly well suited to design structures that can completely burn up during controlled or uncontrolled atmospheric reentry when a mission is completed. AM is therefore an excellent tool to be considered when *design for demise* is an imperative.

Design Tools

Design optimization tools will multiply enormously the potential of additive manufacturing for weight reduction. Specific design tools for weight optimization can be adapted to all manufacturing technologies, but they find an incomparable playground with AM, thanks to the extraordinary freedom offered by powder bed technologies.

The software market now proposes design tools especially relevant with AM because they lead to complex shapes with evolving sections and curves that other processes cannot easily produce. The best examples of optimized designs for a function are trees and skeletons! Nature is a precursor in terms of weight-optimized design. A bone is a wonderful example of shape optimization for a specific function, with great mechanical resistance and light weight, thanks to a solid external wall and internal trabecular structure. AM is a tool that will allow us to design such structures.

Topological optimization software tools are now starting to be commonly used for mass optimization of space hardware. Starting from the volume allowed for the part, possibly including forbidden zones (for instance, for cables routing or close part arrangement) as well as precisely defined interfaces and mechanical loads, the topological software will remove all useless material keeping only that which the function requires. Resulting geometries need to be smoothed by dedicated drawing software tools and the final design must be of course validated by traditional FEM analysis. An example of the design steps for a simple equipment unit is given in Figure 4.

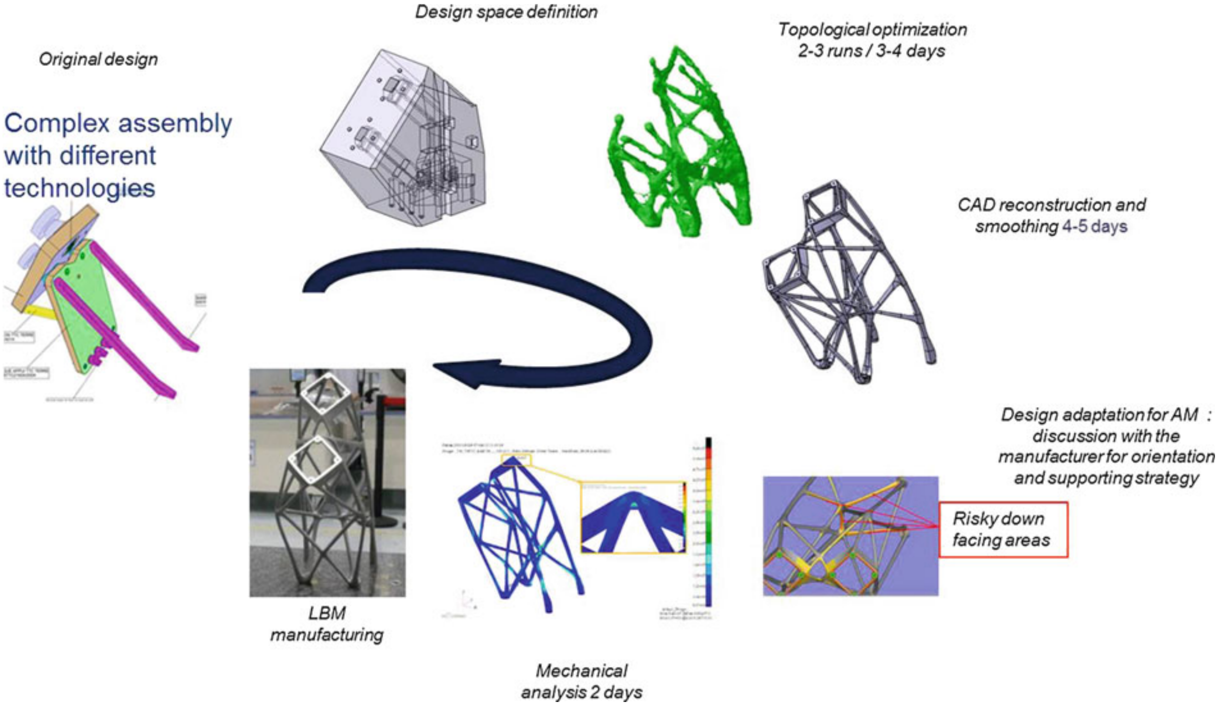


Fig. 4 Typical 3D design optimization cycle for a simple equipment unit

Such design methods lead to weight reduction of 20–50 % when compared to “traditional” design. Even higher weight savings are reached when several functions are merged into a single part. Such design work must be done in tight collaboration between design and manufacturing teams, in order to define together the optimum position in the machine, and the best supporting strategy to achieve accuracy, roughness, or non-distortion requirements.

The powder bed processes offer great freedom to designers but they also have constraints, especially for metallic materials. For example, bars with angles of less than 45°, referring to the building plate, have to be supported to avoid collapsing.

In addition, evacuation of non-melted powders, as well as posttreatment accessibility, must be taken into account. A design well suited for AM will use less powder, less building time, will limit manual operations, and avoid many difficulties during subsequent manufacturing steps.

Post-machining is most often required to reach perfect accuracy for the interfaces. This step can be quite challenging if the part is complex, with unusual or organic shapes.

Panorama of Current Applications

The use of AM for space applications can now be considered mature for low mechanical load parts such as secondary structures, even if the low AM heritage still requires very secured and costly validation. However, the great possibilities of this high-potential technology have imposed its acceptance, and some metallic parts are already now in orbit. Most of these are made of titanium or aluminum, preferred for space lightweight structures and commonly used for satellites

Satellite telecom payloads, for example, the antennas, are specifically designed for a particular mission, corresponding to a service zone coverage. As a result, most of the corresponding structural and RF components are very specific, usually complex in terms of design, and expensive to produce. 3D printing is clearly a high-potential solution, allowing the manufacture of very optimized designs without current development constraints. The metallic structural parts of antennas, currently produced via conventional machining and electrical discharge machining (EDM), using aluminum and titanium, are prime candidates for AM use, with topology optimization of the mass/stiffness ratio. Figures 5 and 6 show examples of such small structures in orbit since 2015.

As already mentioned, AM will offer maximum mass saving and design benefits when used in association with shape optimization, which is what is increasingly done. We can safely predict that in the coming years, there will be an enormous increase in 3D printing use for spaceflight hardware, small MEO satellites, and even larger GEO telecom satellites.

Some other materials are being considered for specific applications such as those requiring high-temperature resistance. Promising on-ground testing was already performed for the combustion chamber of 10 N bipropellant thrusters,

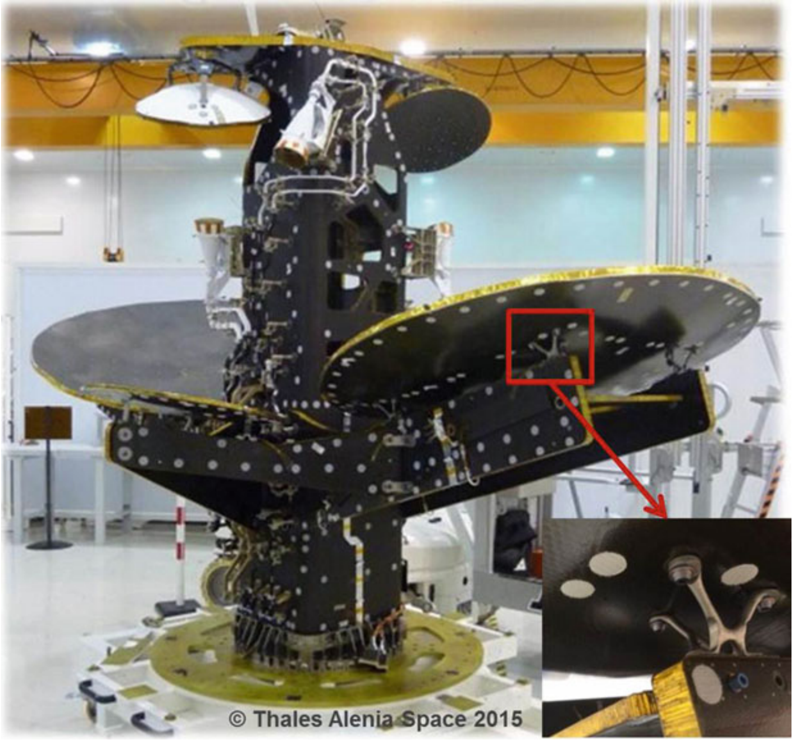


Fig. 5 Example of AM titanium reflector fitting

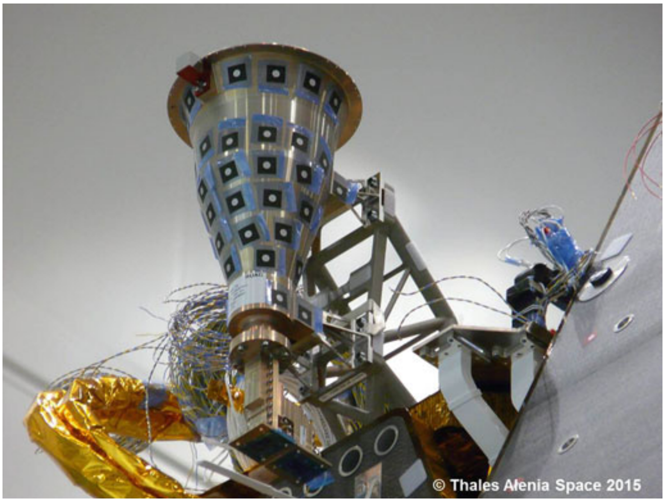


Fig. 6 Example of AM aluminum Ku Horn support

printed in platinum-rhodium alloy using a laser beam-powder bed process. The aim was to demonstrate the feasibility of printing platinum alloys for chambers and nozzles as well as to reduce recurring costs in production and the cost of recycling noble metals. This thruster from Airbus Defense and Space with a 3D-printed combustion chamber and nozzle was successfully test-fired by ESA in 2016.

Aerojet Rocketdyne also recently announced a series of successful hot-fire tests of its RL10 upper-stage rocket development engine, including a core main injector also built with SLM 3D printing. This was part of a USAF/NASA program aimed at demonstrating the capability of additively manufactured complex parts and their qualification for large rocket engines.

CubeSats containing 3D-printed polymer parts, often designed by students, have already been launched. High-temperature range polymers must be used in order to sustain space environment constraints.

Printing in space is also becoming a reality. In the USA, the company Made in Space built a 3D printer for polymers in the ISS, and a similar wire-based 3D printer developed in Europe for polymers (25 cm, 5.5 kg) was launched with the Cygnus resupply vessel toward the ISS (POP 3D) with the objective of giving the ISS crew the capability to build spare or repair parts, or specific tools on demand.

Rapid Prototyping and System Impacts

The interest of AM for the space industry is not limited to flight applications. Mock-ups and prototypes made with polymers by SLA or SLS can provide extremely useful replacements for 3D models in the design phase or used to define complex assembly procedures (Fig. 7).

Assembly, integration, and test facilities are now often equipped with polymer wire or powder-based machines for specific tooling manufacturing. The use of dummies, replacements for real parts coming later in the assembly sequence, or protection structures is now common.

Materials must, however, be carefully selected for specific uses. For example, an equipment destined for use in a thermal vacuum chamber must be considered for outgassing risk of pollution, mechanical behavior not suitable with high temperatures, etc.

Additive manufacturing is not only a new production process; it is also a revolution in the way satellites are designed. Collaborative work, co-engineering design cannot only achieve savings on a single part but will allow for the merging of different parts or even different functions. Satellite architecture can be optimized in the future to generate greater system impact (performance increase and/or weight saving mainly). Complex subsystems, conceived and optimized for their specific functions, will appear and generate cost and schedule savings beyond the 20–30 %, already achieved for single parts (Fig. 8).

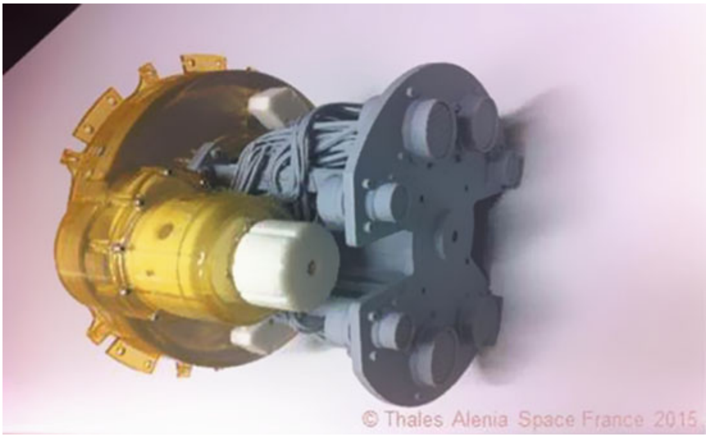


Fig. 7 Example of a multi-material mock-up realized with 3D printing

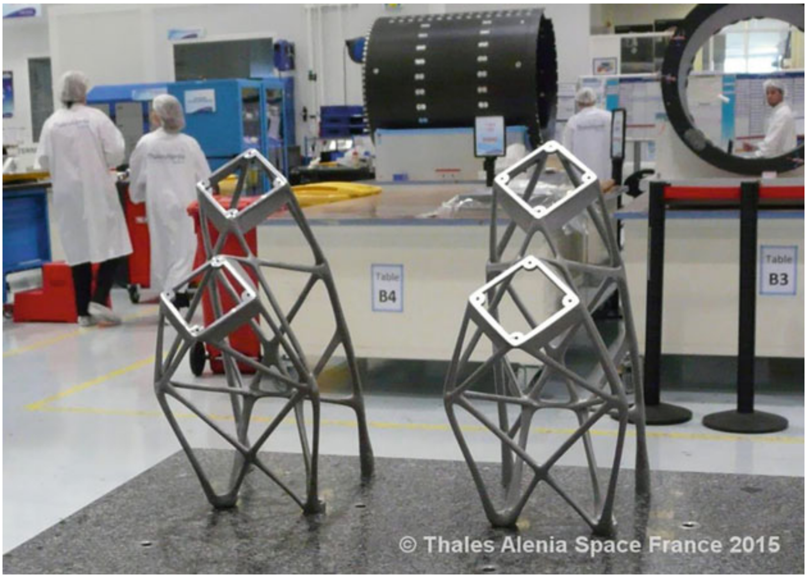


Fig. 8 Aluminum LBM parts ready for flight

Qualification Issues

AM parts validation can be achieved through:

- Acceptance testing of parts in addition to witness samples with destructive testing
- Repeatability demonstration of the process though a multi-batch approach followed by tests on selected samples

AM processes are very close to casting or welding because they involve the melting of material and because qualification status can be obtained only for each material, associated with the corresponding machine and supplier.

The other similarity is the type of defects that can be found on AM-produced parts: porosities, cracks, etc. The validation approach should then be similar to those used for parts produced by traditional processes, meaning a combination of material health inspections and functional testing. The quality of the design and the manufacturer's know-how (part orientation in the chamber, supporting strategy, post-processing, etc.) is a major factor of the quality of the result.

Space applications are very demanding in terms of quality because of the extreme environment and because repair opportunities are very rare. However, compared to aeronautics, the maturation of technology for space applications is much more rapid because the question of fatigue is simplified. It is usually not a major issue for typical satellite applications.

The new challenge for manufacturing is to build larger parts maintaining the necessary high quality level, all the while controlling distortions risks.

In conclusion, there is nothing that prevents this technology from being massively implemented on spacecraft, provided this is done step by step with a controlled risk approach.

Future Capabilities

Addressing larger parts will probably lead to studying new additive processes based on material deposition (powder or wire) to be post-machined and potentially combined with other technologies such as LBM and/or traditional welding. Many patents for these technologies are currently coming into the public domain, which will soon provide a larger market offering.

New function integration may also be a path toward obtaining greater benefits from this technology: equipment or platform thermal control and RF functions could be included in the mechanical support itself.

The implementation of lattice structures will probably be one of the challenges in the coming years, considering the potential mechanical or thermal performance benefits in addition to the possibility of extreme weight reduction.

These ambitious developments are only limited by the capacities of design optimization, easy drawing, and modeling. Such software tools are not yet currently available on the market.

In the future, many parts will be manufactured with AM. The aeronautics sector has already started to produce series with this technology, and this will happen also in satellite manufacturing. Richard Ambrose, executive vice president of Denver-based Lockheed Martin Space Systems declared in 2014, about the A2100 Space Bus: “My goal is to have over 50 % of the structure 3D-printed within 2–3 years.”

Another dream might also become reality: building satellites or spacecraft directly in space, avoiding some launching costs, or even using local material to build on the moon. A lunar base could be built with a material deposition machine,

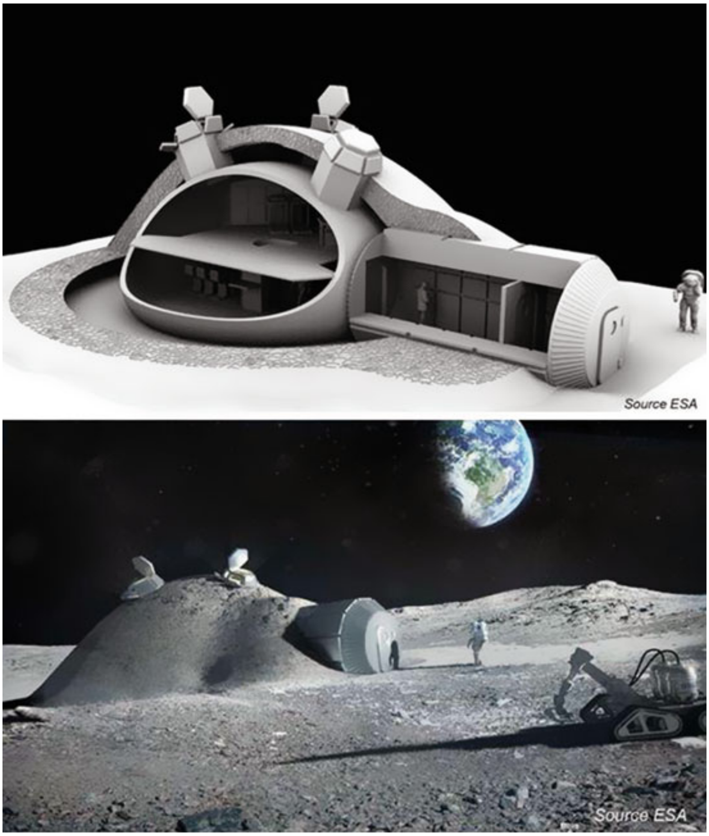


Fig. 9 ESA lunar base project equipped with a protection shield 3D printed with lunar sand

using resin from Earth and lunar sand. ESA has analyzed the possibility of setting up habitat infrastructures with robots using lunar soil as 3D printing material (Fig. 9).

This amazing technology will certainly change our future vision of space missions and may even be the beginning of a new era of human and robotic exploration.

Conclusion

There have been many changes in the approach to the manufacturing of satellites in just the past few years. There have been many innovations in materials, technologies, and processes. In particular advanced manufacturing has moved ahead rapidly due to advanced materials, the use of robotics in both assembly and testing, 3D printing that soon may be at least half of the manufacturing process, rapid prototyping, and advanced artificially intelligent design tools.

These processes cannot only lead to more cost-efficient design and manufacture of satellites but can be very effectively deployed when much larger-scale production is involved by large-scale satellite constellations that can require hundreds of spacecraft of a similar design to be built and deployed in orbit. Significant progress in all these areas have been achieved in only a few years and many more advances are now anticipated in satellite design, manufacturing, and testing. In time these techniques could even be applied to materials processing in space with satellites built not on the ground but in outer space.

Cross-References

▶ Common Elements versus Unique Requirements in Various Types of Satellite Application Systems
▶ Overview of the Spacecraft Bus

References

N. Burger, A. Laachachi, B. Mortazavi, M. Ferriol, M. Lutz, V. Toniazzo, D. Ruch, Alignments and network of graphite fillers to improve thermal conductivity of epoxy-based composites. Int. J. Heat Mass Transf. **89**, 505–513 (2015)

J.J. Butler, G.T. Georgiev, J.L. Tveekrem, M. Quijada, S. Getty, J.G. Hagopian, Initial studies of the bidirectional reflectance distribution function of carbon nanotube structures for stray light control applications. Proc. SPIE **7862**, 78620D-2 (2010)

L. Cornillon, C. Devilliers, S. Behar-Lafenêtre, S. Ait-Zaid, K. Berroth, A.C. Bravo, Silicon Nitride Ceramic Development in Thales Alenia Space: Qualification Achievement and Further Development for Future Applications. International Conference on Space Optics (2014)

W. Debra, Reimagining satellite construction. Aerosp. Am. **53**, 20–25 (2015)

C. Devilliers, T. Lasic, D. Boban, L. Cornillon, D. Tanzilli, S. Aitzaid, K. Berroth, Si_3N_4 ceramic application for large telescope development results. International Conference on Space Optics (2014)

GSTP-6 Element 1 compendium of potential activities advanced manufacturing. ESA (2015)

T. Middelmann, A. Walkov, R. Schödel, State-of-the-art cryogenic CTE measurements of ultra-low thermal expansion materials. SPIE **9574** (2015)

L. Pambaguian, B. Bonvoisin, T. Ghidini, Additive manufacturing technologies. European Technology Harmonisation Technical Dossier, ESA/THAG (2013)

N.T. Panagiotopoulos, E.K. Diamanti, L.E. Koutsokeras, M. Baikousi, E. Kordatos, T.E. Matikas et al., Nanocomposite catalysts producing durable, super-black carbon nanotube systems: applications in solar thermal harvesting. ACS Nano **6**, 10475–10485 (2012)

N. Selvakumar, S.B. Krupanidhi, H.C. Barshilia, Carbon nanotube-based tandem absorber with tunable spectral selectivity: transition from near-perfect blackbody absorber to solar selective absorber. Adv. Mater. **26**, 2552–2557 (2014)

W.O. Soboyejo, J.D. Obayemi, E. Annan, E.K. Ampaw, L. Daniels, N. Rahbar, Review of high temperature ceramics for Aerospace applications. Adv. Mat. Res. **1132** (2015)

X.J. Wang, O.S. Adewuyi, L.P. Wang, B.A. Cola, and Z.M. Zhang, Reflectance Measurements for Black Absorbers Made of Vertically Aligned Carbon Nanotubes. in *Conference on Reflection, Scattering, and Diffraction from Surfaces II*, San Diego (2010)

Wohler Report, 3D printing and additive manufacturing State of the Industry Annual Worldwide Progress Report, Wohlers Associates (2015)

Tracking of Orbital Debris and Avoidance of Satellite Collisions

Joseph N. Pelton

Contents

Abstract

The issue of space debris has become one of increasing concern as the amount of orbital debris, sometimes known as "space junk," has become more severe, especially in low Earth orbit and in the polar orbits used for communications, remote sensing, and meteorological sensing and forecasting. The Chinese missile shootdown of the defunct Fung-yen (FY-1C) weather satellite in 2007 and the collision of the Iridium and Cosmos satellites in 2009 have greatly heightened this concern. Increasingly sophisticated tracking systems have been implemented by the US Air Force Strategic Command, the European Space Agency, and several affiliated national tracking systems to cope with the complex space situational awareness (SSA) challenge that is now presented by rising amount of space debris. A new S-band radar "Space Fence" system and other optical tracking systems in Australia and other parts of the world are being implemented to cope with this task.

J.N. Pelton (✉)
International Space University Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_106

The new S-band Space Fence system that is currently being installed in the Kwajalein Atoll in the Pacific, in particular, will allow an increase in the tracking ability for space debris. This increased tracking ability will thus rise from about 23,000 debris elements that are 10 cm or larger (i.e., about the size of a baseball) in low Earth orbit to well over 200,000 elements that are greater than 1 cm in diameter (i.e., about the size of a marble) in low Earth orbit. The Space Data Association, which has been formed by commercial satellite operators, is also increasingly able to share information among themselves in order to minimize the possibility of collisions and to be aware of close satellite conjunctions in a timely manner. Their tracking capabilities are currently provided by a commercial capability operated by Analytic Graphics Inc. (AGI).

In addition new laws and national regulations as well as guidelines adopted by the Inter-Agency Space Debris Coordination Committee (IADC) and the UN Committee on the Peaceful Uses of Outer Space (COPUOS) to ensure that all satellites are deorbited within 25 years at the end of a spacecraft life represent another key step forward. There are clearly more steps that need to be taken to move toward better collision avoidance systems plus active deorbit and debris mitigation, especially of the largest debris elements from low Earth orbit. It is also key to ensure that the deployment of new large-scale constellations in low Earth orbit is accomplished with strict controls to minimize any new collisions that might occur within these constellations themselves or to avoid collision with defunct debris elements. The addition of constellations with perhaps a thousand small spacecraft or more in just one constellation has given rise to particular new concerns in this regard.

In addition, there needs to be (i) new and better international collaboration to strengthen all elements associated with the more precise tracking of debris in all Earth orbits; (ii) more control processes to prevent debris increase and avoid the formation of new debris elements, including the active deorbit of all launch systems after they have inserted spacecraft into orbit; (iii) better coordination of information among satellite system operators through such mechanisms as the Space Data Association as its membership and participation levels grow; and (iv) new technology and international agreements and perhaps commercial arrangements to incentivize the active deorbit of space debris in future years consistent with existing space treaties and international agreements.

This chapter addresses in some detail the various tracking capabilities that exist or are planned around the world to monitor the orbits of space debris and to provide alerts so as to avert possible conjunctions. It provides information about how these systems are being upgraded and space situational awareness (SSA) capabilities are being coordinated over time. It notes how governmental systems are being augmented by private capabilities that are able to augment space situational awareness and to assist with avoidance of collisions. These systems and processes will perhaps assist with future space debris mitigation and active removal. All of these increasing space situational awareness capabilities are crucial to the future successful operation of application satellites in the twenty-first century.

Keywords

Active space debris removal • Analytic Graphics Inc. (AGI) • European Space Agency (ESA) • EISCAT • Image Information Processing Center and Supercomputing Facility (IIPCSF) • NASA • Optical tracking • Space Data Association (SDA) • Space situational awareness • S-band radar Space Fence • Satellite conjunctions • SPACETRACK • Tracking and Imaging Radar (TIRA) of Germany • UN Committee on the Peaceful Uses of Outer Space (COPUOS) • US Strategic Command (USSTRATCOM) • US Space Surveillance Network • Working Group on the Long-Term Sustainability of Outer Space Activities

Introduction

At the dawn of the space age, a half century ago, the idea that human-manufactured orbital space debris would be a major concern to commercial organizations operating networks of communication satellites, remote sensing networks, navigation satellites, and meteorological satellites was almost unthinkable. Few of these space systems even existed, and the vast reach of outer space was truly enormous. Just the volume of space that surrounds Earth out to geosynchronous orbit represents an astonishingly large 300,000,000,000,000,000 (3×10^{17}cubic kilometers). The space around our planet is a quite vast neighborhood for satellites to populate. And at the outset satellites were quite small and compact – the size of beach balls. But spacecraft and rockets became larger and larger and, more and more satellites were launched.

A lack of care was taken about explosive bolts, upper stage rockets left in orbit, and satellites were launched and then deserted year after year. There were no rules about deorbiting defunct satellites at the end of life. Each year the amount of orbital debris increased, and the situation that Dr. Donald Kessler of NASA warned about back in the 1980s – that of substantial debris buildup that could cascade out of control – is no longer a concern to be ignored but a matter of serious concern.

On January 11, 2007, China conducted a now widely publicized antisatellite missile test to shoot down the defunct Chinese weather satellite, the FY-1C polar orbiting satellite of the Fengyun series. This 750 kg satellite was hit at an altitude of 865 km (537 mi) and was instantly splintered into over 2000 trackable space debris elements. The so-called kill missile was traveling in the opposite direction of the satellite at a speed of 8 km second (or 28,800 km/h).

Then 2 years later, there was a collision of the Iridium 33 and defunct Kosmos-2251 satellite at a relative velocity of 42,000 km/h that occurred on February 20, 2009, at 16:56 Universal Time Coordinates (UTC). This violent intersection also created over 2000 new debris elements. Today the space debris problem continues to increase. According to Dr. Kessler who first predicted the “Kessler syndrome” and the possibility of an ever increasing cascade of space junk, there is now a “likelihood” of a major space collision every 10 years. He also explains that the cascade effect that is now occurring will create more and more debris elements

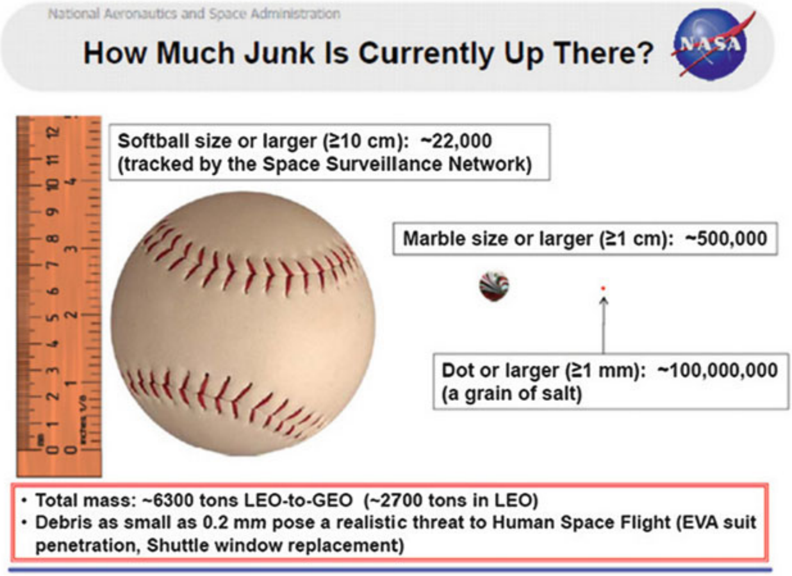


Fig. 1 The NASA Score Sheet Statistics on Space Debris (Graphic courtesy of NASA)

even with no additional launches – and of course we are planning an ever increasing amount. So what are we to do to preserve space operations for the future?

These two major collision events have underlined the dangers and demonstrated the increasing difficulty of controlling the space debris problem particularly in low Earth orbit and the congested polar orbits. The nearly 3000 metric tons of debris in low Earth orbit are thus of an increasing concern (see Fig. 1).

The US Space Surveillance Network detects, tracks, catalogs, and identifies artificial objects orbiting Earth, i.e., both active/inactive satellites, spent rocket bodies, and fragment debris. The system is the responsibility of the Joint Functional Component Command for Space, a part of the US Strategic Command (USSTRATCOM).

The current US standard for protecting astronaut-occupied spacecraft is to maneuver to avoid an object if it is calculated to have a higher than 1:10,000 chance of hitting the asset. The objective is to create a 200 km buffer zone or “bubble of protection” around the International Space Station, for instance. This provides less than 30 s of separation and reaction time between the ISS and crossing orbital debris (Michael Cooney).

The US Space Surveillance Network (SSN) has a specified set of strategic objectives as defined by the US Congress. These explicit duties include:

- Predicting when and where a decaying space object will reenter the Earth’s atmosphere

- Preventing a returning space object, which to radar looks like a missile, from triggering a false alarm in missile-attack warning sensors of the US and other countries
- Charting the present position of space objects and plotting their anticipated orbital paths
- Detecting new man-made objects in space
- Correctly mapping objects traveling in the Earth's orbit
- Producing a running catalog of man-made space objects
- Determining which country owns a reentering space object
- Informing NASA whether or not objects may interfere with satellites and International Space Station orbits

The SPACETRACK program represents a worldwide Space Surveillance Network (SSN). This is a complex network of sensing devices that now includes electro-optical, passive radio-frequency (RF), and radar sensors. The SSN is largely composed of elements owned and operated by governmental agencies, but increasingly its capabilities are augmented by instruments owned and operated by private commercial firms as well. The SSN is tasked to provide space object cataloging and identification, satellite attack warning, timely notification to US forces of satellite flyover, space treaty monitoring, and scientific and technical intelligence gathering. This is a unique combination of tasks that are civilian space activities and US defense and military duties (US Space Surveillance Network).

The continued increase in satellite and orbital debris populations, as well as the increasing diversity in launch trajectories, nonstandard orbits, and more and more satellites – including small, micro-, CubeSats, and so-called Femto (or microsatellites), has made the task of monitoring the skies more and more difficult. This has led to the need to upgrade the SSN to meet existing and future requirements. It has also led to efforts to ensure the cost-effective operation of the SSN through automation where possible.

The prime area of upgrade for the SSN is the near-term creation of the new S-band Space Fence. On June 3, 2014, Lockheed Martin division won a $914 million contract to build this new Space Fence radar system in the Pacific Ocean area along the Kwajalein Atoll in the Marshall Islands. It will serve as the next-generation space surveillance radar system. This nearly $1 billion contract covers the cost of the engineering, manufacturing and development, production, and deployment of the S-band Space Fence. The currently planned initial operational capability date for the installation is in 2018 (Lockheed Martin 2014).

When deployed the S-band ground-based radars will be designed to detect, track, and measure objects in space, mostly especially in low Earth orbit, although it will be able to track larger objects in higher orbits as well. The new radars for the Space Fence program will be able to detect much smaller microsatellites and debris than current systems and speed up detection of possible threats to GPS satellites or the International Space Station or communications and surveillance satellites.

The geographic separation along the Kwajalein Atoll plus the higher wave frequency of the new radar system will allow for the detection of much smaller microsatellites and debris than current systems – down to the size of a marble. It will significantly improve the timeliness with which operators can detect potential threats to GPS satellites or the International Space Station or other space assets and infrastructure (Space Fence 2015).

SPACETRACK in addition to its space situational awareness and debris tracking and collision avoidance capabilities has a clear military-related function. In particular, SPACETRACK has been assigned responsibility to develop the systems interfaces necessary for the command and control, targeting, and damage assessment associated with any potential future US antisatellite weapon (ASAT) system capability. Part of SPACETRACK's capabilities is an image information processing center and supercomputing facility at the Air Force Maui Optical Station (AMOS).

Currently information from the SSN is provided to governmental operators under a memorandum of understanding (MOU) to aid with the avoidance of orbital collision. The sensitivity of the SSN and SPACETRACK's military- and defense-related function was one of the factors that led to the creation of the Space Data Association (SDA) that functions from the Isle of Man and supports satellite operators from around the world. The prime capability of the SDA is to anticipate and help prevent conjunctions of GEO-based operational satellites by operators sharing data about the orbital locations of their satellites. Nevertheless it is also increasing its capabilities to anticipate potential conjunctions in other orbits as well.

The Space Data Association and the Analytic Graphics Inc. Tracking Network

The SDA was formed in 2009 by Inmarsat, Intelsat, and SES to share data. In April 2010, Analytical Graphics, Inc. (AGI) won the contract to design and operate the Space Data Center, SDA's automated space situational awareness system designed to reduce the risks of on-orbit collisions and radio-frequency interference. Initial Space Data Center operations began in July, and full capabilities were online April 2011. The current data base is constructed using AGI's commercial software and is increasingly more capable with data being fed into the system by the member organizations of the Space Data Association (SDA). The data center then provides SDA members networked access to operational capabilities through a service-oriented architecture. The Space Data Center automatically ingests and processes operator-orbital data, performs conjunction assessments, generates automated warning alerts, and supports avoidance maneuver planning and efficient RFI mitigation. It is an enhancement of AGI's SOCRATES-GEO/LEO system (Space Data Association).

Analytical Graphics Inc. is now capable of tracking thousands of space objects with its Commercial Space Operations Center, or ComSpOC, which relies on optical and radio tracking assets and the company's own space surveillance software. AGI is building a catalog of space objects that it calls the SpaceBook (AGI-Lockheed):

Lockheed, in observing AGI's success in providing key services to SDA on an on-going commercial basis, has recently begun branching out to develop optical tracking capabilities in partnership with Australia's Electro-Optical Systems.

Lockheed Martin and Optical Tracking Network

Lockheed Martin Space Systems, working with Australia's Electro Optic Systems, announced on August 25, 2014, that it is planning a new space object-tracking site in Western Australia and hopes to sell the data to the US and Australian governments (http://spacenews.com/41727agi-lockheed-tout-commercial-space-surveillance-systems/#sthash.5CzBha2C.dpuf).

In announcing the new agreement with Electro Optic Systems of Australia, Lockheed Martin clearly indicated that they were, in fact, entering the space situational business. The official announcement said: "Through this agreement with Electro Optic Systems, we'll offer customers a clearer picture of the objects that could endanger their satellites, and do so with great precision and cost-effectiveness." The announced specific objective in using the EOS capabilities in Australia will be to zoom in on specific pieces of debris and determine their content, spin direction, and orbital speed. The data will be used to determine how much of a threat a given piece of debris poses to operating satellites.

The use of optical sensors to track space debris is becoming more and more common. Currently, the ComSpOC has 20 optical sensors and three radio-frequency sensors in operation. ExoAnalytic Solutions of Mission Viejo, California, and the Las Cumbres Observatory Global Telescope Network of Goleta, California, are providing optical sensors. Rincon Research Corp. of Tucson, Arizona, provides data from radio-frequency radar (AGI-Lockheed).

This optical capability is currently seen by US Space Surveillance Network officials as a useful complement to the S-band space tracking radar being located on the Kwajalein Atoll in the Pacific Ocean, near the equator. It has further been indicated that the future planning might lead to a possible second Space Fence site to be located in Western Australia.

Nor is the use of ground-based radar and optical telescopes the only available tools, the US Air Force has also launched tracking satellites into orbit to augment its capabilities on the ground. One of these satellites used for space situational awareness, including the tracking of potential missile threats, is the US Air Force Satellite shown in Fig. 2.

Thus to summarize the situation about current and future space debris tracking capabilities, the following is generally the case. Space surveillance networks are largely limited to larger objects, typically greater than 10 cm in low Earth orbits and greater than 1 m at geosynchronous altitudes. These sensitivity thresholds are by and large a compromise between system cost and performance.

Knowledge of the meteoroid and space debris environment at sub-catalog sizes has up to this point been "calculated" in a statistical manner through experimental

Fig. 2 US Air Force Space Situational Tracking Satellite (Graphic courtesy USAF)

sensors with higher sensitivities. This will change when the Space Fence S-band radar becomes operational in 2018. This new capability is expected to be able to track and catalog perhaps 250,000 elements of space debris in low Earth orbit down to 1 cm – or about the size of a marble. Ground-based optical telescopes can generally detect GEO debris down to 10 cm in size, while in situ impact detectors (detectors flying onboard spacecraft) can sense objects down to a few micrometers in size. And while telescopes are perhaps best suited for GEO and high-altitude debris observations, radars are advantageous in the low Earth orbit (LEO) regime, below 2000 km.

Next Steps in US Laser Ranging Capabilities for Debris Tracking

Laser researchers at NASA's Goddard Space Flight Center in Greenbelt, Maryland, are currently developing a precise "laser ranging" method to define and track orbital debris more accurately. This high-resolution telescope has the ability to do actively determine high-velocity debris orbits by constant ranging. This activity is parallel to the tracking carried out by the smaller Electro Optic Systems telescope in Australia. This would overcome the current difficulties associated with passive optical and radar techniques and provide much greater accuracy information about orbital speeds and other similar data.

The current research is being carried out using the Goddard's Geophysical and Astronomical Observatory (GGAO) which is a 1.2 m (48 in.) with very high resolution.

This device that can transmit outgoing and receive incoming laser beams has been used to provide on-orbit calibration of some of Goddard's spacecraft. This device that has in the past actually obtained the precise distance and velocity and orbits of satellites as far away as exploratory missions to Mercury can provide much greater accuracy in tracking Earth debris (Goddard NASA Team).

ESA and German and European Tracking Activities EISCAT

http://www.esa.int/Our_Activities/Operations/Space_Debris/Scanning_observing

There are other capabilities for tracking and cataloging space debris around the world. One of the key capabilities is those operated via the European Space Agency. ESA radar tracking capability monitor debris increased after the Chinese ASAT test discussed above. These capabilities include the Tracking and Imaging Radar (TIRA) system near Bonn Germany operated by the Institute for High Frequency Physics and Radar Techniques that utilizes bistatic radar scanning.

TIRA debris detection system uses a 34 m dish antenna operating in L-band (i.e., 1.333 GHz) with a 0.45° beam width, at 1 MW peak power. Apart from tracking campaigns, the radar also conducts regular "beam park" experiments, where the radar beam is pointed in a fixed direction for 24 h, so that the beam scans 360° in a narrow strip on the celestial sphere, during a full Earth rotation. With the 24 h data collection, TIRA can detect debris data and determine coarse orbit information for objects of diameters down to 2 cm at 1000 km range. In a bistatic mode, together with the 100 m receiver antenna of the nearby Effelsberg radio telescope, the overall sensitivity increases down to almost 1 cm objects. A special seven-horn receiver, developed for the Effelsberg radio telescope, allows better resolution of object passages, permitting a reliable assessment of the object's radar cross section.

In addition to the German facility near Bonn, there is also the facility in Tromsø, Norway, known as the EISCAT Scientific Association (European Incoherent Scatter Radar). This facility operates a 930 MHz UHF radar and a 225 MHz VHF radar. In addition there is a 500 MHz radar system that consists of a steerable 32 m dish and a fixed 42 m dish nearby.

The primary mission of the EISCAT network is to perform ionospheric measurements. However, following the development of a dedicated space debris computer to run at the back end of the processing units, these radars are now capable of statistical observations of LEO debris down to 2 cm that can be accomplished without diminishing the main EISCAT objectives.

European studies of orbital debris as addressed in the previous chapter suggest that Dr. Kessler's assessment of a debris collision every 10 years may be optimistic and perhaps the result is more likely to be a collision every 5 years. The European Commission and ESA research also emphasize that about 10–15 large objects or about 7 t of debris need to be removed from space a year to reduce the risk of major collisions and damage to other spacecrafts. An object larger than 1 cm hitting a satellite with sufficient relative velocity would likely damage or destroy key subsystems or instruments on board, and a collision with an object larger than 10 cm would destroy the satellite, according to European Commission study findings. The degree of the problem and its urgency are that the study estimates that objects larger than 1 cm are now calculated to reach around one million in 2020 (Sixth European Conference on Space Debris 2014).

Space debris consists of human-made objects in Earth's orbit that no longer has a useful purpose, such as pieces of launched spacecraft. It is estimated that up to 600,000 objects larger than 1 cm and at least 16,000 larger than 10 cm orbit the Earth.

Japanese Initiatives

Japan has also developed tracking capabilities using radar and optical technology but has recently developed a proposal, at the RIKEN research institute a detailed proposal to use lasers in orbit. This would be not only to track debris with some precision but also to use higher-power laser systems to help delete space debris from orbit. In the past it has been suggested that land-based lasers could do this, but Japan has suggested that they might place a test-case laser experiment on the International Space Station to examine how well this work from space.

The concept would be to combine a superwide field-of-view optical telescope to detect objects plus a recently developed high-efficiency laser system, known as CAN, that could track space debris and remove it from orbit.

The RIKEN-developed telescope could be used to find debris with high precision. It was originally planned to detect ultraviolet light emitted from air showers produced by ultrahigh energy cosmic rays entering the atmosphere at night. The proposal is simply to adapt it to the new mission of detecting high-velocity debris in orbit near the ISS.

The CAN laser was originally developed to power particle accelerators. It consists of bundles of optical fibers that act in concert to efficiently produce powerful laser pulses. Combining these two instruments could in theory be used to locate and help with the deorbiting of dangerous space debris in potentially destructive orbits. This system would work only for debris elements, around the size of one centimeter. The intense laser beam focused on the debris will produce ablations. The result would be to reduce the debris' orbital velocity, leading to its reentry into the Earth's atmosphere. The proof of concept proposal would involve a 20 cm telescope and a laser wrapped by 100 fibers. The full-scale operational model would have a 3 m telescope and be wrapped by 10,000 optical fibers. The focus of the operational model would be on eliminating debris from the polar orbit at around 800 km altitude over about a 5-year period (Scientists want).

Active Removal of Orbital Debris

There are more than a dozen concepts about how space debris might either be actively removed from Earth orbit or its orbital affected so that the debris element would be removed more quickly from orbit due to accelerated decay. Increased solar activity during solar max also assists with debris removal. Satellites in orbit below 300 km tend to decay due to natural effects of gravity and atmospheric drag. The largest problems involve large defunct elements in low Earth orbit (especially around 800 km) and especially in polar, sun synchronous orbit.

There are many concepts that have been put forward about active debris removal, and these are discussed in several books; but today, there are no systems actually carrying out such missions not only because of high cost and technical difficulty but also because of issues relating to the Liability Convention that actually creates disincentives to remove debris because of the most legal implications if the removal is unsuccessful (Pelton 2015).

Conclusion

The space situational tracking capabilities to determine space debris orbits have continued to be upgraded in the past 10 years. The S-band radar Space Fence when deployed in 2018 coupled with capabilities such as the Analytic Graphics Inc./Space Data Association tracking system, the new Lockheed Martin and Electro Optic Systems commercial tracking networks, as well as other capabilities around the world such as that of ESA, JAXA, and other space agencies brings a high degree of sophistication to being able to track space debris and detect collisions and potential satellite conjunctions in advance so that evasive action can be taken. This is the good news.

The less favorable news is the degree to which orbital debris continues to increase. The fact that even 1 cm debris elements can do enormous damage when collisions occur at very high relative velocities is quite sobering when it is realized that there are over 250 K of these elements today and that they could number a million by year end 2020. The cascading effect is currently generating more and more debris elements that are quite hard to remove especially using the various techniques that have been suggested to help remove large debris elements. The proposal of the Japanese RIKIN Institute to use laser ablation may be the only cost-effective approach for addressing the smaller and very numerous debris elements. The visual image of debris provided in Fig. 3 below demonstrates the problem visually. It shows quite vividly that the low Earth orbit is truly becoming congested to the extent that it poses real problems with regard to the long-term sustainability of long-term use of outer space for all types of applications including telecommunications, space navigation, remote sensing, and meteorological satellites.

- There are clearly a number of actions that are needed to address this problem. The longer action is delayed, the more expensive, challenging, and difficult the problem becomes. The priority course of action includes the actions that are summarized below:
- Continuing efforts to deploy governmental and private tracking systems using higher frequency and thus more precise radar systems, optical tracking, and active optical ranging plus space-based tracking. This is so as to be able to know the

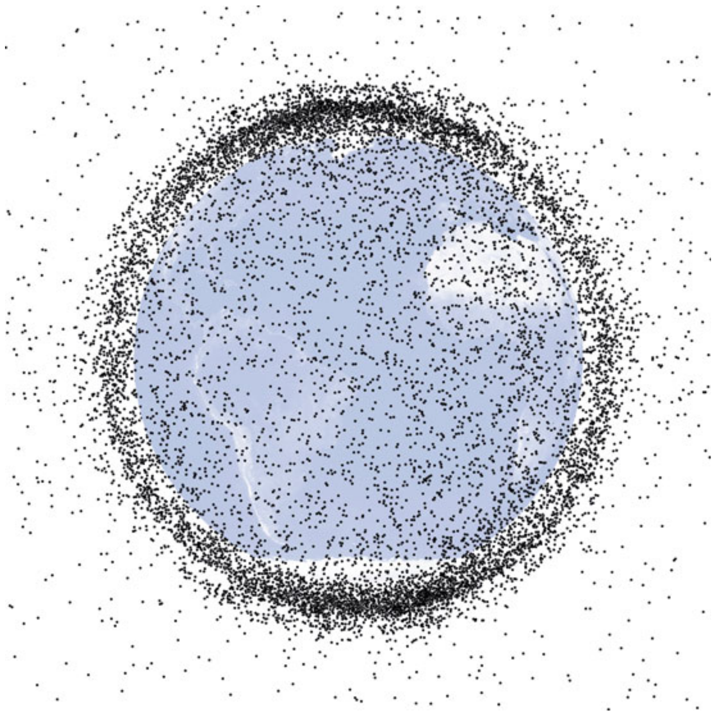


Fig. 3 Orbital congestion in low earth orbit (Image courtesy of NASA)

orbits of all elements in space – active and defunct. These actions are key to being able to undertake evasive actions when possible conjunctions seem likely.

- Undertake to remove as soon as possible the largest defunct debris elements in the critical low Earth orbits/polar orbits since this is the most dangerous situation that risks generating major showers of new debris.
- Make sure that all new satellites are equipped to meet the now well-established international guideline that all space objects are to be removed from orbit within 25 years of their end of life.
- Continue to carry out research as to new technology and systems that might be utilized to remove debris in the most efficient manner possible.
- The industry-created Space Data Association (SDA) should continue all of its various efforts to address this problem and to conduct best industry practices that can help alleviate the space debris problem and exercise operational vigilance so that operating satellites do not collide.
- Consider various ideas and proposals that have been suggested to address the space debris problem. These include such ideas as (i) requiring a separate debris removal capability to go on all new satellites that would not be under the control of the satellite operator, but a separate entity that would control ultimate deorbit, and (ii) considering that, in addition to providing for launch insurance, all entities

launching satellites should going forward pay into a debris removal fund so that resources would be created to clean up space over time.

- The Inter-Agency Space Debris Committee (IADC) has proved a very key resource to address the space debris problem and greatly assisted the UN Committee on the Peaceful Uses of Outer Space (COPUOS) on this issue. It should continue to assist COPUOS and its Working Group on the Long-Term Sustainability of Outer Space Activities that continue to address this now chronic problem.

Cross-References

▶ Orbital Debris and Sustainability of Space Operations

References

AGI-Lockheed Tout Commercial Space Surveillance Systems, http://spacenews.com/41727agi-lockheed-tout-commercial-space-surveillance-systems/#sthash.5CzBha2C.dpuf. Accessed 16 Jan 2016

Goddard NASA Team, http://www.nasa.gov/content/goddard/nasa-team-proposes-to-use-laser-to-track-orbital-debris/. Accessed 16 Jan 2016

Lockheed Martin Lands $914 Million Space Fence Contract. (2014), http://spacenews.com/40776lockheed-martin-lands-914m-space-fence-contract/. Accessed 16 Jan 2016

Michael Cooney, 10 crucial issues around controlling orbital debris, directing space traffic. (Network World, 2014), http://www.networkworld.com/article/2226898/security/10-crucial-issues-around-controlling-orbital-debris–directing-space-traffic.html. Accessed 16 Jan 2016

J.N. Pelton, *New Solutions to Orbital Debris Problems* (Springer, New York, 2015)

Scientists want to blast space debris with a space station-mounted laser, http://www.networkworld.com/article/2914663/security0/scientists-want-to-blast-space-debris-with-a-space-station-mounted-laser.html. Accessed 23 Jan 2016

Sixth European Conference on Space Debris. (European Satellite Agency, 2014), http://www.congrexprojects.com/2013-events/13a09/introduction. Accessed 23 Jan 2015

Space Data Association, http://www.space-data.org/sda/join-sda/. Accessed 16 Jan 2016

Space Fence. (2015), http://www.lockheedmartin.com/us/products/space-fence.html. Accessed 16 Jan 2016

U.S. Space Surveillance Network (SSN), https://en.wikipedia.org/wiki/United_States_Space_Surveillance_Network. Accessed 16 Jan 2016

Part VI

Spacecraft Bus and Ground Systems

Overview of the Spacecraft Bus

Tarik Kaya and Joseph N. Pelton

Contents

Abstract

The evolution of application satellites has hinged on the development of more and more sophisticated spacecraft buses or platforms. The development of three-axis body-stabilized platforms have allowed the deployment of more capable and much higher gain communications antennas, high resolution remote sensing and meteorological sensors, and more precise navigational payloads. The most important development in spacecraft buses has been the development of precisely

---

T. Kaya (✉)
Mechanical and Aerospace Engineering Department, Carleton University, Ottawa, ON, Canada
e-mail: tkaya@mae.carleton.ca

J.N. Pelton
International Space University Arlington, VA, USA
e-mail: peltonjoe@gmail.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_87

oriented body-stabilized platforms that allow the deployment of very high-powered solar arrays and very accurate pointing of high-gain antennas and sensor systems. Other challenges have included developing lower mass and structurally strong spacecraft bodies, improved and longer life thrusters, better performance power systems with greater density of charge, and improved thermal control systems. This chapter explores the development of the spacecraft bus and their technologies. The following chapters discuss tracking, telemetry, and command; reliability testing; and the adaptability of essential multipurpose platforms to different applications.

Keywords
Battery systems • Carbon/epoxy composites • Despun platforms • Fuel cells • Fuel slosh • Heat dissipation • Heat pipes • Inertial wheels • Isotope power systems • Momentum wheels • Nuclear propulsion • Orbital control Power systems • Quantum dot technology • Redundancy • Reliability and lifetime testing • Remote sensing sensors • Solar array • Solar cells • Spacecraft platforms • Spacecraft structures • Thermal control • Three-axis body stabilization • Thrusters

Introduction

Spacecraft Structures

The main goal of building spacecraft structures has been low mass and high strength. The evolution of carbon/epoxy composites and other similar hybrid materials have allowed engineers to construct satellite bodies that are up to three times lighter than the earliest models built with steel and titanium, yet the spacecraft structure can still be as strong as – or stronger than – the earliest satellite designs. The key to developing the strong but lightweight composite materials is to utilize them not only to build the structure that holds the solar arrays, the electronics, the antennas, the batteries, and momentum wheels but to also use these materials elsewhere if possible. Thus these materials may also be used in antenna masts, radiator panels, or in other parts of the satellite. Any mass saved is a true bonus in the field of space applications.

The initial satellite bodies were largely shaped as drums. These satellites maintained their stability for pointing to the Earth by spinning around then using this angular momentum to achieve their orientation – much like a spinning top. In the case of communications satellites, despun antenna systems revolved inside of the structural drum. This drum with the batteries, solar cells, and stabilization and positioning thrusters served as the satellite bus. These “internal antenna systems” spun in the opposite direction to the satellite bus. By spinning at exactly the same rotational speed but in the reverse direction, the effect was a completely synchronized system that could be constantly pointed to the Earth from GEO. Power was supplied to the antenna or sensing system from the bus via what was called a Bearing and Power Transfer Assembly (BAPTA).

This despun spacecraft design worked well and reliably in the early years. But this early design was shown in time to be inferior to the three-axis body-stabilization design for a variety of reasons. First, in the case of three-axis body-stabilized design, the solar array could be deployed to achieve 100 % sun illumination rather than 40 % illumination achieved with the rotational design. With a spinning satellite the solar cells mounted on the exterior drum were illuminated only a part of the time. The rest of the time they were behind the satellite until they reemerged after the outside drum rotated back into a view of the sun. Moreover, the angle of the drum could also shadow the solar cells and prevent optimum illumination.

Secondly, the spinning spacecraft structures were also found to be inferior in design to the box-like structures which could be stabilized on all three axes by means of high-speed reaction or momentum wheels. On the spinning spacecraft, the fuel tanks that supply the onboard thrusters spin with the spacecraft at speeds of up to 60 rpm. The fuel slosh from this action can be sufficient in some instances to cause sufficient precession to occur to send the satellite into a flat spin in the “X” axis rather than the intended “Y” axis. The design of reaction or momentum wheels, with their very high spin rate (i.e., 4,000–5,000 rpm) help to keep the entire spacecraft box and fuel tanks quite stable. Thus fuel slosh is not a problem.

The third reason in favor of three-axis body stabilization – that of greater pointing accuracy – was, in fact, the main driver that led to the design of the three-axis body-stabilized platforms in the first place. The details why ever more precise orientation and positioning capabilities were required for later generations of application satellites will be discussed in the next section.

As a result, the box-like structures have been extensively used in various space missions. The rectangular monocoque body (i.e., one in which the outer structure carries most of the load and stress requirements) results in better strength and stiffness characteristics. The structural design of the satellites is mostly standardized. In the case of some special science missions, very different external configurations are of course possible. However, in general, the satellites are manufactured around a primary structure. The primary structure carries the main external loads and is usually made from a central tube in combination with flat panels. This central tube is usually the major load path between the launcher and spacecraft components. The secondary structure includes mounting platforms, solar panels, and other deployable parts. The secondary structure also serves as a closure panel for the satellite. Finally, for housing electronics and supporting electric cables, etc., some smaller structures are also used (Fig. 1).

Research continues to develop even lighter and stronger materials for spacecraft structures. Also certain problems with epoxy or resin composites in orbit are being researched. Atomic oxygen that reacts with these composites causes microscopic deterioration in certain resins over time. The solution to this problem has been to apply a protective coat to shield the composite materials in the spacecraft structure from the atomic oxygen. Just a thin veneer prevents the atomic oxygen from chemically interacting with the composite materials of the spacecraft structure. Also research continues to develop even more reliable momentum wheels with “magnetic bearings” and thus no rotational friction. These high-reliability

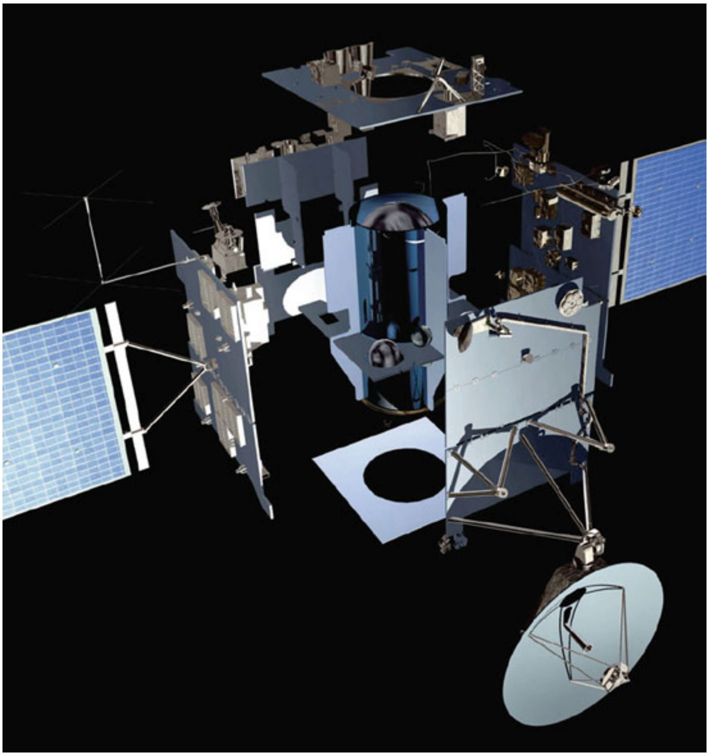


Fig. 1 An exploded view of Rosetta spacecraft with the primary and secondary structures (Graphic Courtesy of ESA)

momentum wheels are designed to maintain quite accurate pointing capability and a high level of stability for the entire spacecraft – i.e., the bus itself, the deployed solar arrays and especially for the payload antennas, sensors, and meteorological and navigational payload systems.

Deployable structures are also a current area of research for several space applications since they can potentially decrease the spacecraft mass and volume, which is very important to lowering launch costs. It is also possible to launch larger antennas by using deployable structures for the GEO communications spacecraft, which can help further reducing size and power requirements of mobile ground receivers.

Orbital Control and Pointing Accuracy

The original development of three-axis body-stabilized design was an attempt to build a more efficient spacecraft, which is capable of very long distance data relays, associated with planetary missions. The greater distance involved in planetary

Fig. 2 A momentum wheel that is used to achieve spacecraft stability (Photo courtesy of NASA) (www.sti.nasa.gov/tto/spinoff1997/t3.html)

research vehicles requires much greater pointing accuracy than could be achieved by a spin-stabilized spacecraft. The creation of a body-stabilized spacecraft created the bonus of allowing the "wings" of a solar power array to be oriented to achieve maximum solar exposure. The use of gravity gradient or spin-stabilized spacecraft are not generally used for most sophisticated and high-value application satellites, simply because they do not offer the pointing accuracy required by most of today's satellites.

The pointing accuracy of a satellite depends not only on the speed and performance of a momentum or reaction wheel but also the ability of the spacecraft to be oriented exactly in the desired direction over time. The very high torque of the reaction or momentum wheel allows very small adjustments. By adding or subtracting a small amount of energy to a reaction wheel on a single axis stability can be achieved. Once this is done in all three axes the satellite can be perfectly stabilized and pointed in the desired direction. A momentum wheel rotates at a higher speed than a reaction wheel (up to 10,000 rpm) to provide the required stability. Momentum wheels are quite small in diameter (i.e., in the range of 15–20 cm in diameter) but their very high rotational speed of 4,000–5,000 rpm give them a great deal of inertial force or torque. The screwdriver in Fig. 2 helps to picture the scale of these momentum wheels.

The initial application satellites were oriented in the desired direction by the use of Earth, sun, and star sensors. These sensors worked reasonably well. In the case, however, of a satellite losing orientation these sensors could lead to difficulties in reestablishing the proper orientation with sufficient speed. The main problem, however, was that such sensors only allowed a pointing accuracy of about 0.5–1.0°. This was all of the pointing accuracy that despun platforms could achieve, but three-axis body-stabilized platforms could be pointed with greater precision. It therefore followed that a more precise system to orient satellites would be needed. This led to the use of RF beacons to align satellites with precise locations on the ground. These RF beacons can allow a body-stabilized spacecraft to be oriented from GEO to a quarter degree of pointing accuracy. Currently there are alternative methods being researched that would use laser systems that could achieve even more precise pointing accuracy.

The other aspect of orbital positioning and orientation depends on onboard thrusters to keep the satellite in the desired orbital position and properly pointed. This is the greatest challenge for satellites in GEO since this type of spacecraft is positioned in equatorial orbit at truly great distances – approximately one tenth of the distance from the Earth to the moon. For many years, orienting thrusters on spacecraft have used hydrazine fuel or a combination of hydrazine and other hypergolic fuels such as nitrogen tetroxide in bipropellant systems. Hypergolic fuels are almost always quite toxic, but have the special feature of exploding with great force on contact with an oxidizer. Vernier thrusters with tiny jets could be used with precision to achieve great accuracy with such propellant systems. There has been research at NASA Glenn Research Center to develop nontoxic and lower cost fuels that still achieve spontaneous combustion. Based on this research, NASA has now patented an innovation that overcomes the problem of using fuels such as methanol as a propellant for satellite thrusters. Their research has developed a satellite thruster that offers the ability to catalytically decompose a reduced-toxicity propellant into hot gases but still having the ability to spontaneously react with an oxidizer to begin the combustion process. This new system can be used for both bipropellant and monopropellant satellite propulsion and in the process this approach can also help to reduce the cost and complexity of satellite missions.[1]

Satellites in GEO are constantly being tugged away from their specifically assigned equatorial position. The gravitational forces of the sun and the moon as well as the Earth's irregular shape and composition serve to move GEO satellites East or West along the equator away from their assigned position. Much more to the point, these various gravitational forces are much stronger in the North and South vertical directions in or out from GEO. These forces are at least ten times more powerful in the North and South direction than in the East and West directions.

The answer to this problem is to develop propulsive systems with much longer lives. The development of electric propulsion and ion propulsion systems have been the answer to longer lived systems. Electric propulsion provides thrust capabilities for much longer, but with much lower levels of force. In an ion engine, ions from a heated gas such as Xenon are accelerated by the electrostatic force to tremendous speeds to provide the thrust (Fig. 3).

In the longer term, ion engines provide more than two to three times the energy per unit of mass than chemical propulsion, but the thrust force at any particular time

---

[1] NASA Glenn Research Center: Reduced toxicity fuel satellite propulsion system including fuel cell reformer with alcohols like methanol. http://technology.grc.nasa.gov/tech-detail-coded.php?cid=GR-50278mini=y

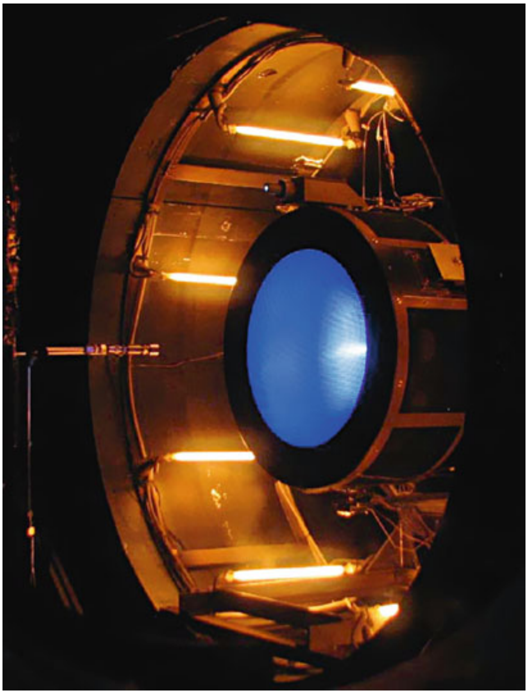


Fig. 3 A long-life Xenon ion thruster (Photo courtesy of NASA) ("Structure and Properties of Matter: Ion Thruster." Ion Thruster from NASA. http://dawn.jpl.nasa.gov/DawnClassrooms/2_ion_prop/index.asp)

is much weaker. This often ends with the need to combine chemical propulsion systems to achieve initial orbital location. It is left to ion engines to keep the satellite oriented and positioned properly over the longer term. Application satellites have grown in sophistication, lifetime, power, and performance and this has also meant that they have grown more massive and large. Although electronics, processors, and application-specific integrated circuits (ASICs) have shrunk in size, antenna and power systems have increased greatly in mass and volume. This means that positioning, stabilization, and orientation systems have also grown in size and performance capabilities have had to improve. Ion engines have certainly increased longer term capabilities, but there has been increased research into the possibility of developing nuclear-powered capabilities. Such nuclear-powered systems could provide for not only orbital positioning but also provide for the power supply as well. There remain serious concerns about nuclear power safety, but nuclear-powered propulsion as well as power systems to support remote sensing (especially radar systems) and very large communications satellites remain a future possibility.

Power Systems

All application satellites require a considerable amount of power to operate and thus providing a reliable, long-lived source of power is critical to mission success. Some types of satellites require greater amounts of power than others. In the case of remote sensing satellites, radar satellites that represent "active sensing," which initiates a signal from the satellite that "bounces back" requires much more power than other remote sensing satellites. In the case of satellite communications, direct broadcast satellites and mobile satellite systems require much more power than fixed satellite systems. All application satellites benefit from a longer lifetime. Thus power systems that can be recharged and last 15 years – if not longer – are certainly welcome.

The most common type of failure in the field of space applications are those related to loss of power – either in space or in ground systems. It is thus important to design systems that have high performance but also have the greatest possible reliability. Most application satellites are powered by solar arrays with backup battery systems that supply power during periods of eclipse or short-term outages that might occur. There has been a constant effort to upgrade the performance of solar arrays that has seen the use of ever more efficient solar cells. The first solar cells were silicon based. These were upgraded from amorphous structure silicon solar cells to structured crystalline structure solar cells that produced higher efficiency conversion of photon received from the sun. In time, even higher performance gallium arsenide solar cells were used. Although these solar cells are more expensive to manufacture the higher performance justifies the investment when the high cost of launch to orbit is considered. Some designs have included solar concentrators that allow the solar array to see the equivalent of two or even three suns. Today there is continuing research to improve solar array performance even further. One of the prime objectives is the reduction of the production cost for high-performance gallium arsenide solar cells. Recent breakthroughs in the production of flexible gallium arsenide solar cells promise the relatively near-term reduction in cost (Fig. 4).

Work at research labs has identified methods that allow the mass production of gallium arsenide cells that can be stripped off and quickly applied to substrates. The production of gallium arsenide solar cells with higher efficiencies and much lower costs are likely within the next 2–3 years.

There are ongoing efforts to increase these efficiencies to even higher levels. Quantum dots that can be "grown" on solar cells to increase their efficiency up to an estimated 50 % are also under development. These quantum dot developments that allow more photons to be absorbed would translate into higher energy conversion efficiencies for high-performance solar cells. This type of design coupled with an addition of more junctions – especially junctions that capture energy in the ultraviolet range – is also in development (Fig. 5).[2]

By choosing the most suitable materials for cell junctions, it has been shown that more efficient multijunction solar cells can be manufactured. The materials are chosen to absorb as much solar energy as possible, thus taking advantage of a larger part of the incoming solar power. In these cells, the junctions are connected in series instead of having a single junction of the conventional solar cell. Thus, the

---

[2] RIT: Solving the world's energy crisis by improving the efficiency of photovoltaics. https://www.rit.edu/showcase/index.php?id=36

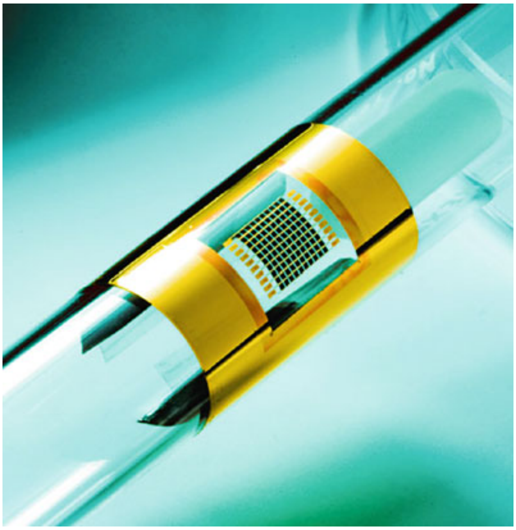


Fig. 4 Rapid production of flexible gallium arsenide solar cells will reduce cost (Photo courtesy of Prof. John Rogers, University of Illinois at Urbana) (Beleicher 2010)

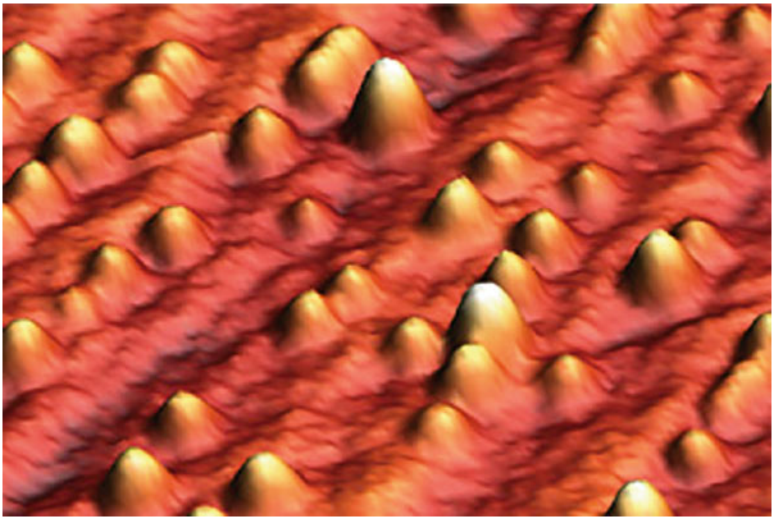


Fig. 5 Quantum dots can be "grown" on solar cells to increase their efficiency (Graphic courtesy of Rochester Institute of Technology Nano Power Research Labs)

multijunction cells generate lower current than the single junction cells but much higher voltage and power. As a result, the triple-junction cells have high conversion efficiencies on the order of 30 %. The triple-junction cells have been used in several space missions. A recent example is NASA's Mars Exploration Rovers: Spirit and Opportunity. Research is under way to manufacture cells with four or higher junctions to increase further cell efficiencies. Multijunction cell efficiencies up to

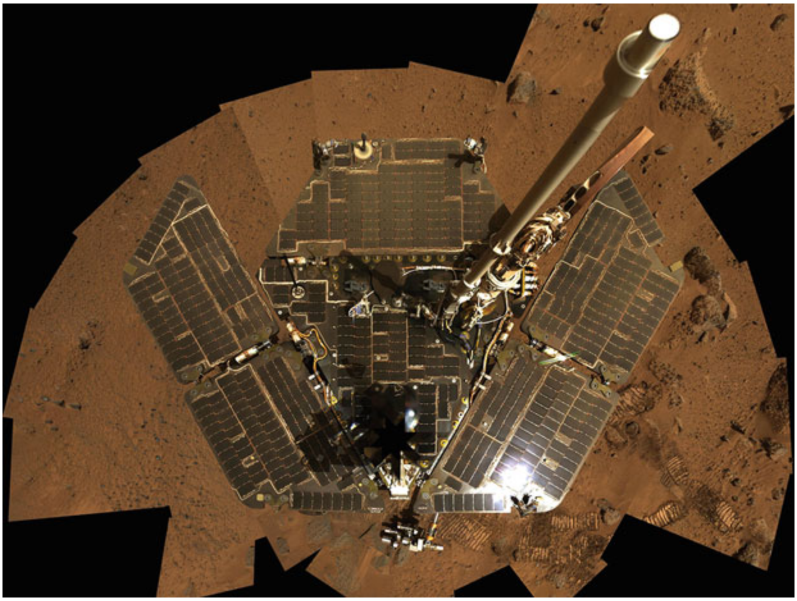


Fig. 6 The high-efficiency triplejunction solar cells were used on the solar arrays of the NASA's Mars Exploration Rover Spirit (Photo courtesy of NASA)

42 % in laboratory conditions have been reported. The combination of the best materials, quantum dot technology, and multijunction solar cells may allow the development of solar cells with over 50 % conversion efficiencies (Fig. 6).

The other objective with regard to solar cell power system is to develop more advanced solar array deployment systems. The latest designs can roll out solar cells on a thin plastic substrate reel that is flexible rather than a rigid one. Another alternative new design would be to have inflatable solar arrays that could be quite low in mass and create effective deployment systems for flexible arrays. Flexible solar arrays, which are deployed with low-mass thin-film solar cells, allow the overall mass of the solar cell system to be less and more reliable. Current state of the art in these flexible solar arrays is on the order of 150–200 W/kg.[3]

Batteries

The onboard batteries are a critical component of the spacecraft's power system. Although a satellite can operate exclusively off of a solar array when it is being fully illuminated, the problem is that satellites of various types and in various types of orbits go into eclipse and then batteries become essential. LEO and MEO satellites

---

[3] P.A. Jones, S.F. White, T.J. Harvey, B.S. Smith, A high specific power solar array for low to mid-power spacecraft. http://www.aec-able.com/corpinfo/Resources/ultraflex.pdf

operating in constellations – especially LEO satellites – are constantly going into and out of eclipses as they move behind the Earth. Thus satellites in LEO and MEO are constantly dependent on their batteries on an orbit-to-orbit basis. The Iridium constellation system in LEO was designed so that when these satellites are in the polar region they can be reconditioned and recharged. For satellites in the LEO and MEO orbits, conditioning of batteries and consideration of when and how to discharge are a much more crucial matter and certainly plays into spacecraft lifetime and reliability considerations.

For GEO satellites the considerations are much different. For long periods of time during the year the satellite is in continuous solar illumination. There is an annual cycle, however, which twice a year brings a GEO satellite into a daily eclipse. During the maximum eclipse period the satellite is without solar-based power for over an hour. Battery-based operation must sustain the satellite operations during these periods. At the beginning of the cycle, there is only a very brief period of eclipse that builds up to over an hour as of the winter solstice, for instance, and then the eclipse gradually subsides.

At the outset of space applications, spacecraft were equipped with nickel cadmium batteries that provided reasonably good power density and lifetime. Over time higher energy density nickel hydrogen batteries were developed and implemented on more and more spacecraft. A shift to different battery types is not a decision taken lightly since the battery system represents a single point of failure. Extensive lifetime testing is undertaken but since spacecraft lifetimes of up to 15 years are desired, this becomes a major barrier to innovation. In the case of batteries, testing under elevated temperatures can be utilized as a means of compressing the reliability and lifetime testing process. Most recently, the transition to lithium ion batteries for spacecraft power storage has taken place. This type of batteries that are in wide usage for cell phones, laptop computers, etc., contain the highest energy storage density and have proved to be reliable over long lifetimes and quite adaptable to recharging.

Batteries are key elements of a satellite design for spacecraft engineers. They represent a critical resource to the overall functioning of the payload and thus a potential single point of failure. They, however, also constitute an element of mass and volume that limits the size of the payload or increases the cost of the launch services. This has led to an ongoing effort to develop batteries that contain a higher energy density, are lighter in weight, and still quite reliable over time with a mean time to failure in excess of 15 years.

There have even been thoughts that if one could develop sufficiently lightweight and high-performance batteries (or perhaps fuel cells as will be addressed immediately below), one could eliminate solar cell arrays and develop satellites that were entirely powered by batteries and/or fuel cells or combine such systems with thermoionic converters. As the power requirements for applications have ascended to perhaps tens of kilowatts, the feasibility of such a design approach has begun to seem more and more unlikely.

The research for the future is aimed at developing new technology for energy storage that could be more efficient than current battery technology as we know it today. Prime in this regard is the concept of developing unitary regenerative fuel cells. Fuel cells typically take hydrogen and oxygen and combine them via an electrochemical system that is triggered by a catalyst to make this process happen. In a way this electrochemical process can be likened to battery, but in this case the product is both electricity and water. The water, within what is called a "regenerative fuel cell," can then be electrolyzed to produce hydrogen and oxygen once again and start the process over again. Recently, researchers have, however, experimented with other possible regenerative chemical interactions. One of these fuel cell design processes has found that hydrogen peroxide (H_2O_2) and sodium borohydride ($NaBH_4$) can produce long-lived and effective reactions. Although this system is currently operating at an efficiency of 1,000 W-h/kg, it is projected that this system (for possible spacecraft use) might be able to operate at efficiencies that are almost three times greater. Other approaches to hydrogen peroxide fuel cells have found that methanol can operate efficiently with a lower cost alkaline-based catalyst.[4]

Various types of regenerative fuel cell systems have been under development for many years and indeed successful fuel cells have been developed for use in ground vehicles, aircraft, and even spacecraft. The prime power storage systems on the Space Shuttle were fuel cells. The most recent research suggests that breakthroughs in the next few years can see fuel cell technologies used not only in space and in vehicles but perhaps to serve as energy systems for buildings and other industrial uses.

The problem to date has been that the catalysts have typically involved quite expensive materials. A quantitative objective of developing a fuel cell that could generate up to a kilowatt of power for under $1000 has long eluded developers working in this field. Recent efforts to develop lower cost catalysts and to perfect systems with longer life and lower mass have increased hopes that viable regenerative fuel cell technology for application satellites can be implemented in near future. Perhaps the hydrogen peroxide fuel cell system might prove both feasible and cost effective for such purposes in the next few years.

Another area of research with regard to cost-effective energy storage is the development of very high velocity flywheels that can store a significant amount of energy. Flywheel systems are increasingly being used on the ground as backup energy storage systems for emergency power restoration in emergency communications systems and even conventional power storage systems. Such flywheel systems do not require an electrochemical process to take place. Nor do these systems require almost continuous operational management with periodic discharge and other maintenance efforts, which is the case with batteries. The most attractive aspect of this approach with regard to application satellites is the possibility that the flywheels could serve a dual purpose of not only storing energy but also utilizing these flywheels as reaction wheels for stabilization and orientation purposes as well. NASA and other space agencies are thus pursuing research in this area with the

---

[4] A methanol and hydrogen peroxide fuel cell using non noble gas catalyst in alkaline solution. http://etd.lsu.edu/docs/available/etd-11052006-193341/unrestricted/Sung_thesis/PDF

objective of in-orbit tests of reaction or momentum wheels that could also serve as energy storage systems.[5]

Nuclear and Isotope Power Systems

The story of application satellites has been in many ways a history of "technology inversion." This means that application satellites have evolved from small, low-powered devices that work with large, sophisticated, and high-powered ground systems to the "inverse situation." This means that today's application satellites, for the most part, are large, sophisticated, and high-powered systems in space that work to increasingly small, mobile, and low-powered systems on the ground. In some cases the ground systems can even be handheld mobile systems. This trend has pushed ever upward the demand for onboard power systems – especially for active radar systems for remote sensing and mobile and broadcast communications satellites. The first communications satellites such as Early Bird (1965) generated 100 W of power. Today's most capable application satellites may typically require on the order of 12–15 kW. The various technologies discussed above have been able to deploy power systems capable of generating such levels of powers, but the future suggests that even higher power levels may well be required.

One possible pathway forward would entail the use of nuclear reactors or radioisotope systems to power the application satellites of the future. As of this time only a few small nuclear reactors designed to power manned missions or planetary explorations have been launched into space. Although Russia has launched as many as 35 compact nuclear reactors to power space missions, the United States has launched only one such system. This was the SNAP 10A (Systems for Nuclear Auxiliary Power) launched on April 3, 1965. This was, in fact, the first nuclear reactor launched into space and the only US mission of this type. The launch of a nuclear reactor into space has been considered politically controversial around the world because of the radioactive contamination that could occur in the event of a launch failure or in the event the spacecraft was not deorbited safely (Fig. 7).

The SNAP 10A spacecraft had several components. These consisted of a compact nuclear reactor, the nuclear reactor controls and associated reflector system, and a heat transfer and power generator system. The nuclear reactors launched by the Soviet Union and Russia are considered to be similar in design. For reasons of safety, cost, and performance requirements it seems unlikely that a complete nuclear reactor system would be used to power future application satellites.[6]

---

[5] Fuel cells and hybrid energy systems. NASA Space Architecture Strategic Research Plan. http://www.macrovu.com/image/PVT/NASA/RPC/uc%3DFuelCellV4.pdf

[6] G.L. Bennet, Space Nuclear Power: Opening the Final Frontier Fourth International Energy Conversion Engineering Conference (IECEC), San Diego, California. http://www.fas.org/nuke/space/bennett0706.pdf

Fig. 7 SNAP 10A, the first US nuclear reactor launched into Earth orbit (Photo courtesy of NASA)

There have been a number of radioisotope-powered systems launched into space by both the United States and the Soviet Union/Russia. These Radioisotope Thermoelectric Generators (RTGs) have been developed under the SNAP Program within the United States. A variety of the RTG systems powered by a number of different SNAP radioisotope fuelled power plants have, in fact, been launched by the United States since the 1960s. In fact, over 40 such radioisotope-powered systems have been launched – essentially all in support of planetary exploration missions where the duration of the mission and remoteness from the sun has limited the feasibility of solar cell arrays serving as the reliable, longer-term power supply.

The New Horizon spacecraft launched in 2006 to study the Pluto system and the Kuiper belt is powered by an RTG. This RTG operates by heating a number of thermoelectric converters that generate electricity with about 7 % efficiency.[7] This low efficiency implies that an RTG will generate more heat than electrical power. The excess heat can be used to maintain certain components at a suitable level of

---

[7] NASA New Horizon: Mission to Pluto and Kuiper belt. www.nasa.gov/mission_pages/newhorizons/main/index.html

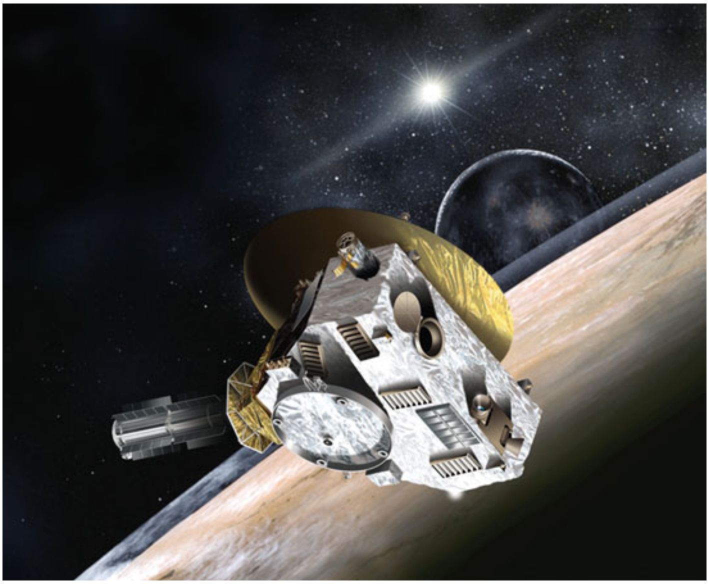


Fig. 8 The New Horizon exploratory spacecraft with an RTG power source (Photo courtesy of NASA)

warmth but in most of the cases the thermal subsystem needs to deal with this waste heat (Fig. 8).

These RTG power sources produce heat as the source of electricity when solar power is not a viable power supply and the mission is of a very long duration. The isotopes used in RTGs can be many different types, but the general purpose heat supplies (GPHS) for the Cassini Mission and the New Horizons probe are Plutonium 238 oxide pellets.

When selecting an isotope for an RTG, it is necessary to consider several mission parameters including mission life, sensitivity of equipment to radiation, and desired temperature range. Plutonium 238 has been the fuel of choice for nearly all RTG-powered spacecraft. Plutonium 238 when decaying emits alpha particles which are easily shielded. In addition, it has a half-life of 87.8 years. Although this is much more than required, it gives the best available power density and half-life time combination. However, there have been difficulties in obtaining the required amount of Plutonium 238 in recent missions, leading to the significant delays. An alternative isotope is Polonium 210, which has in fact much higher power density but the half-life time is much shorter. For missions with long lifetimes, this is a serious limitation. Table 1 provides a list of some common RTG isotopes.

To date, no applications satellite has had sufficient power need or size to justify the use of an RTG.[8] It is believed by some engineers that if and when electric power

---

[8] RTG history and New Horizons. www.osti.gov/accomplishments/rtg.html

Table 1 Information regarding various radioisotopes

Radioisotope	Class of emitter	Half-life	Watts per gram (W/g)
Plutonium 238	α	87.8 years	0.39
Polonium 210	α	138.4 days	140
Curium 242	α	165.6 days	120
Curium 244	α	18.1 years	2.27

The design of an RTG power supply that provides sufficient heat to a surrounding grid of thermoelectric converters is shown in Fig. 9.

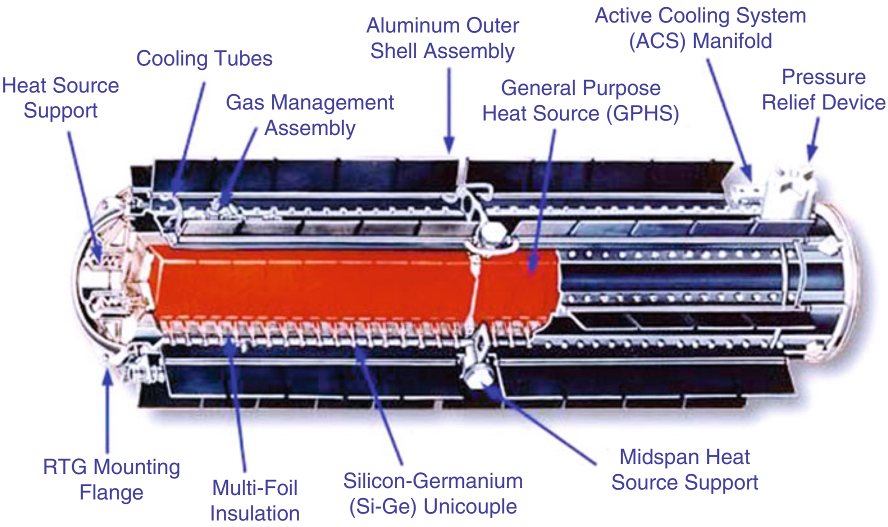


Fig. 9 A schematic diagram of a general purpose heat supply RTG (Graphic courtesy of US Department of Energy, Office of Nuclear Energy, Science and Technology)

system needs reach a level of perhaps 40 kW, radioisotope-powered heat sources working with thermoelectric converters with 10 % efficiencies could justify the use of such devices.

This approach would require additional thermal control systems to dissipate the excess heat and would of course entail special safety procedures at satellite launch and with deorbiting procedures. The current applications that involve sending spacecraft into deep space, of course, do not raise the deorbiting issue.

Thermal Control and Heat Dissipation

The key element to remember about all application satellites is that they operate in a very hostile environment and there are no repair crews to fix a malfunction. One of the major environmental problems that satellite designers must address is the thermal environment. Regulating the thermal environment is key because many electronics

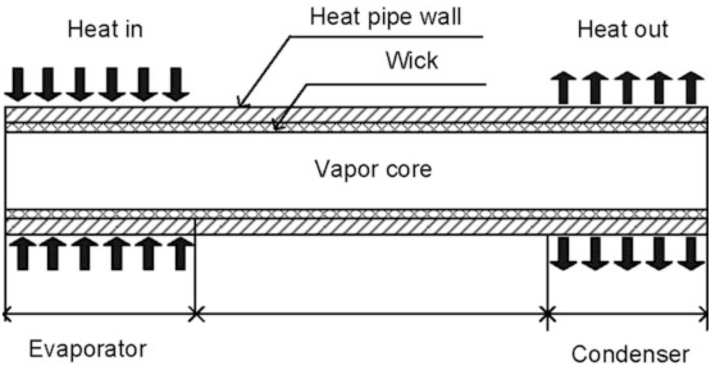


Fig. 10 The basic schematic concept for a heat pipe

must be kept within a modest temperature range to operate reliably. Further there are special conditions that must be considered. One of these is the sharp thermal gradient that occurs when a satellite moves from an eclipse condition to full solar illumination where the outside of the satellite may heat rapidly within a few minutes. There can also be special conditions where a remote sensing instrument must be kept in cryogenic conditions to obtain precise information.

Outer space is cold but an object in space is warmed by the sun's irradiation. Thus the spacecraft must be designed so that it absorbs some degree of heat but it also reflects some of the heat so that the spacecraft remains neither too hot nor too cold. Further, the power system of the satellite generates electricity and heat. The interior heat of a satellite can build up over time unless there is a mechanism to transfer the heat inside the spacecraft to the outer shell. There are several ways of achieving this goal. One of the most efficient method uses a "heat pipe."

A conventional heat pipe mechanism is depicted in Fig. 10 that indicates how these devices operate. Heat is applied to the outer casing of the evaporator side, leading to the evaporation of the liquid. The resulting vapor is pushed toward the condenser. The meniscus formed at the surface or inside the capillary structure naturally adjusts itself to establish a capillary head that matches the total pressure drop. The subcooled liquid from the condenser returns to the evaporator as a result of the capillary pressure, completing the cycle. Thus, the heat pipe operation is passive (no moving parts). One of the main advantages of these types of systems is high-reliability operation without consuming a great deal of the mass budget for the satellite with a minimum demand on the volume that these devices consume. Although the passive and active elements for thermal control and regulation of a satellite design typically involves only about 2–3 % of the total mass budget, there are still research efforts to reduce the mass and volume of these systems so as to increase the satellite's payload.[9]

---

[9] Spacecraft thermal control. http://webserver.dmt.upm.es/isidoro/tc3/STC%20missions%20and%20needs.htm

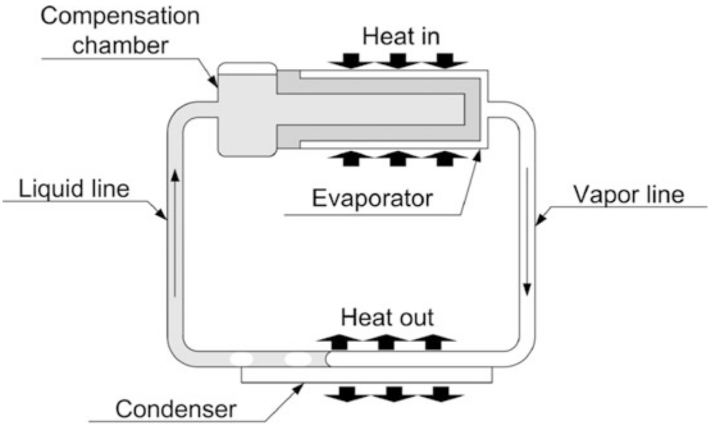


Fig. 11 Loop heat pipe (LHP) schematic (not-to-scale)

On the other hand, if the thermal controls do not work properly then the entire mission can be jeopardized. The first operational communications satellite with a despun antenna, the Intelsat III, had a spinning antenna bearing freeze that was thought to be, in part, the fault of inadequate thermal control. In this case, the entire satellite was declared a complete failure and the satellite's bearing and thermal control system had to be rapidly redesigned against production and launch schedule constraints.

Heat pipes have been standard tools in spacecraft thermal control and there is sufficient flight heritage. One important drawback of the heat pipes is that they do not work at adverse elevation under gravity. This is an important limitation for ground testing and may also limit operations, e.g., exploration missions on planetary surfaces or adverse acceleration forces during the spacecraft maneuvers. A more recent version of the heat pipes is a loop heat pipe (LHP), which was invented in Russia in the late 1970s. The operation principle of an LHP is similar to the heat pipe described above. In addition, the reservoir or compensation chamber as shown in Fig. 11 provides additional volume for the fluid volume changes, allowing variable conductance operation. The porous structure is only limited to the evaporator section, thus very high capillary pumping can be obtained by using metal-sintered wicks without introducing unnecessarily high-flow resistance. LHPs offer several advantages which make these systems very attractive for the thermally demanding missions. In addition to the passive operation, they can work against gravity, up to several g loads. They exhibit diode behavior (i.e., if the condenser is warmer than the evaporator, the fluid circulation automatically stops). Therefore, they can be used as a thermal switch. For instruments requiring fine temperature control, a tight temperature range as narrow as ± 0.1 K is also feasible.

LHP transport lines (liquid and vapor lines) can employ flexible tubing. This allows easy integration and more importantly this design allows deployable radiators. The Boeing-702 communications bus was the first US commercial satellite to

Fig. 12 A small HP unit for spacecraft thermal control (229 mm × 127 mm) (Photo courtesy of NASA)

employ a LHP-based deployable radiator system. Research in this area continues because of the importance of this technology. The development efforts are focusing on the multiple evaporator and multiple condenser systems, operation at higher heat flux, and miniaturization. Figure 12 shows a small LHP unit.

Onboard Heaters and Cooling Systems

One of the greatest challenges of maintaining a long life in a spacecraft that is equipped with a large amount of electronics is keeping the temperature within a range that is actually quite narrow. In most cases it is desirable to keep electronic gear within a range of about 5–25 °C. There is a large temperature range when the satellite must make the transition during launch from quite hospitable temperatures to the coldness of outer space in a matter of only a few minutes. Once a satellite is in orbit, it may fly around the Earth in LEO that every 90 min or so brings it from full sun illumination to a very cold environment as the satellite travels behind the Earth. The use of passive techniques such as multilayer insulation (MLI) on the exterior of the satellite and the use of heat pipes can help a good deal, but most application satellites must have more elaborate heating and cooling devices to help keep the satellite's interior in a safe and viable temperature range. These can be closely akin to electric heaters and refrigerant systems used on the ground except they are subjected to extensive lifetime testing processes to insure long life and the ability to withstand the forces of liftoff. There are also often power converters from DC to AC current and these can also serve the dual purpose of assisting with heating devices.

Conclusion

The optimum design of an application satellite is not only a difficult and demanding process but one that is constantly evolving as research into a wide range of technologies and materials continues over time. This chapter has briefly addressed spacecraft structure and materials, positioning and orientation systems, sun, star, and Earth sensors, and other systems for maintaining spacecraft pointing, power systems, and thermal control systems.

New materials not only for the spacecraft bus structure but also antennas and solar cell deployment arrays have allowed application satellites to become lighter (with regard to these component parts) and thus allowed more mass (and sometimes volume) for mission payload. Alternatively these advances in structural materials have allowed for reduced payload service costs. Spacecraft are also better able to be accurately positioned and pointed to allow higher precision monitoring or for higher gain telecommunications antennas to function more effectively. Currently the most precisely pointed application satellites, operating from GEO, can be pointed with a directional accuracy of up to 0.25°.

The objective for positioning and orientation systems is not only precision but long life as well. High-performance ion engines that are capable of sustained operation over long periods of time allow precise deployment and operation. Over the longer term, ion engines may not only provide station keeping and pointing control but also assist with deployment of satellites from LEO to GEO over a longer term period. Overcoming and solving the orbital debris problem is becoming an increasingly serious issue – especially in LEO. This slow deployment from LEO to GEO, which increases the risk of collision, poses a deterrent to such a slow spiraling orbital deployment method.

One of the most critical and demanding aspects of an application satellite is reliable and sustained electrical power supply. Power is key to all types of application satellites and failure of spacecraft or short-term outages are most frequently related to power supply difficulties. Over time, power supply technologies have increased greatly in capability. Solar cells are lighter and convert energy with higher efficiency. Flexible array structures are lighter in weight. Batteries now have greater power density and last longer, but there are new power storage technologies under development. These new technologies include regenerative fuel cells, flywheel systems, and possibly radioactive thermoelectric generators and thermo-ionic converters.

Finally, the performance and reliability of application satellites are dependent on the thermal environment of the spacecraft being maintained within a narrow range of temperature gradients. The thermal vacuum testing of integrated spacecraft and components and thermal control systems have proven valuable over the years. Efforts continue to develop thermal control systems that can handle higher power densities as well as to develop improved passive and active units to help control temperature extremes. Research also continues to develop thermal control systems which can provide temperature control within a very narrow range (0.1 K) for demanding payloads.

The Tracking, Telemetry, Command, and Monitoring (TTC&M) systems on spacecraft are critical to their operation and this subject will be addressed in the following chapter. There will be additional chapters that address the reliability testing and engineering of satellites to achieve longer life, the problems of orbital debris, and other issues that are critical to performance of application satellites.

Cross-References

- ► Common Elements Versus Unique Requirements in Various Types of Satellite Application Systems
- ► Lifetime Testing, Redundancy, Reliability, and Mean Time to Failure
- ► Telemetry, Tracking, and Command (TT&C)

References

A. Beleicher, Rubber-stamping makes creating solar cells, transistors, and infrared detectors easy (IEEE Spectrum, 2010), http://spectrum.ieee.org/semiconductors/materials/thinfilm-trick-makes-gallium-arsenide-devices-cheap. Accessed May 2010

Telemetry, Tracking, and Command (TT&C)

Arthur Norman Guest

Contents

Abstract

The telemetry, tracking, and control (TT&C) subsystem of a satellite provides a connection between the satellite itself and the facilities on the ground. The purpose of the TT&C function is to ensure the satellite performs correctly. As part of the spacecraft bus, the TT&C subsystem is required for all satellites regardless of the application. This chapter describes the three major tasks that the TT&C subsystem performs to ensure the successful operation of an applications satellite: (1) the monitoring of the health and status of the satellite through the collection, processing, and transmission of data from the various spacecraft subsystems, (2) the determination of the satellite's exact location through the reception, processing, and transmitting of ranging signals, and (3) the proper control of satellite through the reception, processing, and implementation of commands transmitted from the ground. Some advanced spacecraft designs have evolved toward "autonomous operations" so that many of the control functions have been automated and thus do not require ground intervention except under emergency conditions.

A.N. Guest (✉)
International Space University, San Francisco, USA
e-mail: guest.arthur@gmail.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_69

Keywords
Command • Communications • Computer • Control • Data handling • Guidance • Navigation • Range • Satellite • Telemetry • Tracking

Introduction

Regardless of the application it is being used for, satellites represent a complex system of hardware and software. However, the satellite is only one component in the larger system required to provide the service for which it was built and launched. There are three specific segments that must work together for the larger overall system to provide communication, navigation, or any other service of interest (Army Space Reference Text 1993):

- The space segment consisting of all satellites required for the application and the launch vehicles used to deliver those satellites to orbit.
- The command segment consisting of all the personnel, facilities, and equipment that are used to monitor and control all the assets in space.
- The user segment consisting of all the individuals and groups who use and benefit from the data and services provided by the payloads of the satellite and the equipment that allows this use (Fig. 1).

Onboard each satellite, the connection between the spacecraft and the command segment is achieved by the telemetry, tracking, and control (TT&C) subsystem. As can be deduced from its name, this subsystem has three specific tasks that must be performed to ensure the ability of the satellite to successfully achieve any application:

1. Telemetry. The collection of information on the health and status of the entire satellite and its subsystems and the transmission of this data to the command segment on the ground. This requires not only a telemetry system on the spacecraft but also for a global network of ground stations around the world to collect the data, unless, of course, the application satellite network includes intersatellite links that are capable of relaying the data to a central collection point.
2. Tracking. The act of locating and following the satellites to allow the command segment to know where the satellite is and where it is going. Again this requires a ranging system on the spacecraft and a data collection network on the ground that allows this ranging and tracking function to work.
3. Control. The reception and processing of commands to allow the continuing operation of the satellite in order to provide the service of interest. Again a ground system is required.

These tasks must be performed for both of the major components of the satellite: the payload and the spacecraft bus. Earlier chapters detailed the principles and technologies related to specific payloads and applications, and the previous chapter introduced the key subsystems present in any spacecraft bus required to support the

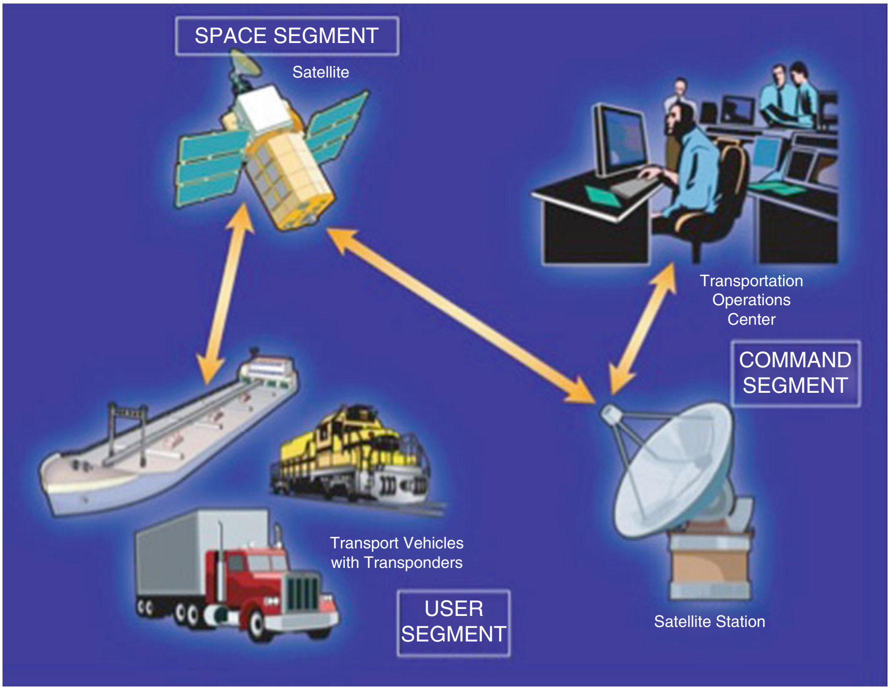


Fig. 1 An example of the three segments of a space system (Adapted from energy.gov)

payload (structure, attitude control, power, thermal controls, etc.). The TT&C subsystem must be able to gather information from and transmit commands to all of the subsystems in the spacecraft bus as well as to the payload itself. The following sections in this chapter focus on the principles and factors involved with each of the three tasks that are performed by this subsystem.

The design of TT&C systems becomes more and more complex as one moves from a single geosynchronous satellite operating over a single country to a global geosynchronous network involving a number of such satellites. The most complex and demanding TT&C network involves the operation of a global network of satellites in medium earth (MEO) or low earth orbit (LEO). This is because there are many more satellites in orbit and the orbital trajectories of the satellites are more complex. The relay of data from such MEO and LEO systems is more demanding in almost every respect. The particular strategies that are used to cope with TT&C operations with MEO and LEO constellations are described below.

Telemetry: Providing Health and Status Updates for the Satellite

It is the task of the command segment located on the ground to provide the commands that will keep the satellite operating as required. However, before any commands can be issued or even chosen, the team on the ground must know what the status of the satellite is at any given time. Telemetry is the collection of measurements and onboard instrument readings required to deduce the health and status of all of the subsystems on the satellite. The TT&C subsystem must collect, process, and transmit this data from the satellite to the ground.

The first step in providing status updates to the ground is the collection of the measurements required by the command segment. Measurements related to the health and status of the satellite include:

- The status of resources (e.g., propellant supply and the health and charging status of batteries)
- The attitude of the satellite (e.g., the readings from sun and star trackers or RF tracking systems)
- The mode of operation for each subsystem (e.g., the on/off state of a heater)
- The health of each subsystem (e.g., output from the solar panels)

These measurements are not only necessary for the spacecraft bus but also for assessing the health of the payload. On a communications satellite, the telemetry data would include information such as the switching configuration for the routing of signals, power output of the transponders, the direction the antenna is pointed in, or the health and status of imaging systems. All of these measurements are collected with various sensors such as thermometers, accelerometers, and transducers that provide outputs in such forms as measured resistance, capacitance, current, or voltage. The design of a spacecraft for the collection of data from such physical sensors and the associated wiring required for gathering this information concerning the health and status of the spacecraft and payload can lead to a noticeable mass and cost penalty. For example, some larger communications or remote sensing satellites can have up to 500 temperature sensors onboard the spacecraft. The collection of data from such a large number of sensors can lead to an extensive wiring harness (500 temperature sensors × 2 wires = 1,000 wires). This has led to ongoing research related to new alternatives for collecting information such as the European Space Agency's Fiber Optic Satellite (FOSAT) project which uses fiber optics rather than conventional wiring in order to more efficiently gather data on the health of a satellite (Fig. 2).

The use of these sensors to gather the required measurements is only the first step in providing telemetry from the satellite to the command team on the ground. The second step is the processing of the measurements. This processing includes the conversion of analog measurements to digital information as well as formatting of all the measurements for effective, and if required redundant, transmission to Earth. The processing of telemetry data involves two key factors. These two factors involve the nature of the automation patterns of the spacecraft and the data storage algorithms. These will typically be different for each particular application satellite network. These will be based on the specific applications and mission parameters for each application satellite system.

Automation refers specifically to the ability of the satellite to interpret and respond to the telemetry measurements without interaction with the command

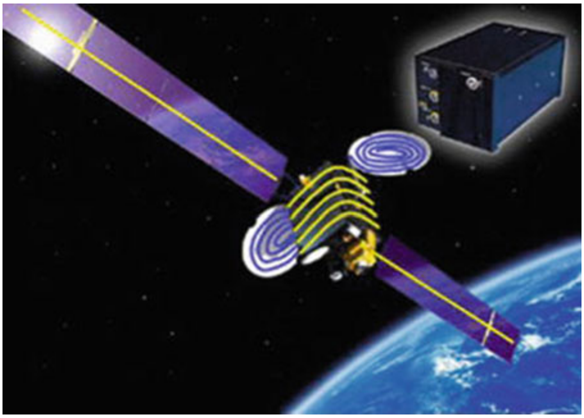


Fig. 2 Artist depiction of FOSAT using fiber optics for health monitoring (Courtesy of ESA) (ESA: FOSAT – Fiber Optic Sensing Subsystem for Spacecraft Health Monitoring in Telecom Satellites, http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=28652) (Ecke et al. 2001)

segment. This allows the satellite to issue commands to the subsystems directly versus receiving and processing the commands from the ground. Automation can typically be found in response to predicable or common faults in certain subsystems that require actions such as placing components on standby when they operate outside of a range of parameters. Automation allows instantaneous actions to be taken onboard of a satellite and the degree of autonomous operation can vary widely among various applications satellites and the sophistication of their software. Automation can lead to three particular requirements for the TT&C subsystem as noted below (Pisacane 2005):

1. First and foremost is the ability of the onboard software to properly identify telemetry that indicates a subsystem is acting incorrectly and the corresponding ability to identify and process the correct response.
2. Closely associated with the first degree of automation is the related ability for the telemetry and the command components to communicate with each other and pass information back and forth on a nearly instantaneous basis.
3. Finally, it is important for the software to have a diagnostic capability that allows the telemetry system to determine if an abnormal reading is caused by another subsystem or by errors in the TT&C subsystem itself.

Data storage of telemetry may also be required as transmissions down to the ground may not be feasible at any given moment. It is extremely costly (prohibitively so) to establish enough ground facilities for a global satellite system deployed in low earth orbit (LEO) to have constant contact with the command team. The exception would be in the case that there are intersatellite links onboard all satellites that would allow the relay of telemetry data and commands via these links. Due to the difficulty of continuous access the satellite software and onboard computers must be able to process and store the data from the sensors on board while waiting for a viable window of communications to open up. The instrument readings and measurements are then transmitted along with other pertinent information such as when each measurement was collected. The cumulative data and exactly when it was collected are essential information for the ground team's data analysis particularly in case of anomalies.

An alternative to data storage is the use of a constellation of satellites that can relay telemetry from any given satellite to specific locations on Earth. One example is NASA's Tracking and Data Relay Satellite System (TDRSS) that consists of nine on-orbit satellites that relay communications from any other LEO satellite to its ground segment known as the White Sands Complex. This system can also be used for providing uplinks necessary for issuing commands to a satellite. European and Japanese systems have developed similar capabilities.[1]

The final step the TT&C subsystem must perform in providing telemetry from the satellite to the command or ground segment is transmitting the data to the Earth. The principles, concepts, and hardware used for transmitting this type of data such as processing, commutation, multiplexing and antenna, and transponder design are described in chapters "▶ Regulatory Process for Communications Satellite Frequency Allocations," "▶ Satellite Radio Communications Fundamentals and Link Budgets," "▶ Satellite Communications Modulation and Multiplexing," "▶ Satellite Transmission, Reception, and Onboard Processing, Signaling, and Switching," "▶ Satellite Communications Antenna Concepts and Engineering," "▶ Satellite Antenna Systems Design and Implementation Around the World," "▶ Satellite Earth Station Antenna Systems and System Design," "▶ Technical Challenges of Integration of Space and Terrestrial Systems," and "▶ Overview of the Spacecraft Bus."

The communications system used for downlinking telemetry may be the same system as that used for communicating the payload data or it may be an independent system depending on the satellite's application. Typical frequencies for the telemetry system include: S-band (2.2–2.3 GHz), C-band (3.7–4.2 GHz), and Ku-band (11.7–12.2 GHz). Other frequency bands can also be employed for different types of application satellite systems. Telemetry communications tend to have a bit-error-rate of approximately 10^{-5}. Telemetry systems that utilize Ku-band frequencies need to make some allowance for rain attenuation in the design of the TT&C (Larson and Wertz 1999).

Tracking: Locating and Following the Satellite

In order to communicate with a satellite, whether it is to receive telemetry or send commands, the command segment must be able to locate, range, and track a satellite accurately. These ranging functions are part of the task of tracking which is performed by the TT&C subsystem. The satellite must first be able to locate and lock onto transmissions between the ground station and satellite. Once the satellite is

---

[1] NASA Space Communication: TDRSS https://www.spacecomm.nasa.gov/spacecomm/programs/tdrss/system_description.cfm

locked on, the TT&C subsystem determines the range, or lineofsight distance between the satellite and the radial velocity of the satellite. This allows the command segment to know where the satellite is and where it is going.

The process of locating and locking onto a satellite from a ground station is known as carrier tracking. This is most commonly accomplished in application satellite operations by using a principle known as phase coherence. This involves creating what is called a "two-way-coherent." The typical operational mode in this respect involves establishing the downlink communication frequency at a predetermined ratio of the uplink frequency. This allows a synchronization of their phases. Initially, the satellite searches and validates a connection to the uplink frequency based on the predetermined parameters. The TT&C subsystem then implements commands to set the frequency of the downlink communications, so it is related to the uplink frequency through a prespecified ratio. This not only allows a ground station to lock onto a satellite, but it allows it to do so quickly as the expected downlink frequency is already known. There are standards for this process. For instance, the transponders tracking setting for the NASA Ground Spaceflight Tracking and Data Network (GSTDN) which sets the downlink/uplink ratios 240/221 (NASA Space Communication: TDRSS https://www.spacecomm.nasa.gov/spacecomm/programs/tdrss/system_description.cfm).

Determining the range between a satellite and a ground station is typically achieved through the use of tones or pseudo-code. The tone or code is modulated to the uplink frequency and when the satellite recognizes it, the TT&C subsystem adds the same tone or code to the downlink. The command segment can then calculate the round-trip time required for that tone and use that information to calculate the distance between the ground station and satellite. With the range (i.e., the distance to the spacecraft) established, the actual location of the satellite can be determined by using the pointing information of the satellite to determine the satellite's azimuth and elevation angles (NASA Space Communication: TDRSS https://www.spacecomm.nasa.gov/spacecomm/programs/tdrss/system_description.cfm).

An alternative means for determining the range also allows for the radial velocity of the satellite to be determined. This method uses the Doppler shift of the frequencies of the uplink and downlink to determine the satellite's location and velocity. As discussed in earlier chapters, the Doppler effect is the change in frequency of the transmissions caused by the relative movement between the transmitter and the receiver. When the satellite is approaching a ground station, the frequency it receives is higher that the frequency transmitted. When the satellite is moving away from the ground station, the frequency it receives is lower than the frequency transmitted. This is also true for the frequency of the transmissions going from the satellite to the ground station. One issue with using the Doppler effect to determine the location of a satellite is that there are always two locations, the true or nominal location and the virtual or mirror location that are possible at any single point in time. In order to account for this, the TT&C subsystem has to apply processing algorithms to determine which location is correct. Satellites, such as Argos, use two positioning algorithms: least squares analysis and Kalman filtering. The Doppler shift method is best applied to satellites in relatively low orbits (Fig. 3).

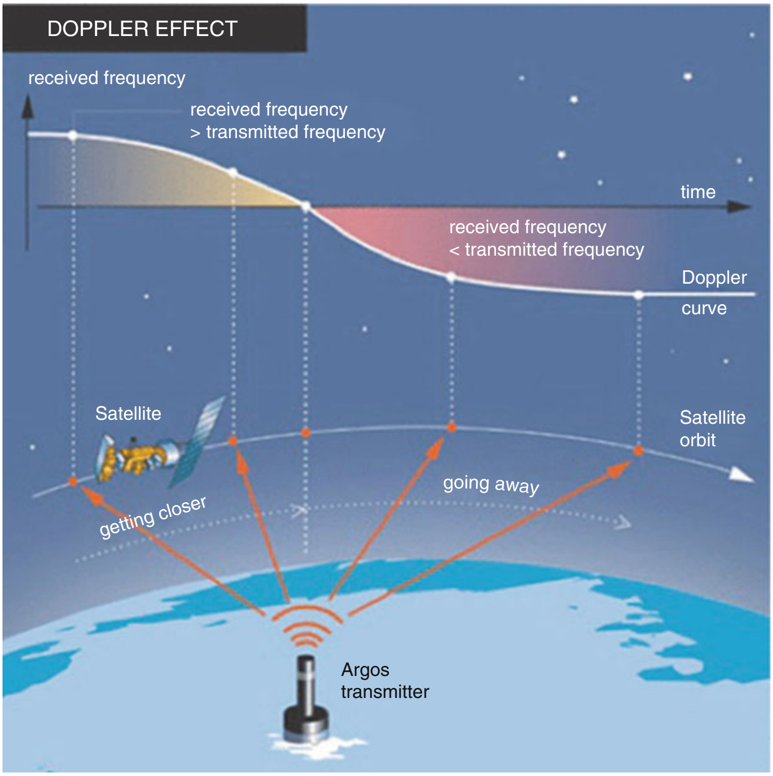


Fig. 3 Using the Doppler effect for tracking a satellite (Courtesy of http://www.argos-system.org/) (Argos User's Manual http://www.argos-system.org/manual/index.html#3-location/32_principle.htm)

Control: Commanding the Spacecraft Bus and Payload of the Satellite

Control or command is the third task of the TT&C subsystem, and it is the act of ensuring the satellite's spacecraft bus and payload do what is necessary to meet the objectives of its particular mission. Allowing control of the satellite requires that the TT&C subsystem receive, process, and implement the commands required by the command segment on the ground. As discussed briefly earlier, some commands may be automated through the use of onboard software that implements predefined commands upon recognition of specific conditions. Some satellites designed for "autonomous operation" carry this degree of automation to very sophisticated levels.

Commands are used to reconfigure a satellite or its subsystems to respond to mission conditions. Commands may include switching subsystems and components between on and off states or changing the operating mode in another manner. The commands may be used to control the spacecraft guidance and attitude control or

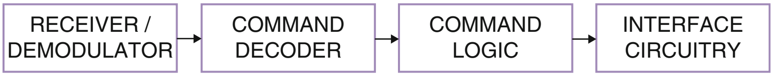


Fig. 4 Command system block diagram (Derived from information provided at http://ocw.mit.edu)

deploy structures such as solar arrays or antennas. Finally, commands may come in the form of software programs that are to be uploaded into the onboard computer to control components on an ongoing basis.

The first step of the command system is to receive the data from the ground through its communication system. This system uses the same principles and technologies as described in chapters "▶ Regulatory Process for Communications Satellite Frequency Allocations," "▶ Satellite Radio Communications Fundamentals and Link Budgets," "▶ Satellite Communications Modulation and Multiplexing," "▶ Satellite Transmission, Reception, and Onboard Processing, Signaling, and Switching," "▶ Satellite Communications Antenna Concepts and Engineering," "▶ Satellite Antenna Systems Design and Implementation Around the World," "▶ Satellite Earth Station Antenna Systems and System Design," and "▶ Technical Challenges of Integration of Space and Terrestrial Systems." Typical frequencies for the command system include: S-band (2.2–2.3 GHz), C-band (3.7–4.2 GHz), and Ku-band (11.7–12.2 GHz). Control communications tend to have a bit-error-rate of approximately 10^{-6} (Keesee 2003).

This rate is an order of magnitude less than that noted for the telemetry communications due to the importance of ensuring that the commands issued by the ground are recognized correctly by the TT&C subsystem. Typical data rates required for command systems range from 500 to 1,000 kb/s.

Once the satellite has received and demodulated the uplink command transmissions (or a command is produced by the onboard computer), the command system includes three additional segments: the command decoder, the command logic, and the interface circuitry. The decoder reproduces the command messages and produces the lock/enable and clock signals required. The command logic validates the command and rejects it if there is any uncertainty regarding its authenticity. This logic then gets implemented through the interface circuitry that connects to the other systems on the satellite. Some application satellite systems have sophisticated security processes in place. These might require that at least some ground commands be authenticated from another location in order to be implemented. There have been instances where spurious commands (whether intentional or unintentional) have disabled application satellite networks (Fig. 4).

The command decoder collects and processes all incoming commands from sources such as the ground and the onboard computer. The decoder includes an arbitration scheme that determines how each command is given priority in the processing queue.[2] Due to the criticality of the uplink commands, they are often encrypted and as noted above might also require authentication from another TT&C

---

[2] Op cit., Wiley.

ground facility. Typical command messages include input checkerboard bits, synchronization bits, command bits, and error detection bits. The command itself includes the spacecraft's address and the command type. In virtually all operational application satellite systems both elements of the command must be verified. Command types include commands to flip a relay in a system, to pulse a piece of electronics, to change the output level of a component, or to request or send data to or from a component.

In some cases, the command may involve a whole sequence of events. In the case of the deployment of the Light Squared Land Mobile Satellite that was deployed in 2011, the 20 m antenna system did not normally deploy. In this case, a series of commands were sent to fire thruster jets in a sequence to "joggle" the spacecraft so that the antenna finally deployed.

The command logic in the TT&C subsystem must verify and validate the command. This includes ensuring that the commands are being sent to the correct spacecraft or that the command itself is valid. Additionally, the timing of the command must be valid and the command itself must be authenticated. Once the logic is used to process the command, the TT&C subsystem activates the interface circuitry as necessary depending on the type of command. In the case of trouble shooting or failure-recovery operations, there may be a need to override constraints in the onboard software to allow higher risk commands to be executed.[3]

TT&C System Design Aspects

Satellites must be designed to operate correctly for the entire life of their mission. The TT&C subsystem is a critical part of ensuring that the satellite performs as required and can react to changes in conditions at the satellite (either internal or external). Because of this, the system must go through stringent quality control and testing of its components before they are used on a satellite. Additionally, key portions of the subsystem are designed to have redundant components to ensure that if one part fails, another is still available for use. If the satellite is designed for a 15-year time for instance, the TT&C subsystem might even be designed with a mean time to failure of 18–20 years.

One of the major trade-offs in designing a TT&C system is how complex the system must be to meet the goals of the mission. The more complex the system, it will be able to provide more telemetry and process more commands. The disadvantage is that this complexity typically leads to more components and therefore more mass and cost to the overall spacecraft. Most of the complexity of a TT&C subsystem is actually contained in the software that includes diagnostics to determine if a particular pathway in an onboard switch is somehow defective. The good news is that one can often update or reengineer software so that improved software can be uploaded after launch.

---

[3] Op cit., Keesee.

Connected to the concept of complexity is the aforementioned use of automation. A design must look at the requirements for the mission and determine how much telemetry and command should be dealt with onboard the satellite and how much should flow through the command segment on the ground. A thorough review of how the processing scheme for each command impacts the overall mission reliability and overall cost should be performed prior to launch and reviewed periodically to see if new software could help ensure improved and more reliable performance. In some cases, it may be more cost-effective to deal with situations through onboard computing, while other situations may be more cost-effectively dealt with through analysis on the ground.

Conclusion

The telemetry, tracking, and control subsystem enables the critical connection between a satellite and the ground segment. Regardless of the application of the satellite in question, the TT&C subsystem must perform all of its tasks in order to have a successful mission. The three major tasks of the subsystem are:

1. Telemetry. The collection of information on the health and status of the entire satellite and its subsystems and the transmission of this data to the command segment on the ground in an accurate and consistently reliable manner.
2. Tracking. The act of locating and following the satellites to allow the command segment to know where the satellite is and where it is going and with a high degree of precision. Geosynchronous satellites that can be 40 time further out in space than low earth orbit systems require more exacting methods to determine range because of the greater distances involved.
3. Control. The reception and processing of commands to allow the continuing operation of the satellite on an uninterrupted basis. Protection of the control commands to prevent spurious commands is just one of the elements that designers of satellite networks must take into account.

The next chapter discusses the importance of lifetime testing, redundancy, and reliability. Each of these aspects should be taken into consideration when designing a TT&C subsystem.

Cross-References

- ▶ Common Elements Versus Unique Requirements in Various Types of Satellite Application Systems
- ▶ Fundamentals of Remote Sensing Imaging and Preliminary Analysis
- ▶ Introduction to Satellite Navigation Systems
- ▶ Lifetime Testing, Redundancy, Reliability, and Mean Time to Failure
- ▶ Overview of the Spacecraft Bus

- ▶ Satellite Communications Antenna Concepts and Engineering
- ▶ Satellite Communications Modulation and Multiplexing
- ▶ Satellite Radio Communications Fundamentals and Link Budgets
- ▶ Satellite Transmission, Reception, and Onboard Processing, Signaling, and Switching

References

Army Space Reference Text (US Army Space Institute, Fort Leavenworth, 1993), www.fas.org/spp/military/docops/army/ref_text/index.html

W. Ecke et al., Fiber optic sensor network for spacecraft health monitoring. Measure Sci. Technol. J. **12**(7), 974 (2001)

C.J. Keesee, *Satellite Telemetry, Tracking, and Control Subsystems* (Massachusetts Institute of Technology, 2003), http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-851-satellite-engineering-fall-2003/lecture-notes/l20_satellitettc.pdf

W. Larson, J. Wertz, *Space Mission Analysis and Design*, 3rd edn. (Microcosm Press, Hawthorne, California, 1999)

V. Pisacane, *Fundamentals of Space Systems* (Oxford University Press, New York, 2005)

Lifetime Testing, Redundancy, Reliability, and Mean Time to Failure

Joseph N. Pelton

Contents

Abstract

The environment of outer space is quite hostile to the many spacecraft that are now deployed in Earth orbit and beyond. There are many hazards in terms of severe thermal gradients, space weather from the sun and beyond, and intense radiation from the Van Allen belts as well as strong magnetic forces. Today, application satellites also must plan to cope with man-made hazards that arise from space debris, RF interference (RFI), and other possible hazards such as spurious commands. There are also risks associated with the launch and

---

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_70

deployment of satellites since there are strong "g forces" during launch and difficulties that can arise from the unfolding, roll-out, or explosive or spring-loaded extension of solar arrays, antennas, and other systems that must be deployed in space in response to remote command. This complex series of hazards requires extensive testing of application spacecraft that are deployed into Earth orbit with the hope of extended lifetime operation. These hazards and difficulties of space operations increase the importance of lifetime testing. It also demands the design of application satellites to be rugged and to have reasonable levels of redundancy so that service can be maintained if various components happen to fail. In the case of application satellites, rugged design, redundancy, and demanding lifetime testing of applications satellites and its subsystems and components are of utmost importance simply because there is little opportunity for repair or refurbishment operations in space. Without these precautions, a very expensive application satellite that requires perhaps an even larger investment to launch it into space could be lost to the satellite operator and thus require replacement at very high cost either to the satellite operator or to the companies that have insured the launch and operation of the satellite.

In recent years, there has been an alternative approach taken in terms of deployment of large constellations of small satellites in space as an alternative to a few large satellites designed and tested for long life. These small satellites have been built at much lower cost using off-the-shelf components and most frequently by advanced manufacturing techniques that include 3-D printing. These have frequently been launched as "piggyback" missions and thus at much lower cost.

Networks such as Skybox Imaging, Planet Labs, PlanetiQ, Dauria Aerospace, Tyvak Nano-Satellite Systems, NovaWurks, and GeoOptics have all emphasized this approach that involves miniaturization, low-cost satellites, and associated modest launch costs over larger and more capable satellites that subjected to extensive lifetime testing prior to launch. This new paradigm is also now being tested by new communications satellite operators such as OneWeb that proposed to nearly 800 mass-produced satellites plus spares to create a network optimized for Internet-based services, and a megaLEO constellation by SpaceX might ultimately involve thousands of small satellites. For this type of alternative design architecture, the replacement of failed satellites with a ready supply of spares is the key to achieving system reliability. This approach is seen as the alternative to stringent testing and flight-qualified components with proven long-life capabilities in a stringent space environment.

The following text discusses all of these strategies for coping with and minimizing risk for the satellite application industry although the much greater emphasis is on the stringent reliability and long-life design approach, since the ventures employing a constellation of small satellites largely depend on a robust sparing effort.

Keywords

Accelerated testing • Acoustical testing • Anechoic chamber testing • Autonomous operation • Constellation of small satellites • Deployment risks • Inclined orbit operation • Independent verification and validation (IV&V) • In-orbit incentives • Launch insurance • Launch risks • Lifetime testing • Mean time to failure • Pogo effects • Radiation redundancy • Reliability • RF interference • Robust sparing of small satellite constellations • Satellite operational and launch insurance • Space debris • Subsystem testing • Thermal vacuum testing • Van Allen belts

Introduction

This chapter addresses the various strategies that operators of application satellite systems can undertake to increase the reliability of their space assets, ground systems, and overall system operations as well as to manage risk against various types of losses that can occur. These strategies include the following: (1) constant identification of and knowledge about various types of risks and hazards – natural and man-made (i.e., space debris, conjunction of other spacecraft, and RFI); (2) design of space and ground systems to withstand these risks, to have significant link and operating margins, or to have redundancy or spare components to restore failed systems; (3) design to eliminate as many single points of failure as possible; (4) carry out sophisticated testing strategies to identify weak or flawed elements against manufacturing mistakes, and to check space systems against a lack of tolerance to mechanical, vibrational, electronic, radiational, thermal, RF interference, power outages, as well as conduct deployment tests for antennas and solar arrays (this is done either through component, subsystem, or fully integrated satellite testing); (5) provide for independent expert oversight of manufacturer design and testing; (6) write contractual provisions for incentives for reliable performance to manufacturers – including in-orbit incentives for successful operation; (7) provide for various forms of launch and operational insurance; (8) employ extensive computerized monitoring systems to track spacecraft health and operational parameters and where possible to evolve toward Autonomous Operation of spacecraft where onboard artificial intelligent or expert systems can anticipate problems and maintain maximum reliability of operation; (9) take evasive action and precautions such as avoidance of known space debris, powering down of spacecraft when major solar flares are ejected from the sun; (10) design of ground systems with a high level of redundancy, backup power, security codes against spurious commands, backup tracking, telemetry, command and monitoring (TTC&M) facilities and redundant areas of global coverage as well as constant training and education of satellite operators.

Alternatively one can deploy constellations of small satellites using off-the-shelf conventional components and rely on extensive sparing to recover from satellite failures. This approach depends on low-cost “piggyback” launches and carries with it the danger of the increasing build up of space debris – especially in low Earth orbit.

These efforts to provide for reliability, longer spacecraft and system lifetime, better management oversight, better trained personal, redundancy, and various forms of risk management including insurance coverage almost always add to cost. Commercial operators are thus constantly struggling against trying to add reliability, redundancy, and lifetime while minimizing cost. Different satellite system operators have thus not surprisingly devised different strategies to minimize risk and cost while maximizing reliability. Some have put the greatest stress on design, while others have pursued accelerated or altered forms of testing, while some others have relied more on various forms of insurance or contractual solutions such as having the contractor deliver the spacecraft in orbit after check-out and verification tests.

Satellite and Subsystem Lifetime Testing

Hazards (Natural and Man-Made)

The key element in trying to design in reliability for space assets is to understand as well as possible potential risks. In space the risks are many and they vary over time. Risks from space weather include various types of radiation and solar storms that follow an 11-year cycle that are the greatest risk during solar max. The Van Allen belts between low Earth orbit and medium Earth orbit represent a major hazard to application satellites that operate in these types of orbit. Medium Earth orbit satellites that operate just above these belts, for instance, often add additional glass coating on top of the satellite solar cell arrays to mitigate the deterioration over time of the solar cell effective power output. Even geosynchronous orbit satellites although safe above the highest Van Allen belts must travel through these high-radiation zones during the launch operations. Military and defense satellite systems are often designed as so-called radiation-hardened (Rad-Hard) facilities. These “radiation-hardened” satellites have many design features to protect the electronics and power systems of these spacecraft. These designs are intended to provide protection against natural radiation as well as man-made radiation. Some systems have been designed with a “Faraday cage” to protect against the radiation and electromagnetic pulse (EMP) of a nuclear blast in outer space (Jakhu et al. 2009).

There are also systems that monitor the sun for evidence of intense solar flares. Since it takes 8 min for light to travel from the sun, but quite a bit longer for the particles from a solar storm (i.e., typically in the range from 24 to 56 h), there is often ample time to “power down” satellites and to go into “safe mode” when such a violent solar blast occurs. Most application satellites are thus designed to provide at least some degree of protection against radiation, electromagnetic disturbance, thermal extremes of hot and cold (through thermal reflectivity, heat pipes, etc.). They are also tested against the violent shaking and vibrations that occur during rocket launches, including the so-called pogo effect or oscillations that can occur between the various stages of a launch vehicle and the fairing structure that contains the payload. The design is also constructed so that the antennas of the spacecraft can be tested on the ground to verify their gain and beam shaping capabilities and their electronics and filters tested to screen out unwanted RF interference and background noise. The most recent hazard that has become of concern to application satellite operators is the increasing amount of orbital debris – particularly in low Earth orbit. Currently the US Space Command is monitoring on the order of 20,000 pieces of orbital debris which is of the size of a human fist or larger, and there is an estimated 500,000 pieces of orbital debris of the size of 1 cm or more. Such debris, moving at super high orbital speeds can still do significant damage. Efforts have now been made to create data centers whereby satellite operators in GEO orbit such as Intelsat, SES Global, and Inmarsat can share data about their satellites' orbits and times of close conjunction of their spacecraft. This coordination process continues to expand and soon it will cover spacecraft operation in the lower orbits.

Testing Strategies

The initial testing strategies for verifying the reliability and projected lifetime of application satellites were those techniques developed by the early Intelsat system and the manufacturers of their spacecraft. Intelsat, and their early system manager, Comsat, decided to rely on a combination of techniques rather than a single strategy. Components, subsystems, and fully integrated satellites were subjected to vibration and acoustical testing, thermal vacuum tests, accelerated lifetime testing of units such as spacecraft batteries, etc. These tests were carried out by satellite manufacturers, but Intelsat engineers monitored the tests and provided oversight of the design, engineering, and testing process. Tests were also carried out to test the performance of antennas and RF electronic systems in anechoic chambers and testing range sites. Intelsat also provided for increasingly more comprehensive launch insurance against launch failure. It also structured its contracts so that the manufacturer was paid so much against delivery milestones, but it reserved a large final incentive payment after the satellite was launched and had performed successfully in orbit. The reasoning was that the manufacturer thus had not only an encouragement to build the spacecraft but to see that it operated successfully in space. Experience over time showed that if a spacecraft could operate successfully over the first month in orbit, it would likely remain successfully in operation until worn-out components begin to fail toward the end of the satellite's life. This lifetime projection curve is sometimes called the "bathtub curve" because failures occurred swiftly at the beginning, and then there was a smooth steady state for years, followed by a rapid increase of failures at the end of life – much like the shape of a traditional bathtub.

The first satellites had only limited lifetime expectancies in the range of 18 months to 5 years. Over time as the satellites became larger, more sophisticated and with lifetimes up to 15 years and longer, more elaborate and demanding tests were developed. The testing process often comprised 20–35 % of the total spacecraft manufacturing cost. With accumulated experience, the satellite industry and the satellite manufacturers began to think of different strategies for lifetime testing and for reliability engineering.

New Concepts

As new satellite operators began to deploy satellite systems, the quest to find ways to deploy reliable satellites while containing costs as much as possible began in earnest. Some satellite operators deserted their own research and development programs and abandoned the use of oversight engineers and thus left all reliability concerns to the manufacturers. Some moved to buying their satellites based on a proven spacecraft bus series. In this process, they benefited from the economies of scale since the initial nonrecurring design and engineering costs had already been recovered. Further they thought that if they were buying not the first three of a satellite series but instead buying three satellites very much like the ones that had already been manufactured more than a dozen times, then weaknesses of design or component manufactured would have already been corrected.

Other concepts were more daring. The Iridium satellite system which was a large constellation with some 66 operational satellites, a dozen spares, and a final production run of some one hundred satellites decided that it could streamline its engineering, manufacturing, and manufacturing process by designing in the quality of its component and component testing and then carry out an accelerated testing program. At the end of the Iridium manufacturing process, a complete satellite was being produced in less than 5 days. This stood in contrast to large geosynchronous satellites that might be in production and testing for well over a year and sometimes over 2 years. Although there were a number of early satellite failures due to mechanical and electronic failures as well as operator errors, the final Iridium constellation was able to achieve a combined network lifetime reliability record that exceeded over 500 satellite years in orbit.

Increasingly the satellite industry has come to rely on the satellite launch and operational insurance industry to provide a key element of its risk management strategy. Initially the insurance and risk management companies were reluctant to insure satellite launches. This was because launchers were much less reliable than today, the cost of the launchers and satellites was high, and the industry had little experience with this type of high risk coverage. Initially, insurance coverage only applied to two launches in succession and the premiums were high. By the 1980s, however, the launchers were more reliable and the insurance industry had become more familiar with space industry practices. Thus the reinsurance process was able to spread the risk over many different companies so that the exposure by any one company was comparatively low. Organizations with lots of launches, good technical oversight, and a good track record were able to get launch insurance for any particular launch for as low as around 7 % or 8 % of the risk exposure. Smaller organizations with fewer launches without a known track record of course paid higher rates. Today a typical launch operation requires satellite operators to pay in the range of 15–20 % of amount seen as the “total risk exposure” for an application satellite launch.

Organizations insure not only against a launch failure and the loss of a spacecraft, but they also take on liability insurance against some form of catastrophic loss such as an event where a launcher goes off course and lands in a city with a huge loss of life and property. In many cases, governments that host launch operations provide some level of liability coverage against such a catastrophic loss above a level such as $100 million. Some organizations today simply plan that they will pay a premium of up to 20 % for insurance against a launch failure and loss of a spacecraft. This is seen as particularly prudent in cases where the entire satellite network consists of only one, two, or three spacecraft. A few organizations such as Intelsat have felt when premiums seem at their highest to self-insure against a launch failure. It is possible to insure against other risks such as a “crash between satellites” or orbital debris. The cost of this type of insurance was once quite modest such as $50,000 per satellite per year, but these premiums have risen sharply in recent years.

Satellite Lifetime and Mean Time to Failure

The world of satellite reliability largely operates in the domain of collective probabilities. The projected “mean time to failure” (MTTF) comes from combining multiple risk factors based on assessed risk of components and subsystems. If a satellite has 1,000 parts each with a 0.999 assessed reliability for 7 years, this might appear to constitute a very high overall reliability. The combined risk, however, falls significantly when the 1,000 component parts are combined to calculate a cumulative rate of failure. The actual risk assessment is far more complicated because some components are more reliable than others and have a proven track record in space. Some other elements may constitute a single point of failure, while others may be backed up by redundant components.

In such cases, it may be straightforward to replace a failed power amplifier by simply switching over to a backup amplifier to continue reliable operation. In the case of a solid-state power amplifier that is very light in mass, redundancy is a reasonable design choice. In the case of a momentum wheel for stabilization of a spacecraft, the cost, weight, and complexity of design lead to this key subsystem to be a single point of failure. The process of reliability assessment based on component and subsystem design specifications produces a bell curve of projected times of failure.

The mean time of these various projections is when the satellite is seen as most likely to fail. A satellite with a mean time to failure of 15 years will have sufficient batteries, solar cells, station-keeping fuel, and backup components to last much longer than its MTTF. In fact the satellite may last for 5, 7, 10, 15, 18, or 20 years due to millions of different factors, but typically today the 15 years MTTF will be the best projection for GEO satellites by satellite designers on the basis of data available at the time of launch.

One of the key factors in satellite lifetime involves the orbit. GEO satellites can last much longer in their orbit since they are almost out of the Earth’s gravity well. LEO satellites are much harder to maintain in their orbits because the gravitational attraction is much stronger and the earth’s atmosphere is working to degrade its orbit – especially during solar max periods. In short, the orbit constitutes on the other key elements in projecting satellite lifetimes.

Key Components, Subsystems, and Lifetime Expectancies

The design of a space system for reliability has several key elements. One of the hardest challenges is to recognize those elements that represent a potential single point of failure on a spacecraft and to ensure the optimum design for those parts and effective test of those components prior to launch. Another key element is to provide for redundancy of parts or elements where it is feasible and cost-effective to do so. Yet another element is to provide for sufficient margin in expenditures or for systems that wear out or deteriorate in performance over time. Although these concepts seem straightforward in terms of designing for reliability, they prove difficult in practice.

In the case of single points of failure, some elements such as the TT&C system, the deployment of antennas, and the stabilization and pointing system are obvious and are given a great deal of attention in the design, engineering, and testing process. Where things can go wrong in this process is unfortunately very numerous. Some tasks such as electronic power converters can seem very easy because it has been done many times before. Thus, such tasks can be given to junior engineers. But if the power system fails, the entire satellite fails as was the case with power converters on the Intelsat V series. Thus one key rule is to let single point of failure “trump” assumed ease of design and lack of engineering complexity. Design reviews should consider single points of failure and not skip over simple elements like power converters just because they do not seem to present a challenge. Also it is important to recognize that two or more failures can combine to create a catastrophic failure.

The more complex a design is and the more subsystems or components included in a satellite or launch vehicle, the more likely that multiple failures can combine to create a total system failure. In the case of a US military satellite, a low-gain omni TT&C antenna was eliminated from a satellite for budgetary reasons. This satellite because of a problem due to fuel sloshing combined with satellite commands went into a flat spin (rather than a vertical spin). In this mode the satellite antennas were spinning around instead of pointing constantly toward Earth. With only the high-gain antenna available for commands, it was not possible to command the firing of jets to allow the satellite to be recovered from flat spin. This particular combination of problems and design changes resulted in a catastrophic failure of the satellite.

The addition of redundant parts that can be called into service by a command to the satellite sounds like a very good idea. The problem is that every addition to the satellite increases its volume, mass, and complexity. The challenge is to decide where the best and wisest investment in redundancy makes sense in terms of prolonging the useful life of a satellite without unduly increasing its mass and associated launch costs. Redundancy in the crucial communications electronics of a communications or navigation satellite or having backup components in the key sensors of a meteorological or remote sensing satellite makes sense in that these are key to mission’s performance and electronics and computer chips are lightweight and small in volume (Williamson 2006).

The “sparing philosophy” for application satellites is closely tied to probability analysis. At one time designers opted for one redundancy in crucial electronics. Later, based on in-orbit experience, they began to opt for one backup repeater on communications satellites for every two devices with the ability to switch to the backup device when either one of the two repeaters failed. In the case of remote sensing satellites, space navigation, and meteorological satellites, the sparing philosophy is more complicated in that there tends to be one sensor of each type rather than multiple transponders. Nevertheless, one can still provide for redundancy within the electronics associated with these sensing or transmission devices.

Finally, there is the issue of providing additional performance margin in elements of the satellite that deteriorate over time or involve expendables. In the case of power systems, it is not possible to provide completely redundant systems, but one can provide additional solar cells to support peak power requirements for the projected end of life. Likewise, it is possible to provide additional battery capacity to support power needs at the projected end of life. Also it is possible to add fuel to stabilization and positioning systems. Additional fuel can be added to let a satellite be maintained in orbit more accurately and for longer periods of time. When a satellite is constructed, it is designed against a set of mass and volume constraints with some margin as needed as it is finally engineered and manufactured. At the time of launch, there is usually some mass margin left, and at that stage additional fuel is often added up to the lift capacity of the launching system.

In the case of GEO satellites, there are also operating strategies that can be employed to extend lifetime. The gravitational pulls on a satellite in GEO orbit are ten times stronger in the north to south (i.e., latitude) direction as opposed to the east to west or longitudinal direction. Some operators as a satellite moves toward its end of life allowed a GEO spacecraft to go into a slightly "inclined orbit" that moves in an "S-shaped curve" above and below the equator while maintaining the specified and assigned longitude position. This allowed lifetime extension at quite low cost. As long as the inclination did not build up to over 5° off of the equator, this added only minor risk to the continuity of service.

Optimum Lifetime Engineering and Testing

The world of application satellite engineering is quite different from the world of human space flight. In the case of application satellites, the objective is to design and manufacture satellites of high reliability over reasonably projected lifetime without incurring excessive costs. In contrast, the "man rating" of a spacecraft has typically involved adding demanding margins on top of demanding margins and almost endless testing. In the case of the Lunar Excursion Module, the main contractor, Grumman, was required by NASA to construct ten major test articles of the complete craft. In fact, Grumman, in its zeal to produce a no-fault vehicle ended up constructing 29 test articles including complete structure, thermal and electrical models (Williamson 2006).

In the world of manufacturing application satellites, the spacecraft designers and manufacturers strive for extremely high quality and engage in extensive testing, but to do so in a cost-effective manner with only a reasonable number of tests. Current strategies used by manufacturers include the use of standardized platforms for various types of application satellites sized to different launchers and fairing enclosures. This not only reduces the engineering and manufacturing costs but also allows reduced testing after a particular platform design has been tested on the ground – and in space – a number of times.

The same is true for components such as batteries, heat pipes, thrusters, solar cells, application-specific integrated circuits (ASICs) as well as their substrates. Such component parts are increasingly standardized in design and manufacturing techniques for similar reasons. The problem standardization of space qualified and proven components in terms of their design, manufacture, and testing is that this tends to block design innovations. Once a component or subsystem has been standardized and proven in space in a number of missions, it becomes "locked in" and design upgrades become difficult. One finds anomalies such as several generations' old computer chips on spacecraft that have been "space qualified" some time ago. Thus one can find contemporary spacecraft with very slow processors even though one can find off-the-shelf processors commercially available that can work at much faster speeds and purchased at much lower costs.

The design of a reliable spacecraft thus involves a number of judgment calls. Does one use an older and well-qualified ASIC or perhaps use a redundant new ASICs that can process information four times faster? How much glass coating should be applied to solar cells to slow the effects of space-based radiation? Does one specify multiple and redundant testing of components, subsystems, and a fully integrated satellite or can some tests be skipped?

If one is deploying a large-scale constellation of LEO satellites with scores of satellites in the production process, can one eliminate qualification tests if earlier satellites have passed with flying colors? The logic of such an accelerated testing program in such a case could be that there are multiple spares being launched, and thus if one or two satellites should fail in orbit, spares can quickly replace the failed spacecraft. This thought process of significant reliance on spares to restore failed satellites can become even more predominant if the LEO constellation population is increased from around 50 to 70 such as Iridium or Globalstar to huge numbers such as hundreds to even thousands – as anticipated by the OneWeb and SpaceX constellations. This approach of manufacturing lower cost satellites with off-the-shelf components on an assembly line basis in bulk with a significant number of spares is currently unproven since this approach is just starting to be employed. In another five to ten, the implications will be much clearer in terms of reliability, cost-effectiveness, and other issues such as increasing the problem of orbital debris removal.

Strategies to Extend Satellite System Lifetime

As noted in the previous section, different types of systems can be designed for different levels of reliability and projected lifetimes based on the orbit used, the type of network deployed, and how it is operated. GEO satellites are much more expensive to launch; it can be designed for longer life because much less gravitation pull degrades their orbit, in contrast to LEOs; and far fewer of them are needed to deploy a viable operating network. In fact, one GEO can create a fully operational system for a country, a region, or even transoceanic service. For all of these reasons, these satellites tend to be designed for the highest levels of reliability and the longest lifetime. Key strategies that are available for lifetime extension include (1) inclined orbit operation, (2) sparing philosophies (especially for large-scale LEO constellations), and (3) constellation deployment schemes (including “piggyback” opportunity launches of small satellites).

Inclined orbit operation was briefly described above as a way to extend the life of GEO satellites and does not apply to MEO or LEO constellations. This mode of operation is simply a strategy to conserve onboard station-keeping fuel. This is a viable strategy for extended lifetime operation only if the satellite is otherwise functional in terms of its power systems, antenna and electronic capabilities, and TT&C systems. Operators of GEO satellite networks are assigned locations in the orbital arc between 0° and 360° in the equatorial plane, and they are expected to maintain their satellites within a half degree east or west of their assigned location. Excursions in the north or south direction (i.e., latitude) are not as rigidly controlled.

Operators are normally expected to maintain their satellites within 5° north or south of the equator. The tilt of the Earth’s axis and the gravitational forces of the Sun and Moon, however, make it challenging to keep a satellite exactly on the Earth’s equator. Relaxing the “box” within which one seeks to keep a GEO satellite and allowing inclination to build up can add years to the end of life for the spacecraft. The alignment problem for Earth stations pointed to such a GEO satellite becomes most difficult for those in the subsatellite location in equatorial countries, especially where narrow spot beams are in use.

There is a simple mechanical device that can be added to the pointing mechanism on all affected Earth stations. This mechanism allows “tracking” of the satellite as it moves above and below the equatorial plane on a 24-h-a-day cycle. Organizations such as Intelsat, SES Global, etc., have used this technique to extend the lifetime of their GEO satellite by 1, 2, or 3 years. Usually other components of the satellite die out, such as the power system, and thus end the spacecraft life. Nevertheless, these types of inclined orbit satellites can be used as emergency spares or to provide service until replacement satellites can be launched. With today’s increased concern with orbital space debris, there is currently increased concern that GEO satellites retain at least a sufficient amount of fuel in their tanks to raise them out of the GEO orbital arc.

Sparing philosophy addresses not the reliability of a satellite but the ability to restore service in case a particular satellite should fail. In this regard, different types of applications have different degrees of service standards, and LEO and MEO constellations have different requirements than GEO systems. Telecommunications satellites and space navigation satellites have the highest requirements for continuity of service in that even the slightest moments of outages can have the most significant requirements. Real-time communications satellite networks seek to achieve Integrated Service Digital Network (ISDN) standards of 99.98 % reliability that allow for only about 100 min of outage in a year’s time.

Space navigation satellite systems seek an even higher standard in that systems such as the Global Positioning Satellite system is used for such vital functions as assisting with the takeoff and landing of aircraft. Telecommunications and space navigation satellites have live spare satellites in orbit with the ability for fast and hot switch over from an operational satellite to a spare. There are also mutual aid working group (MAWG) procedures that allow for rapid transfer of telecommunications traffic from fiber-optic systems to satellites and vice versa. Soon there will be a varied of space navigation satellite service (SNSS) networks in orbit that can mutually reinforce one another for vital navigational services.

In essence these various networks such as the US GPS, the Russian GLONASS, the Indian Regional Space Navigation System, the Compass/Beidou System of China, the Quasi-Zenith System of Japan, and the Egnos/Galileo System of Europe will, in time, work to mutually reinforce one another (although each will have their own sparing capabilities and sparing philosophy). These attempts to coordinate the interoperability of these various systems are being carried out through the International Committee on Global Navigation Satellite Systems (ICGNSS). These systems, working in tandem, will essentially allow the entire collection of space navigation satellites to achieve 100 % availability (UN Office of Outer Space Affairs).

Remote sensing, Earth observation, and meteorological satellite systems, of course, also seek a high degree of continuity of service, but because of cloud cover, the need for processing and "ground truthing" of data, the demand for absolute and total system availability, and thus the need for extensive "live sparing" of satellites in orbit are less than the case with telecommunications and space navigation systems. The various systems that exist around the world help to supplement one another and provide a reasonable degree of backup, especially in the case of a satellite failure among governmentally operated systems. In short remote sensing and meteorological satellites tend to be launched on a scheduled replacement timetable. If there are short periods of time where a system is not fully populated, then sharing of data among various international systems tends to provide a reasonable degree of backup. The demands for more and more extensive data monitoring related to climate change are altering this picture, and an ever larger number of climate change monitoring satellites are planned to be launched in the next few years.

Satellite Design and Redundancy

Crucial systems engineering and optimization decisions are made at the outset when various types of telecommunications, space navigation, remote sensing, and meteorological satellites are being designed. Initial decisions are made as to the expected performance and capabilities the satellite will have and the anticipated mean time to failure objective set. GEO orbit satellites today tend to have lifetimes in the 12–18 years range. MEO satellite constellations are roughly in the 10–15 years range, and LEO satellites because of gravitation pull and atmospheric drag have the shortest lifetime of only 8–10 years. There are exceptions to the above lifetime ranges for application satellites. Nevertheless, these are "normal" lifetime expectancies based on the design of power systems, fuel for thrusters, and other systems that degrade in performance over time. Only small incremental investments in terms of batteries, solar array cell size, fuel for orientation and station-keeping thrusters, and redundant transponders or electronics are needed to extend the lifetime of a satellite.

One might, therefore, assume that operators of application satellite systems would automatically design for the longest possible lifetime. This, however, is not the case. The reasons "against lifetime extension" are severalfold. Succinctly stated these reasons are:

1. *Obsolete technology.* Satellites are very much like specialized digital computers in orbit. The resolution of sensors improves. ASIC chip technology races ahead. Solar cells become more efficient. In light of rapid technical innovation in this field, a satellite can become obsolete in less than a decade. Lifetime extension in such conditions becomes a questionable proposition.
2. *Unanticipated failure modes*. One can add a number of elements that can potentially add years of life to a satellite. The problem is that there can be unanticipated failures due to a violent solar flare, an electric power connection to a momentum wheel, etc., that turns into a single point of failure that disables the entire satellite. The addition of fuel, longer-lived power systems, and redundant transponders do not guarantee longevity in the harsh environment of space.
3. *Higher launch and satellite costs*. Additional redundancy or add-ons to extend satellite lifetime can lead to higher launch and satellite costs. Satellite operators, with an eye to profitability, might be willing to add fuel up to the launcher's capacity or might even add some redundant light weight computer chips or solid state amplifiers, but this is about the limit. The investment in additional batteries or solar cell arrays is seen as adding unnecessarily to the mass and cost of the satellite and thus also adding to the launch cost. System designers look at features such as lifetime extenders or capabilities such as inter-satellite links as an "opportunity cost." In short one can invest in features such as more batteries, solar cells, or inter-satellite links, but this prevents the "opportunity" to have a bigger payload for remote sensing, meteorological imaging, telecommunications services, or higher power space navigation signals.

Some believe that, particularly with low Earth orbit constellations – where many dozens of satellites are launched to complete a system – the key to extending system lifetime is in the sparing philosophy. The concept is to launch a number of operational spares and to have more on-ground spares that can be launched as needed. In this scenario, fuel can be added but the lifetime extension is largely accomplished through sparing philosophy. This, in any event, was largely the approach utilized with the Iridium and Global Space Mobile satellite systems in the cases of their LEO constellations. The point is that in designing satellite systems and planning for their reliability, redundancy, sparing philosophy, and lifetime extension concepts, a wide range of options are considered against cost optimization formulas. This economic optimization involves not only the design of the satellite themselves in terms of capacity, performance, reliability, etc., but also the optimization of the orbital configuration that is used in terms of orbital height, number of satellites and spares, and special requirements such as inter-satellite link (i.e., cross-link) communication.

Redundancy and Single Points of Failure

After the first level of system design and optimization is completed that results in an initial satellite and system network design, the second level of analysis focuses on what are called “single points of failure.” A three-axis body-stabilized spacecraft has a single momentum or inertia wheel that spins at thousands of revolutions per minute to keep the satellite constantly pointed toward Earth. Redundancy of this massive system is not cost-effective to consider, and thus great pains are taken to ensure that this system is highly reliable, tested on the ground and on previous satellite missions, and engineered so that it never loses power. The deployment of the satellites’ solar arrays and the communications and/or TTC&M antennas, the release of the satellite from the launch vehicle, and a number of other possible failure modes are identified as ways in which a single mishap can completely end the mission.

The experience of launching application satellite over the past half century lends important insight and is highly instructive in identifying these critical failure modes. Electronic power converters, exploding batteries, stuck bearings, solar storms, antennas, or solar cell arrays that will not deploy, loss of a command channel to a satellite, overheating or freezing of a satellite’s electronics, and even miscalculation as to whether a launcher is deploying one, two, or many satellites can represent the difference between success and mission failure. These various critical mission systems, components, or operation are thus given intensive attention at the design, engineering, manufacturing, qualification testing, and deployment and operation stages. One of the key elements of independent verification and validation is the sharing of data through knowledge-based information (KBI) networks so that lessons learned from earlier failures or problem recovers can be shared. Thus, standards for design and testing come from experience gained from application satellite programs around the world. Each time a satellite is manufactured, launched, and operated, it adds to the knowledge about vulnerabilities and ways to make future satellites more reliable.

TTC&M

The payload that provides communications, space navigation, remote system, or meteorological imaging is, of course, why an application satellite is launched. Nevertheless, it is the spacecraft bus that supports the successful operation of the spacecraft payload. In terms of an analogy, the platform or spacecraft bus is like the “school bus” that allows the delivery of its passengers to the right location in good health. If the “bus” breaks down or malfunctions, the “critical services” do not get delivered. The key means by which the “bus” is able to operate successfully is via the tracking, telemetry, and command systems that allow the operators on the ground to know where the satellite is, whether the spacecraft is oriented in the correct position, whether the power and thermal systems are performing correctly, or whether there are any subsystem anomalies or failures. The command system allows thrusters to be fired or backup components to be activated in case of electronic or mechanical failures.

In many satellites, there is also a "monitoring" function that checks that the payload (as opposed to the platform) is performing correctly and is not experiencing interference, system overload, or pointing errors. Some of the most advanced satellites have been designed with expert systems and/or artificial intelligence to allow the satellite to engage in "autonomous or semiautonomous operation." Such satellites are thus more independent of ground control and troubleshooting capabilities. For most application satellites, however, the loss of TTC&M function – even for a short while – can spell the death of a satellite. This is why the design and operation of the TTC&M subsystems to have continuous access from the ground control centers to the spacecraft's TTC&M antennas remains so critical. This required to always stay connected – either to ground controls or computer monitors – which is why there are higher gain TTC&M antenna systems plus backup "omni antennas" that can be commanded from any angle in order to help recover from outage with the higher gain antenna systems.

There is another type of concern involving satellite commands. This is the problem of "hackers" or other organizations or people with hostile intent sending spurious commands to application satellites that would temporarily or even permanently disable them. To protect against unintentional or intentional sabotage of satellites, most operators have not only special codes that must be employed but also require that the properly encoded commands be sent from at least two rather than a single location. There have been recorded cases of "hacker attacks" against commercial application and military satellites to send them commands that resulted in their loss of pointing orientation or to power down their operation. Fortunately, these have resulted in temporary loss of service and no permanent satellite failures.

Autonomous Operation and In-Orbit Servicing

The planners of next-generation satellite systems have given increased focus to the possibility of autonomous operation and in-orbit servicing. In the case of "autonomous operation," there are dual objectives of reducing operating costs while at the same time increasing reliability. The truth of the matter is that a number of satellite failures have been due to operator errors when a number of operations are required in a condensed period of time.

The thought has thus been to automate as much of the TTC&M operation as possible and to have onboard computers (or their backup processors) to be able to manage the operation of the spacecraft power, thermal, and pointing functions as possible. These systems would work to keep onboard systems always within a specified range of parameters and to operate with prearranged recovery procedures based on years of operator experience.

Although there can be emergency operation and ground-based overrides, it is thought that computer-driven satellite operations supported by millions of lines of code can allow a very great reduction in the need for large ground crews working 24/7 shifts while at the same time actually decreasing operator error. This would seem to be particularly true with regard to the operating of very large-scale LEO satellite constellations where a large number of satellites are under active management at all times. The same objectives for autonomous operation also apply to MEO and GEO systems.

Civilian Versus Military Satellite Design Strategies

Today, there are a large number of military communications, remote sensing, and space navigation systems that in many ways resemble commercial and civilian application satellites, but they are different in their design, engineering, and manufacture in many ways. The military systems tend to be significantly more expensive. This is because of efforts to make the military systems more robust in the case of attack and where possible to make them more reliable. Some efforts such as radiation hardening of military systems represent clear-cut differences. These satellites also operate in different frequency bands and are often required to interface with telecommunications or other facilities on the ground, on the seas, and in the skies. Some satellites are equipped with Faraday cages to protect against electromagnetic pulse (EMP) and associated nuclear explosive effects. The commercial alternative is to provide for more spares (i.e., both in-orbit and on the ground) as well as to rely on insurance coverage as a way to protect against various types of risk factors. Consequently the commercial systems without the protection and special redundancy are generally less costly.

To date there have not been specific attacks on commercial application satellites consistent with the nonmilitary uses of space as recorded in the Outer Space Treaty of 1967 and the five resolutions of the United Nations Committee on the Peaceful Uses of Outer Space (COPUOS) (Jakhu et al. 2009). Nevertheless, there continue to be concerns about the various tests of antisatellite weaponry that have been carried out by the United States, Russia, and China to verify their ability to destroy in-orbit assets. Recent tests by China and the United States have increased concerns not only about the military use of antisatellite weapons but also concerns that the use of such techniques can greatly expand the problem of orbital debris and the sustainability of space. These issues are being addressed in the context of United Nations discussions regarding the demilitarization of space in Geneva, Switzerland.

Conclusion

Reliability is absolutely key in the operation of application satellites because they are providing a vital service. Telecommunications satellites and space navigation satellites do not tolerate even brief outages due to the types of real-time services that they provide. Remote sensing, Earth observation, and meteorological satellites can tolerate somewhat more extended outages, but even these satellites need to be as absolutely reliable as possible since there are today no available means to do carrying out on-orbit repairs and refueling or installing of new batteries or solar cell systems. This situation is changing and on-orbit repair, refueling, and retrofit are quickly becoming an option to support not only extended reliable satellite operation but system upgrade.

The harsh environment of space includes hazards like radiation, sharp thermal gradients, a nearly complete vacuum, and shifting orbital mechanics. These conditions plus the remoteness of the satellite from Earth demand that all application satellites be designed to the highest levels of reliability, tested under extreme conditions, and provided with critical backup components where and when possible. It also requires a tracking, telemetry, command and monitoring system that allows operators to know exactly where the satellite is at all times, how it is oriented, how the satellite components are functioning, and to be able to command the satellite subsystems. Remote commands can activate backup subsystems, fire jets to reorient the satellite, recharge batteries, reorient solar arrays, and carry out a wide range of activities to recondition the satellite or rescue it from a variety of hazards. Satellites can be powered down during solar flares or coronal mass ejections. Backup transponders or ASIC components can be activated or a fault switch located. Once an application satellite is deployed and tested out, it has a remarkably good chance of operating for its full lifetime. A part of its capabilities such as a solar array, a battery, or a transponder or one of its sensors might be reduced in performance, but the odds are, based on 50 years of experience, that the satellite will have over a 90 % chance of achieving its projected lifetime. Many satellites continue to function well past their mean time to failure dates.

For the future, autonomous operation, in-orbit servicing, and improved component and subsystem design will likely assist in extending lifetime, improving reliability, and otherwise enhancing the cost-effectiveness of service.

The most significant new development in the satellite applications, in terms of reliability, is the new approach to deploy low-cost small satellites with off-the-shelf components in larger-scale constellations with greater reliance on significant sparing. In these small satellite constellations, the emphasis is not on individual satellite reliability and instead on overall system reliability and performance.

Cross-References

- ▶ Common Elements Versus Unique Requirements in Various Types of Satellite Application Systems
- ▶ Overview of the Spacecraft Bus
- ▶ Telemetry, Tracking, and Command (TT&C)

References

R. Jakhu, J. Logsdon, J. Pelton, Space policy, law and security, in *The Farthest Shore: A 21st Century Guide To Space*, ed. by J.N. Pelton, A.P. Bukley (Apogee Press, Burlington, 2009), pp. 306–314

United Nations Office for Outer Space Affairs: Report on current and planned global and regional navigation satellite systems and satellite–based augmentation systems, http://www.oosa.unvienna.org/oosa/SAP/gnss/icg.Html. Last accessed 14 Jan 2016

M. Williamson, *Spacecraft Technology: The Early Years* (IEEE, London, 2006), p. 312

Ground Systems for Satellite Application Systems for Navigation, Remote Sensing, and Meteorology

Scott Madry, Joseph N. Pelton, and Sergio Camacho-Lara

Contents

Abstract

The technology, the applications, and the economic forces that have driven the design, functionality, and performance of ground systems for satellite communications have been very closely mirrored in the other major application satellite services. It is for this reason that this chapter combines consideration of the ground systems for satellite navigation, remote sensing, and meteorology. In essence, all the ground systems for the various applications are communication systems. Although the radio frequencies, modulation, and multiplexing methods and encryption schemes utilized vary for a variety of reasons – including defense

---

S. Madry (✉)
Global Space Institute, Chapel Hill, NC, USA
e-mail: Scottmadry@mindspring.com

J.N. Pelton
International Space University, Arlington, VA, USA
e-mail: peltonjoe@gmail.com

S. Camacho-Lara
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Santa María Tonantzintla, Puebla, México
e-mail: sergio.camacho@inaoep.mx

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_11

and military-related consideration – all application satellites employ satellite communications between the spacecraft and the ground system. Some systems are broader or narrower in bandwidth and some only involve downlinks, while others are more interactive with up- and downlinks.

The common elements that range across the ground systems for all application satellites include the following:

- All application satellites have become higher in power, more accurate in their stabilization and pointing of their onboard antennas, and better able to deploy higher gain and larger aperture reflectors. This has allowed ground systems to be smaller, more compact, lower in power, lower in cost, and more widely distributed.
- Downlinked information is often encrypted to protect the integrity of information and data relayed from the satellite – particularly if there is a proprietary or defense-related application for the downlinked information.
- Solid-state digital technology associated with integrated circuitry, application-specific integrated circuits (ASICs), and monolithic devices that have allowed the ground systems to be more highly distributed.
- There are essentially two tracks in ground system development – one where geosynchronous satellites are involved and the ground system can be constantly pointed toward a single fixed point in the sky and the other where the ground system must have the ability to receive signals across the horizon and capture signals from a satellite that moves across the sky. Both types of ground receivers suited to “fixed” or “non-fixed” signal reception are needed in satellite communications, remote sensing, meteorological satellites, and satellite navigation.
- In addition to the user terms associated with different types of applications, there is a need for a tracking, telemetry, and command system to ensure the safe operation of the application satellite.

Despite these elements of commonality, there are indeed differences in the ground systems, the antenna characteristics, their tracking capabilities, the frequency utilized, the degree to which the data is protected by encryption, and the need for expert analysis of the data received from the spacecraft.

Keywords

Application-specific integrated circuit • Autonomous control • BeiDou space navigation system • Compass space navigation system • Encryption • Galileo space navigation system • GLONASS space navigation system • GPS receiver • Indian regional space navigation system • Monitoring • Monolithic devices • Navstar • Omni-antennas • Quasi-Zenith space navigation system • Spacecraft performance monitoring • Squinted-beam antenna • Telemetry and command • Tracking

Introduction

The first type of commercial satellite applications was for satellite communications. The types of ground systems evolved as the satellite communication industry evolved to cover fixed satellite services, broadcast satellite services, mobile satellite services, and store and forward data relay services. The use of satellites for remote sensing and meteorological services followed fairly closely in time, but these operations remained largely governmentally operated because there was not an established commercial market for these services. Only in the last few years have commercial remote sensing satellites evolved, but they remain heavily dependent on governmental and defense-related clients. The last of the satellite applications to evolve is that of satellite navigation.

In recent years, the number of types of application satellites has continued to multiply rapidly. In all types of satellite applications, the evolution of ground systems for users has been remarkably parallel. Satellite systems have become larger and more powerful and thus space sensors and payloads have become more capable. The bottom line is that ground systems for satellite application users have become smaller, lower in cost, and more accessible and allowed the range of applications and services provided by these ground systems to diversify and thrive (Pelton 2012). The need for expert analysis of remote sensing data and meteorological satellite data has lessened the trend in these areas, but an ever-increasing number of new digital applications is fueling this trend even in these areas. Despite this overall trend toward moving satellite applications to the "edge" with smaller and lower-cost user terminals in play, each type of application still has its own types of user devices. The one thing that all application satellites truly have in common on the ground is the tracking, telemetry, and command (TT&C) systems.

TT&C Ground Systems

The most vital part of any application satellite network is its tracking, telemetry, and command (TT&C) system. This is simply because without a functioning TT&C system, the satellite is lost in space and essentially not able to function. One can think of the TT&C system as both the brains and guidance control capability of the satellite. As noted in chapter "▶ Telemetry, Tracking, and Command (TT&C)," "the three major tasks that the TT&C subsystem performs to ensure the successful operation of an application satellite are: (1) the monitoring of the health and status of the satellite through the collection, processing, and transmission of data from the various spacecraft subsystems, (2) the determination of the satellite's exact location through the reception, processing, and transmitting of ranging signals, and (3) the proper control of the satellite through the reception, processing, and implementation of commands transmitted from the ground." These functions remain constant regardless of what type of application the satellite might be involved.

There is another function, however, that is very much dependent on the type of application. This is what is called the "spacecraft performance monitoring function," and this capability allows ground controllers to operate the satellite correctly, but the electronic monitoring can also allow operators to determine if the satellite is actually delivering the service for which the satellite was intended in a proper fashion or if there is radio-frequency interference that is impairing the service. Thus, for communication satellite, monitoring would determine if the quality as measured in bit error rate was appropriate and if there was undue interference occurring. In the case of remote sensing and meteorological satellites, the performance of the onboard sensors would be measured and calibrated. In the case of satellite navigation, the transmit signal would be measured and calibrated and interference detected.

The TT&C design for various types of application satellites is governed by only a few factors. These are:

- *Global, Regional, or Domestic Network*: If there is a global network, then TT&C facilities need to be available around the world. If the network is polar orbit and sun synchronous, then the TT&C facilities need to be located differently than if the network is geosynchronous. When application satellites were first deployed, each system operator had to arrange for or build their own TT&C network of expensive ground stations at appropriate locations around the globe. In time, however, as more and more systems were deployed, arrangements were made so that TT&C facilities could carry out these operations for multiple systems under an appropriate fee basis.
- *Assigned Frequencies for TT&C Operations*: A variety of frequencies were assigned to carry out TT&C operations. Most of these were in the very high-frequency (VHF) and ultrahigh-frequency (UHF) bands since these did not require broadband or high data rate to carry out tracking, ranging, telemetry, or commands functions.
- *Encryption*: Encrypted signals in special frequency bands are used for military and defense-related satellites, but as application satellites have gotten larger, more powerful, and broadband, application satellites of all types have moved to more and more secure modes of operation. Today, TT&C operations for all types of satellites are typically carried out with encrypted signals, and commands may not only be digitally encrypted, but there may be special security techniques employed such as requiring all commands to be confirmed from a separate TT&C site.

The high cost of TT&C operations is a matter of concern for all satellite application systems. This issue of cost has been addressed by automating many aspects of the TT&C operation with alarms sounding when a certain set parameter has been exceeded. The further step is to move toward what is called “autonomous operations.” This means that an artificially intelligent (AI) onboard computer assumes control of the satellite and responds to routine technical issues based on expert systems designed to adjust satellite settings or employ backup components and subsystems. With autonomous control, a satellite’s operation can become largely “autonomous,” and ground control is reassumed only in the case of major difficulties.

Ground Systems for Satellite Navigation

In a rapidly expanding field of applications of the signals provided by the GNSS, it is essential that the user community and receiver-producing industry have a clear and consistent description of the global and regional systems that are currently operating and that which will operate in the future. To this end, the United Nations Office for Outer Space Affairs, in its role as Executive Secretariat for the International Committee on GNSS (ICG), prepared a publication on the planned or existing systems and on the policies and procedures that govern the service they provide (United Nations Office for Outer Space Affairs 2010). To reflect future changes, the publication will be updated and will be available on the website of the ICG. The following paragraphs describe the ground segments for the existing and planned GNSS and their augmentations.

Because of the large number of satellites that comprise the global and regional navigation satellite systems and their augmentations (chapter "▶ Current and Future GNSS and Their Augmentation Systems"), their ground segments involve a larger number of antennas and monitoring stations. Figure 1 shows a typical GNSS antenna.

Fig. 1 Unified state ground control network GLONASS receiving station (Courtesy of Russian Space Agency Roscosmos)

The global positioning system (GPS) of the US operational control segment consists of four major subsystems: a master control station, an alternate master control station, a network of four ground antennas, and a network of globally distributed monitor stations. The master control station is located at Schriever Air Force Base, in Colorado, USA, and is the central control node for the GPS constellation. The master control station is responsible for all aspects of constellation command and control, including routine satellite bus and payload status monitoring, satellite maintenance and anomaly resolution, management of signal-in-space performance, navigation message data upload operations, and detecting and responding to GPS signal-in-space failures.

In September 2007, the GPS operational control segment was modernized by turning to a distributed system resulting in increased capacity for monitoring GPS signals, from 96.4 % to 100 % worldwide coverage with double coverage over 99.8 % of the world. All the current GPS interface control documents can be downloaded from GPS.gov, the official US government webpage for GPS.

The ground segment of the Wide-Area Augmentation System (WAAS) is operated by the Federal Aviation Administration (FAA) of the USA There are 38 wide-area reference stations throughout North America (in Canada, Mexico, the USA, and Puerto Rico). The FAA plans to upgrade the wide-area reference stations with receivers capable of processing the new GPS L5 signal.

The Local-Area Augmentation System is a ground-based augmentation system that was developed to provide precision-approach capability for categories I, II, and III approach procedures. It is designed to provide multiple runway coverage at an airport for three-dimensional required navigation performance procedures and navigation for parallel runways with little space between them and “super-density” operations.

The Nationwide Differential Global Positioning System is a national positioning, navigation, and timing utility operated and managed by the US Coast Guard. It consists of 50 maritime sites, 29 inland sites, and 9 waterway sites. The system provides terrestrial services to 92 % of the continental USA with 65 % receiving dual coverage. The system is used in surface and maritime transportation, agriculture, environmental and natural resource management, weather forecasting, and precise positioning applications.

The national network of continuously operating reference stations, coordinated by the National Geodetic Survey and tied to the National Spatial Reference System, consists of more than 1,300 sites operated by over 200 public and private entities, including academic institutions. Each site provides GPS carrier phase and code range measurements in support of three-dimensional centimeter-level positioning activities throughout the USA and its territories.

The Global Navigation Satellite System (GLONASS) of the Russian Federation ground segment consists of a system control center; a network of five TT&C centers; the central clock; three upload stations; two satellite laser ranging stations; and a network of four monitoring and measuring stations, distributed over the territory of the Russian Federation. Six additional monitoring and measurement stations are to start operating on the territory of the Russian Federation and the Commonwealth of Independent States in the near future.

The Unified State Ground Control Network[1] (USGCN) is designed to control automated spacecrafts, manned spacecrafts, and space stations. USGCN solves movement control problems for spacecrafts of different purposes during all flight and descent phases, control of operation of all their equipment and systems; scientific, meteorological, communication, television, navigation, and topogeodesic data reception; manned spacecraft crew radio communication; and carrier vehicle launch measurements. USGCN is a combination of hardware components and facilities located at the head test center of spacecraft testing and control (HTCTC SF) of the Ministry of Defense (MoD) and at ground-based detached command and measurement complex (DCMC). These facilities are connected by data and control communication channels into the unified automated control complex. Facilities inside the USGCN designed to control single spacecrafts or constellations of similar spacecrafts form ground control complex (GCC), which together with onboard control complexes comprise an automated control network. The document that defines requirements related to the interface between the space segment and the navigation user segment is the interface control document (version 5.1, 2008).[2]

The Galileo ground segment controls the Galileo satellite constellation of the European Union, monitoring the health status of the satellites, providing core functions of the navigation mission (satellite orbit determination, clock synchronization), determining the navigation messages and providing integrity information (warning alerts within time-to-alarm requirements) at the global level, and uploading those navigation data for subsequent broadcast to users. The key elements of those data, clock synchronization and orbit ephemeris, will be calculated from measurements made by a worldwide network of reference sensor stations. The current design of the system includes 30–40 sensor stations, five tracking and command centers, and nine mission uplink stations. The present Galileo Open Service Signal-in-Space Interface Control Document (OS SIS ICD) Issue 1[3] contains the publicly available information on the Galileo Signal In Space. It is intended for use by the Galileo user community, and it specifies the interface between the Galileo Space Segment and the Galileo User Segment. As the Galileo constellation is placed in orbit, the interface document is subject to evolution, and the information contained in it may change.

The EGNOS ground segment is mainly composed of a network of ranging integrity monitoring stations, four mission control centers, six navigation land Earth stations, and the EGNOS wide-area network, which provides the communication network for all the components of the ground segment. Two additional facilities, the performance assessment and system checkout facility and the

---

[1] http://www.spacecorp.ru.

[2] http://www.spacecorp.ru/en/directions/glonass/control_document/index.php?sphrase_id=3633.

[3] http://ec.europa.eu/enterprise/policies/satnav/galileo/files/galileo-os-sis-icd-issue1-revision1_en.pdf.

application-specific qualification facility, are also deployed as part of the ground segment to support system operations and service provision.

The ground segment of the Compass/BeiDou Navigation Satellite System of China consists of one master control station, upload stations, and monitor stations. Compass/BeiDou user terminals are intended to be “compatible” with GPS, GLONASS, and Galileo receivers.

The ground segment of Japan’s Quasi-Zenith Satellite System (QZSS) performs the tracking, computation, updating, and monitoring functions needed to control all of the satellites in the system on a daily basis. It consists of a master control station in Japan, where all data processing is performed, and some widely deployed monitor stations in the area that are visible from the space segment. The monitoring stations passively track all satellites in view and measure ranging and Doppler data. These data are processed at the master control station so that the satellite’s ephemerides, clock offsets, clock drifts, and propagation delay can be calculated and are then used to generate upload messages. This updated information is transmitted to the satellites via TT&C and to the navigation message uplink station at Okinawa for subsequent transmission by the satellites as part of the navigation messages to the users. The interface specification (IS-QZSS) document[4] defines the interface between the space segment (SS) provided by the Quasi-Zenith satellites and the user segment of the QZSS.

The ground segment of the augmentation system MSAS consists of two master control stations (one at Kobe and one at Hitachiota), two monitoring and ranging stations (one in Australia and one in Hawaii), and four ground-monitoring stations (at Sapporo, Tokyo, Fukuoka, and Naha). The master control stations generate augmentation information based on the GPS and MTSAT signals received at the ground-monitoring stations and the monitoring and ranging stations. The ground-monitoring stations monitor GPS satellite signals and transfer the information to the monitoring and ranging stations.

The ground segment of the Indian Regional Navigation Satellite System (IRNSS) is responsible for the maintenance and operation of the constellation. This segment comprises nine IRNSS TT&C stations, two spacecraft control centers, two IRNSS navigation centers, 17 IRNSS range and integrity monitoring stations, two IRNSS timing centers, six CDMA ranging stations, and two data communication links. As part of the ground segment, 15 Indian reference stations for monitoring and collecting the data, two master control centers, and three uplink stations are planned for the GPS-Aided GEO-Augmented Navigation (GAGAN) system.

The range and diversity of satellite navigation units that are available today are quite large. Among the units dedicated to use with the US Navstar system, GPS alone is staggering huge. There are units optimized for truckers, for boaters, for hikers, and even for golfers. There is a unit that is loaded with the layout of thousands of golf courses. These are low in cost, and these consumer-oriented units can cost in the range of US$100 (in used condition) to up to US$1,000. The

---

[4] http://qz-vision.jaxa.jp/USE/is-qzss/DOCS/IS-QZSS_14D_E.pdf.

Fig. 2 Garmin© Montana GPS receiver for hikers (Graphic courtesy of Garmin©)

price variation depends on the software for visual display, storage capability, touch screen capability, etc. (see Fig. 2).

There are then much higher-end GPS units that are designed for aviation applications and can provide 3-axis orientation and real-time display for pilots that shows the location with respect to the Earth (see Fig. 3). There are also units that are designed to operate in the frequencies for the GPS network as well as the GLONASS system. These are larger and more expensive (see Fig. 4). Finally, there are units that can access not only GPS and GLONASS but are designed to flexibly access other satellite navigation systems as they are deployed such as the European Galileo, the Chinese BeiDou/Compass system, the Japanese Quasi-Zenith system, and the Indian Regional Navigation Satellite System (see Fig. 5). These multiuse units that can utilize signals from all the current and planned satellite system cost in the range of about US$10,000–US$12,000. As the operators of the GNSS move toward interoperability (chapter "▶ International Committee on GNSS"), the receiver systems should become less expensive.

The above are only some of the ground receivers for space navigation systems now broadly available. There are even more sophisticated GPS receivers that are designed for space experiments and for activities or experiments where greater 3-axis spatial accuracy is required. In such cases, for instance, a group of four

Fig. 3 The Helm X650 "True Map" GPS that provides a 3-axis display for aircraft applications (Graphic courtesy of Helm©)

Fig. 4 The LaiPac Tech dual-mode receiver for both GLONASS and GPS signals (Graphic courtesy of LaiPac Tech)

GPS receivers might be configured in close proximity to establish greater range accuracy. Such a configuration might be used for aeronautical or space applications (Lachapelle et al.). At the other end of the spectrum, there are quite low-cost Argos omni-receivers available that are used in such applications where there is not a requirement for quite high spatial or geographic accuracy. These applications using Argos receivers might include activities such as locating ocean buoys associated

Fig. 5 Tokay professional satellite navigation receiver is equipped for all satnav frequency bands (Graphic courtesy of Tokay)

with ocean-based experiments.[5] The number and types of space navigation receivers will doubtlessly increase as more satellite navigation systems such as the Galileo, the Compass, the Quasi-Zenith, and the Indian Regional Navigation Satellite Systems continue to be deployed in the coming years.

The ground units for space navigation must be able to receive signals from orbiting satellites that are typically in quite high-medium Earth orbit or in the case of the Quasi-Zenith network (a geo-orbit that is tilted 45° to the equator). All of the ground receivers now in operation are essentially very low-gain omni- or squinted-beam omni-devices. These terminals, since they do not have an active tracking capability, are designed to capture transmitted signals from medium Earth orbit satellites anywhere above the horizon. The key to these devices, from a technical performance perspective, is a highly specialized application-specific integrated circuitry (ASIC) that allows digital processing algorithms to augment the ground unit's ability to receive a very low-level signal from the navigation satellite that is orbiting many thousands of kilometers above.

Ground Systems for Meteorological and Remote Sensing Satellites

The ground systems for meteorological and remote sensing satellites are, for the most part, quite different from satellite navigation satellites. This is because most of these ground systems are designed to receive signals from meteorological or remote sensing satellites at special facilities designed for data analysis by trained specialists. Essentially, meteorological satellites are just a specialized form of remote sensing satellite, and only geosynchronous meteorological satellites are a separate case. In some cases data from a remote sensing satellite can be relayed via a data relay satellite to provide the data for analysis on an accelerated basis. This could be in the

---

[5] Argos Receiving Antenna. http://www.telonics.com/products/argosReceivers/.

case of a military or defense-related application or in the case of a hurricane, monsoon, or typhoon in order to mitigate a disaster or provide accurate warning notices.

In essence, virtually all ground stations supporting remote sensing or meteorological activities are not intended for mass consumer use although there are lower-cost “hobbyist ground systems” that can be acquired at rather reasonable cost that will be discussed below. The number of professional or ground receivers located at universities for training purposes is much fewer than is the case with space navigation receivers. In short, while there are tens of millions of GPS and GLONASS – and soon other types of space navigation terminals – there are only thousands of professional high-gain meteorological or remote sensing receiving ground stations with rapid tracking capabilities. For polar-orbiting meteorological and remote sensing satellite operations, these will typically have rapid tracking capabilities since the relatively low orbits involve passes over the ground station and data processing center in a matter of just a few minutes.

All types of these ground antennas for meteorological and remote sensing operations (including geo- and polar-orbiting systems) will have increased gain and thus relatively antenna apertures. As remote sensing and meteorological satellites have been equipped with more and more types of sensors – spectral, hyper-spectral, infrared, radar, monitors of lightning strikes, etc. – the ground systems have been upgraded to receive more data more efficiently via enhanced digital transmission capabilities. During a single satellite pass, many gigabytes of data might be gathered at a single processing center. The initial ground systems have huge antenna systems since the transmission capabilities of early satellites were limited. The early TIROS ground systems represent some of the largest ground systems constructed during the 1960s (see Fig. 6).

The design of some remote sensing systems such as the so-called “Mission to Planet Earth” system was scaled down in terms of the amount of data collected via various sensors not because of the ability of ground stations to collect the data, but due to the ability to accurately process the many terabytes of data that could be collected by a network of scores of remote sensing and meteorological satellites. Today, the ground systems for collecting data from remote sensing systems such as Spot Image, GeoEye, and other remote sensing satellite networks are still quite capable of rapid tracking across the sky but nevertheless much smaller than the giant TIROS stations shown below (see Fig. 7).

The first step in processing data at a ground receiving complex is “unpack” the incoming data and store it for processing. The data processing of remote sensing data is often divided into “preliminary processing” that is then followed by “thematic processing.” The preliminary processing involves “unpacking.” This involves converting the raw data, received by the ground station, into products suitable for storage and further thematic processing. This “preliminary processing” can include a number of steps. These can include radiometrical calibration, geolocation, and geometric correction of images (Gershenzon and Kucheiko).

Later thematic processing involved the detailed interpretation that is carried out to accomplish specific tasks associated with agriculture, forestry, fishing, mining, urban

Fig. 6 The TIROS ground stations from the 1960s (Photo by Frank Vosk)

Fig. 7 A spot image remote sensing receiving station (Spot Image Receiving Station, http://www.astrium-geo.com/files/pmedia/public/r2184_9_spot_receiving_station_antenna.jpg) (Graphic courtesy of Astrium)

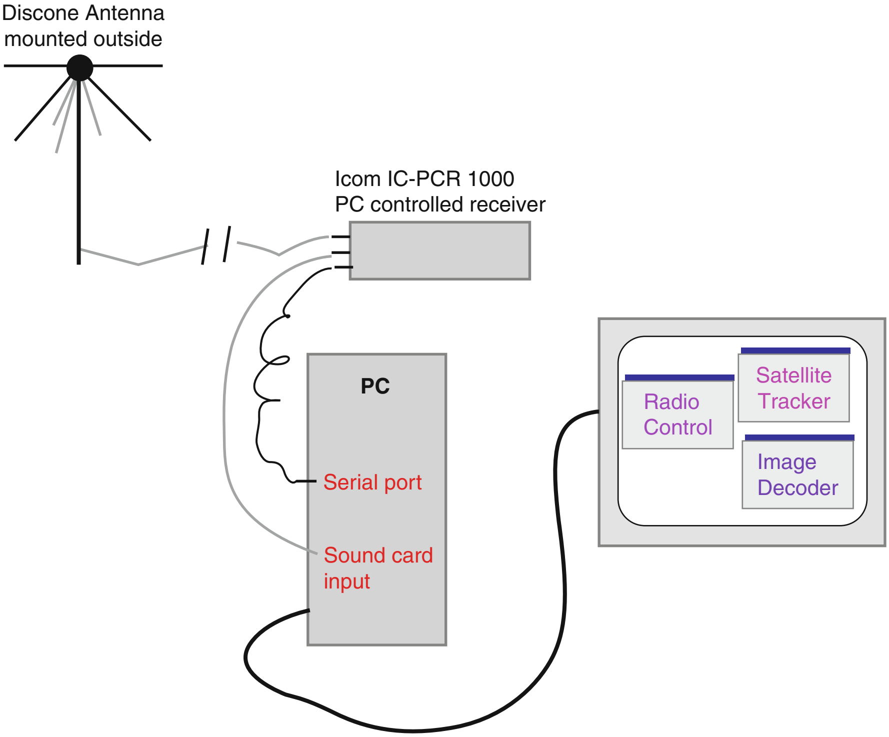


Fig. 8 Diagram for a hobbyist version of a simple ground station to receive remote sensing or meteorological data (Graphic Courtesy of HobbySpace)

planning, or even disaster recovery. There is today a wide range of ground stations capable of reception and preliminary processing of data, and these vary in their characteristics if they are receiving from geo- or polar-orbiting satellites in terms of their tracking capability. Also, there are some differences as to the receiving data rates in terms of the governmental, military, or commercial applications and especially with regard to the resolution of the sensing. Obviously, the higher the resolution, the more data that is captured to be transmitted to the ground.

There are much simpler, non-tracking receiver antennas that can capture far less data, but still could be of interest to the hobbyist who is interested in space-based meteorology or remote sensing. The following diagram outlines a schematic for such a low-resolution ground receiver that could be purchased or constructed from components by a hobbyist for a few hundred dollars (US)[6] (see Fig. 8).

---

[6] Building a Weather Satellite Station. http://www.hobbyspace.com/Radio/WeatherSatStation/.

Future Trends

The future trends for ground systems for all application satellites will tend to follow the pattern seen over the past 50 years. This trend will be for smaller, less expensive, and more widely available and user-friendly ground systems. The fields of satellite communications and satellite navigation have seen this development quite successfully accomplished since software and ASIC chips have allowed the consumer to utilize the transmission directly from the satellite with increasing ease.

Recently there have been a number of new type of satellite constellations based on the use of small satellites with a larger number of spacecraft in the constellation. This development is driven by the availability of new commercial launch vehicles that can insert small satellites into low Earth orbit at lower cost and 3-D printing and advanced manufacturing techniques that can create highly capable small satellites at lower cost. This requires greater capability to capture signals from rapidly orbiting satellites. Part of this evolution has been the development of “smart” ground systems that have the ability to track “electronically” satellites as they pass over. This new development of electronically tracking ground systems is still new and rapidly evolving, but these ground receivers will increasingly depend on programmable application-specific integrated circuits that will allow electronic rather than physical tracking of satellites as they pass overhead.

As the volume of receiver terminals in these areas has risen to many millions, the cost of these devices has continued to drop so that consumer TVROs for direct broadcast television and GPS receivers are now in the hundreds of dollars (US) range. The challenge for these small ground units with electronic tracking capability is to be manufactured and sold at reasonable small cost. This will be one of the major challenges for the 2017–2020 time frame.

Today, meteorological and remote sensing satellites still require a significant amount of professional formatting and analysis (preliminary processing and thematic processing) that creates the need for larger receiving stations with relatively complex operation and maintenance requirements. Pressures to develop software so that remote sensing and meteorological data can be brought closer to the “edge” – particularly to support warfighters in the field – have continued the trend toward decentralization and the use of expert systems in mobile computers and even personal data assistants for weather and remote sensing data assessment. The complexity of thematic analysis will prevent the complete decentralization seen in satellite navigation and satellite communications, but the future trends toward smaller, simpler, and lower-cost receivers seem to be universal throughout the industry.

Conclusion

The ground systems for all types of application satellite systems will be increasingly smaller, simpler, and highly distributed with communication satellite and space navigation units continuing to lead the way. There is speculation that such devices might someday evolve into wearable antennas that might take the shape of wristwatches or be embedded in a shirt or jacket. The evolution of meteorological and remote sensing ground systems will take a slower and more diverse route. This could be a full spectrum of ground systems with complex antennas continuing to be linked to analysis centers for more complex operations related to map making, resource prospecting, and high-value remote sensing applications, but other simpler applications being much more distributed through the use of artificially intelligent or expert system software allow data going directly from satellites to handheld or lower-cost mobile units.

All of the applications will continue to need tracking, telemetry, and command plus performance monitoring to operate the application satellites in the sky. Indeed, to allow the ground units to shrink in size and cost, the satellites will need to be higher in power and increasingly capable. This means that the satellites and the TT&C facilities will need to be more capable to compensate for the simpler units on the ground or onboard vehicles (including aircraft).

Cross-References

- ▶ Current and Future GNSS and Their Augmentation Systems
- ▶ International Committee on GNSS
- ▶ Introduction to Satellite Navigation Systems
- ▶ Satellite Antenna Systems Design and Implementation Around the World
- ▶ Telemetry, Tracking, and Command (TT&C)

References

V.I. Gershenzon, A.A. Kucheiko, *Remote Sensing Data.* http://www.scanex.ru/en/publications/pdf/publication4.pdf

G. Lachapelle, H. Sun, M.E. Cannon, G. Lu, Precise aircraft-to-aircraft positioning using a multiple receiver configuration. http://webone.novatel.ca/assets/Documents/Papers/File19.pdf

J.N. Pelton, *Satellite Communications* (Springer, New York, 2012)

United Nations Office for Outer Space Affairs, *Current and Planned Global and Regional Navigation Satellite Systems and Satellite-Based Augmentation Systems*, ST/SPACE/50 (United Nations, New York, 2010). http://www.icgsecretariat.org

Common Elements versus Unique Requirements in Various Types of Satellite Application Systems

Joseph N. Pelton and Scott Madry

Contents

Abstract

The concept of developing a handbook on satellite applications is based on the concept that all of the commercial and practical applications of space have many elements in common. In fact very similar power systems, spacecraft platforms, stabilization and positioning systems, and tracking, telemetry, and command

---

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

S. Madry
Global Space Institute, Chapel Hill, NC, USA
e-mail: Scottmadry@mindspring.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_71

systems are used for the various types of application satellites. It is the payloads that tend to be quite specialized. Even in the field of telecommunication satellites, quite different antenna systems and communications subsystems are now developed and deployed for various satellite systems for satellite broadcasting, fixed, or mobile services. It is equally true that different types of remote sensing, space navigation, and meteorological satellites can and do have different payload designs. The purpose of this chapter is to contrast and compare different types of application satellites to note major areas of similarities as well as how and why differences occur. Such an analysis is useful to understand where the most promising common elements lie in order to aid identifying new potential synergies for future research and development in order to seek out improved methods for common forms of reliability testing, sparing and redundancy strategies, as well as lifetime extension and reduced operating and monitoring costs.

Keywords

Three-axis stabilization • 3D Printing AC/DC current • Antenna systems • Ariane launch vehicle • Carbon/epoxy composites • De-spun platform • Earth and sun sensors • Energy density • Heat pipe • Intellectual property • Launch services • Lifetime testing • Lithium ion batteries • Payload design • Power systems • Solar cell • Space agencies • Spacecraft bus • Spacecraft structures • Thermal control • Thrusters • Tracking, telemetry, and command (TT&C) • Vernier jets • World Trade Organization (WTO)

Introduction

There are clear parallels between the various types and classes of application satellites. The various sections of this book are organized to indicate first the various ways that telecommunication satellites, remote sensing satellites, space navigation, and meteorological satellites are different in terms of payloads and antenna systems as well as market trends and structure. This is followed by noting the various elements that tend to be common in terms of spacecraft structures, power systems, orientation and positioning systems, and launch arrangements.

This chapter has two main objectives. The first of these objectives is to seek to provide an analysis of the various technical, operational, market, and business aspects of the overall satellite applications field in a manner to show those elements that are quite parallel and common and which would benefit from future technical research and development (R&D) and management strategies that allow mutual benefit across the entire field. Examples of this would be things like improved and longer-life batteries or fuel cells; improved inertial or momentum wheels; lower-cost and more precise atomic clocks; improved tracking, telemetry, and control systems; improved storage and buffering systems; better orientation and positioning systems; improved data relay systems; low-cost and more reliable launch systems, etc.

The second of these objectives is to clearly identify areas within the overall field of satellite applications where individual and unique requirements exist and separate approaches are needed to develop appropriate and improved new technology, operating capabilities, and/or business and management techniques. Examples of unique requirements that are particular to a specific application area would be the following: (i) extremely large aperture antennas for mobile communications; (ii) the improved design of synthetic aperture radar systems for remote sensing; or (iii) new algorithms and software for tracking of high altitude platforms via GNSS navigation satellites.

There is a temptation to assume that this sort of dichotomy with regard to common elements and unique elements can quickly be identified and sorted out from one another. Unfortunately this is not always the case. If one takes the example of onboard autonomous control, there may well be common elements that could reduce the cost of operating the various types of application satellites. It is also true that a technique developed within an artificial intelligence or expert system software might be very well applied to satellite telecommunications, for instance, but might also produce an unintended or harmful result in a space navigation system or remote sensing satellite.

The main point to emphasize is that even if there is a presumed benefit from a new capability developed for one type of application satellite, there must be very careful consideration as to how or why it might be applied to another type of system. This caution even applies to a particular subfield of satellite applications. A technique that is appropriate for broadcasting (or a one-way service where transmission delay is not a key issue) may not work well for interactive systems that provide mobile satellite communications, fixed satellite services, or defense-related applications. This is, in short, a much more difficult process than one might at first think. It is much harder than the children's game of "which of these things is not like the other" that is played on education television. Nevertheless, the following analysis seeks to create an analytic framework for identifying the most promising areas for future development where application satellites might have quite common goals for improved performance, reliability, or cost reduction.

Common Technical Elements in Application Satellite Programs

There are many aerospace manufacturers around the world and their expertise and research capabilities are central to the supply of quality application satellites – past, present, and future. A number of space agencies around the world add some research funds and expertise in the application satellite field, although support for application satellite research has waned as industrial and commercial capabilities in these areas have strengthened over the years. Today there is still active and meaningful support for research in the application satellite area. Notable R&D programs are pursued by the Brazilian Space Agency (INPE), the Canadian Space Agency (CSA), the Chinese National Space Agency (CNSA), the European Space Agency (ESA) (especially the TIA-ARTES research program), the French Space Agency (CNES), the Germany Space Agency (DLR), the Indian Space Research Organization (ISRO), the Italian Space Agency (ASI), the Japanese Space Agency (JAXA) (especially the Engineering Test Satellite (ETS) research satellite series), the Russian Space Agency (Roscomos), and even the US Space Agency (NASA) (which was one time active with regard to telecommunications satellite research but now only carry out programs regard to remote sensing and meteorological satellites). These various government-sponsored research programs – and even some defense agency research support – add quite useful supplemental research funding and technology development. Nevertheless, it is the major aerospace industries that are critical to the future development of application satellite technology. Major firms such as Ball Aerospace, Boeing, Lockheed Martin, Northrop Grumman, Orbital Sciences, Space Systems/Loral (in the USA); Alcatel, BAE Systems, EADS/Astrium, Paradigm, Siemens, and Thales Alenia (in Europe); NEC, Toshiba, and Mitsubishi Electronics Company (MELCO) (in Japan); the Chinese Aerospace Science and Technology Corporation and the Great Wall Industry Corporation (in China); Comdev, MacDonald Dettwiler, and MDA Space Missions (in Canada); and Hyundai (in Korea) represent some of the most important suppliers of the world's application satellites.

The above listing of aerospace manufacturers is, of course, only intended as partial listing of the entirety of global application satellite system suppliers. This is nevertheless indicative of the largest suppliers. There are important emerging suppliers in Brazil, Israel, India, and Russia, as well as entities like Surrey Space Technologies in the United Kingdom and several US suppliers that specialize in microsatellite design and manufacture. The appendices to this handbook provide a more comprehensive listing of launch vehicle suppliers and satellite manufacturers.

The basic point is these commercial spacecraft manufacturers actually design and manufacture the overwhelming number of telecommunications, remote sensing, space navigation, and meteorological satellites. Any transfer of technology-related technical, operational, or business systems for application satellites must take into account the pool of satellite system suppliers and intellectual property protections and patents that would be central to this process (Ippolito and Pelton 2004).

Spacecraft Structures and Bus Platforms

A very high percentage of application satellites today are built on "bus platforms" equipped with high-speed, magnetically suspended momentum wheels that allows the entire system to be stabilized in all three axes and very accurately pointed. The "boxed shaped" bus is capable of supporting a wide range of antenna structures and deployable solar power arrays that can be oriented to achieve maximum sun exposure. In the very early days of satellite design and deployment, satellites came in a wide range of shapes and sizes and many satellites from start to finish were either one of kind or a few of a kind. This rather chaotic design environment tended to drive up sharply the initial design and engineering costs, i.e., the nonrecurring cost component, for each new satellite or satellite series. In the age of the "de-spun satellite design," there was a move toward standardization of different classes of satellites, i.e., small, medium, and large configurations (Pelton 2006).

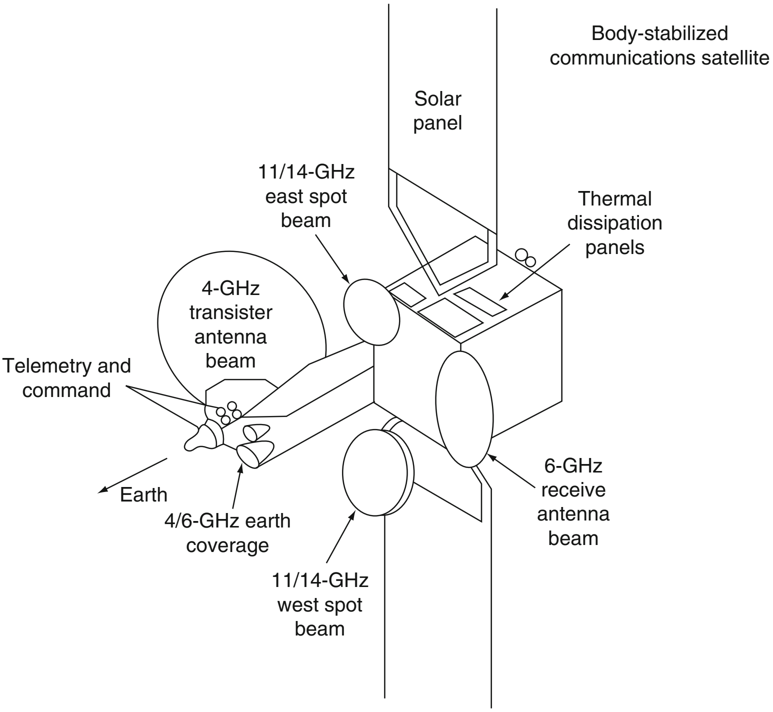


Fig. 1 The three-axis body stabilized application satellite with a box-like platform is now the industry standard for virtually all commercial spacecraft (Graphics Courtesy of J. N. Pelton)

The advent of the three-axis body stabilized spacecraft that used high-speed momentum or inertia wheel to achieve much more precise pointing of space systems in Earth orbit revolutionized the space applications industry. This approach led to a typically standardized “box design” for the basic spacecraft unit from which antennas, solar power arrays, and imaging systems were deployed. Once this concept of how to design a satellite that could precisely point antennas or imaging devices and allow solar arrays to be deployed with maximum efficiency, the wisdom of developing specific classes of “platforms rather naturally evolved to even more sophisticated levels” (Fig. 1).

Just as automobile manufacturers tend to have just a few standardized chases (or platforms) on which to design cars and trucks, the satellite industry found scale economy and more rapid build and test efficiency to have only a few platforms to support telecommunications, remote sensing, space navigation, and meteorological satellites. In fact these same platforms might indeed be used as the basis of exploratory, scientific, or even defense-related satellites as well.

Other common elements in spacecraft bus design evolved as well. Most spacecraft buses and “masts” for antennas or sensors were built from ultrastrong but quite lightweight carbon/expoxy composite and acrylic fiber structures. The electrical systems used within the spacecraft bus also tended to be similar “direct current

(DC)” based designs, although highly specialized systems in rare instances also use “alternating current.” In general, AC to DC conversion tends to be minimized where possible. Passive and active thermal control systems from reflective exterior materials to heat pipes could be used over again with confidence since common design allowed for use of design that had been actually tested in space.

Tracking, telemetry, and command (TT&C) systems were often attached to the bus in a similar fashion. Sensors for detecting the spacecraft orientation were likewise common as well as the thrusters and Vernier jets that maintained the proper position accuracy to the Earth below (Pelton 2006).

All of these similarities in spacecraft bus design for application satellites provide powerful economies of scale. The creation of just a few different sized bus platforms not only allows for economies in terms of design, engineering, and production, but it also contributes to reliability in terms of being able to eliminate design flaws, reduce manufacturing errors, and improve testing techniques. The use of common batteries, momentum wheels, and thruster jets allows improved lifetime testing practices.

Power Systems

Satellite power systems, because they are crucial to performance and vary greatly in terms of requirements from satellite to satellite – even from different telecommunication remote sensing, space navigation, or meteorological satellite – deserve particular consideration. Power failures are a very key issue for application satellites. This is because both in space and on the ground, this is one of the most frequently experienced problems. While active components fail, they can often be backed up by redundant units that allow rapid restoration of service. If a solar power array fails to deploy or if a battery system fails, the mission is essentially a loss. Some telecommunication satellites (especially direct broadcast or mobile satellite systems requires quite high power level in the range of 10 kW or above) as to active remote sensing systems such as radar satellites that must generate power to be reflected back to the satellite. Other micro satellites for communications or remote sensing as designed by the Surrey Space Technology Center can have relatively low power requirements. Thus power systems for application satellites constitute a particular area of focused research. There are several key factors of commonality.

- *Solar cell performance*: Launch service costs are quite high. It may well cost as much to launch a satellite and insure against its failure than to manufacture it in the first place. Since launch costs are high, one desires solar cells that are quite efficient in converting solar energy from photons into electrical energy. Thus low efficiency by low-cost amorphous silicon cells are typically not used on application satellites. There is often a choice between lower-cost structured silicon solar cells (around 15 % or efficiency) and higher-cost gallium arsenide solar cells (around 20–25 % efficiency). Clearly if one could develop much higher efficiency solar cells while keeping costs low, this would be a boon to all forms of application satellites. (Indeed, such developments could spin off to Earth-based solar energy generating systems.) This is why there is a good deal of research in new high valence materials for solar cells as well as multijuncture solar cells (i.e., the violet solar cell and/or the "rainbow" solar cell). So-called quantum dot technology that could be up to three times more efficient than solar cells is subject to intensive R&D. Others believe that reflective surfaces or solar concentrators that allow a solar cell to "see" the equivalent of three or indeed many suns might provide an important pathway forward.

- *Battery and fuel cell systems*: One of the common problem experienced by application satellites is that the sun's energy is at one time or another blocked by the sun. The Earth blocking the sun from illuminating solar arrays is the most common problem. In order to provide power during eclipses or other short-term outages that may occur, one must have an alternative power supply and energy storage supply. There has been steady improvement in battery performance in terms of energy density and reliability. Lithium ion batteries are now the most common satellite energy storage and power supply system. Progress continues to develop unitized and regenerative fuel cells that could be used on satellites or even Earth-based applications. Some shorter-lived satellites deployed in LEO orbits might have much different requirements from longer-lived satellites in GEO orbits. Satellites in polar orbit that have maybe 35–40 min of eclipse out of a 90 min orbit will have different battery requirements from a GEO satellite that only experience only seasonal eclipses and even then for a maximum of no more than a hour out of a 24 h day.
- Despite these various differences, all types of application satellites could benefit from improved battery or fuel cell development and/or increased lifetime and reliability. Some engineers believe that in time batteries, fuel cell, or even nuclear power sources may evolve to such an improved state in terms of performance, cost, and reliability that one might even deploy a satellite without solar power arrays and rely exclusively on one of these onboard power sources that does not require the risk of deployment of a solar array and does not entail the considerable mass of having both solar and power storage systems onboard the satellite. Technology to support such a design concept is still in the future.
- *Protecting satellite power systems from failure*: There are ongoing research and development (R&D) programs directed toward enhancing the reliability of satellite power systems. Satellites that fly through the van Allen belt and especially medium earth orbit satellites typically have their solar cells coated with a silicon veneer to protect the lifetime effectives of these devices. There are efforts to find solar cells that are more resistant to radiation as well as to find lightweight coating systems to protect the solar arrays. There are also new designs to deploy solar arrays more reliably. Instead of "accordion-like" extension systems some of the newer designs have solar arrays that can be rolled out to full extension (Sachdev).

For the reasons noted above, there is ongoing research and development to extend the performance, reliability of satellite power systems, as well as efforts to reduce the cost and mass of the various sources of power supply. The satellites that require the most power such as direct broadcast satellites, mobile satellite communications, and

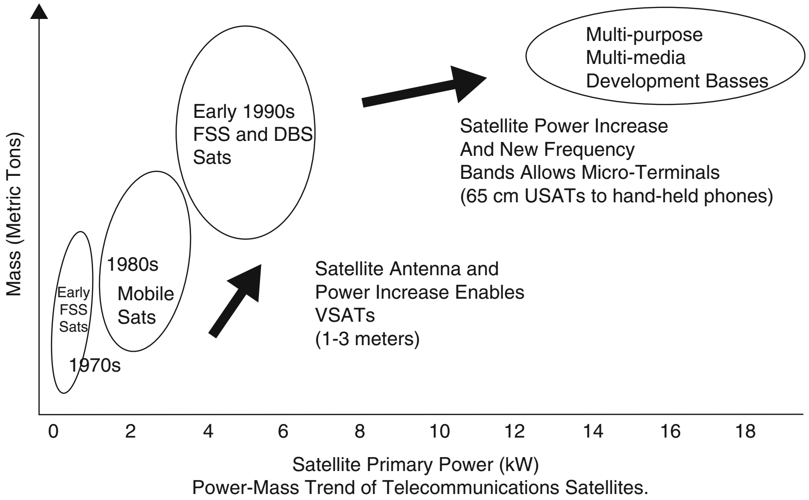


Fig. 2 The continuing trend toward more powerful application satellite (Graphic Courtesy of J. N. Pelton)

radar sensing systems because they represent the greatest challenges will likely lead the way forward. As can be seen in Fig. 2, the trend line has been and continues to be larger and more powerful satellites can beam information to smaller and more mobile user terminals.

Tracking, Telemetry, and Command

All forms of application satellites require a tracking, telemetry, and command (TT&C) system. The different types of applications satellite systems for various services have different requirements and challenges in terms of TT&C. The precise ranging and tracking methods used to detect the precise position of a GEO satellite that is one tenth of the way to the Moon are much more challenging than carrying out this function for a LEO satellite that is in an orbit that is 40 times closer to Earth. On the other hand, a LEO constellation with perhaps 50, 60, or more operational satellites in a global network plus spares makes tracking of such a large number of spacecraft much difficult to “see” from the ground. This is simply because the Earth blocks the ability to track the satellite from a particular location except for a brief period of time.

The problem of tracking a LEO constellation (as well as collecting telemetry data or sending it commands) breaks down into three different solutions. All of these are rather expensive. One solution is to have a large number of TT&C facilities on the ground to carry out these functions. This could require as many as 100 locations with up to three TT&C facilities ability to track in different directions at each location. (In practice, the actual number is quite a bit less because of the geography of the oceans, polar caps, etc. The Global star system, for instance, has an extensive terrestrially-based TT&C system to track its constellation that extends from 55° north latitude to 55° south latitude.) The second solution is to equip a LEO constellation with intersatellite links such as the iridium system that has each satellite with links to four other satellites in a highly symmetrical constellation. In this case, the links are to the two satellites that are either ahead or behind and another two are to satellites that are across. This allows the tracking, telemetry, and command functions to be carried out at only one site with another backup site. The third solution, which has been used by a number of the space agencies such as NASA, ESA, and JAXA, has been to deploy in GEO orbit a tracking and data relay satellite network. Three such satellites can collect information from the LEO or MEO satellites and then relay the information to a global command and control center. Again the satellites in LEO or MEO orbit would have to be equipped with TT&C antenna systems to accommodate such a space-based solution.

The bottom line is that TT&C systems are technically challenging and expensive to engineer and operate, but they are nevertheless essential to reliable functioning of any satellite application network. Despite the differences between orbits and mission requirements, here is a great deal of common technology with regard to TT&C that can be shared.

Ground and User Systems

The general trend in telecommunication satellites with regard to user antennas, handheld transceivers, and receive-only terminals has been to migrate from centralized systems utilizing very large and expensive antenna systems to more and more decentralized units that consumers can purchase and operate themselves. This same trend has now obviously continued with regard to space navigation systems. In the area of remote sensing and meteorological satellite systems, this same decentralization process is now also beginning to occur. In the case of remote sensing and meteorological systems, the practice for decades has been to relay the incoming data from satellites to processing centers that subsequently release data to various institutions and the public. The advent of expert system and artificial intelligence has led to new more immediately distributed applications.

For instance, research is ongoing that would allow airline pilots to receive information in the cockpit during flights in “near real time” about ambient weather conditions that is relayed from meteorological satellites. Remote sensing data and space navigation information can now be increasingly acquired directly by defense forces operating in the field via deployable mobile receiving stations. The overall trend seems to be to move voice, video, and data acquired directly from all types of application satellites and relayed to the end user at the “edge” of various satellite distribution or broadcast networks. This means that compact, transportable, and low-cost transceivers will be increasingly in demand. Critical to this development will be lightweight, high-energy density batteries that can be easily recharged as well as more and more capable application-specific integrated circuits and monolithic devices that allow more and more miniaturization. Improved antenna systems, both on application satellites as well as on user terminals will also be a critical development path forward. In summary, there are key parallel trends across the applications satellite industry to develop better user terminals that are lower in cost, more transportable, equipped with improved batteries and antenna systems that allow higher throughput rates and higher transmission speeds.

Launch Services

Launch services, in terms of cost, reliability, and flexible availability, are critical to the success of the applications satellite industry. For many years, it was the demand for telecommunications satellite launches that drove the commercial launch services industry. For the years from 2005 through 2008, Ariane-5 launch vehicles placed 40 telecommunications satellites in orbit and only two non-telecom payloads. Prior to 2005, of the 155 satellites successfully launched by Ariane-4 in the course of its operation, 139 were telecommunications satellites. Other providers of launch services in China, India, Japan, Russia, the Ukraine, and the United States likewise were heavily oriented toward the launch of communications satellites. Although the growing diversity of commercial markets for applications will continue to expand the demand for commercial launches, telecommunications satellites will likely predominantly drive the market for years to come.[1]

This is to say that the sizing of payloads to accommodate various classes of launch vehicles and the design of future direct broadcast and mobile satellite communications systems will drive the design of the largest spacecraft platforms and their sizing to meet the “fairing dimensions” with the largest launch vehicle available. At the other end of the spectrum, the platforms used on microsatellites will likely be designed for telecommunications satellites and then adapted for use by other satellite applications. Launches of smaller satellites could be multiple launches at the same time. For instance, multiple Orbcomm satellites were simultaneously launched by the Orbital Sciences Corporation’s Taurus launch vehicle and Globalstar unsuccessfully attempted to launch seven of its satellites on a Zenit launcher. In other cases such launches might be “piggyback” launches accomplished on larger vehicles with special configurations to launch one or two major payloads plus a several other microsatellites at the same time. Regardless of the configuration and the payload, the launch industry is most likely to design lift capabilities optimized by demand from the communications satellite industry because of the relative size of the market demand.

The history of launch vehicle development has been dominated for the last 50 years by chemically fueled launchers – using either liquid or solid fuels. This development has led to more and more reliable lift capabilities but only modest

---

[1] European Space Agency (ESA) Telecommunications: The Satellite Market http://www.telecom.esa.int/telecom/www/object/index.cfm?fobjectid

reduction in launch services costs. Although launchers have become more and more reliable, launch insurance with premiums in the range of 12–20 % of the cost of the insured risk have also remained relatively high.

Currently there is a great deal of research to develop new launch or orbital insertion technologies. Much of this focus is on higher and higher performance ion engines and electrical propulsion. These efforts are aimed at adding reliability and cost reduction, but today these systems only have limited thrust capabilities suitable for positioning, station-keeping, or orbital insertion after liftoff via a chemically fueled rocket. Other lift capabilities such as nuclear propulsion, use of tethers for orbital elevation, and higher performance electrical propulsion systems are under R&D investigation. Other unconventional approaches such as “rail-guns,” “space elevators,” lighter-than-air “dark sky” stations as liftoff sites for ion engine propulsion, etc., are also under study. Any new lift capability that would increase reliability, decrease cost, lessen the cost of launch insurance, and perhaps also reduce the polluting effects of chemically fueled launches – particularly solid fueled systems – would be a boon to the entire satellite applications industry.

Common Operational and Regulatory Aspects of Application Satellite Programs

All application satellites must have some degree of operational management and control to maintain the satellite’s health. There must be conditioning of batteries, monitoring of solar array performance and orientation, firing of jets to maintain proper orbits, monitoring of thermal control systems to see that temperatures are maintained within appropriate limits, etc. Most importantly, there is an ongoing review of payload performance to ensure that the desired information is flowing to and from the satellite without significant interference and that key operational components are performing correctly.

Most tracking, telemetry, command, and operational monitoring functions are automated with alarms sounding if a fixed parameter limit happens to be exceeded. In such cases, alarms sound and trouble-shooting efforts begin to identify corrective measures. This might be to reorient the space craft in the case of thermal difficulties or to command a switch to a back up sensing device or communications transponder. Since application satellites are constantly subject to in-orbit hazards such as solar flare, a component failure, or even a possible collision with space debris or another satellite, operators must have a 24/7 capability to respond to any difficulty that may arise. Operational control and management for geosynchronous satellites are generally easier since these satellites are well above the Van Allen Belts, have relatively stable orbits, have less risk of physical collision and only need to connect with only one TT&C facility on the ground. The most difficult satellite systems to manage and control are very large constellations with multitude intersatellite links and multiple TT&C facilities connecting with the space-based network. In the initial deployment of such systems, there can be so-called cockpit errors when satellites are being tested, initial parameters checked, and multiple anomalies being addressed at once.

The increasing complexity of satellites, particularly in terms of communications satellites that may require the instant interconnection of literally hundreds of spot beams, is driving the operation of such networks toward increasing automation of the operating process as well as sophisticated self-diagnostic capabilities to identify system faults. For instance, in the case of the Intelsat 4 satellite, where there are on the order of 240 global, zonal, and spot beams to interconnect in real time, the number of potential pathway interconnections is $n/2(n - 1)$ or $120 \times 239 = 16{,}780$ pathways. Without digital processing support a rapid determination of a pathway fault would be almost impossible.

The most ambitious operational objective is what is called autonomous operation. This is to design the software on the satellite with as much "intelligence" as possible. The idea is not only add onboard switching and signaling capability but also design the satellite so that it can operate as independently as possible with very little ground-based monitoring and control. Experience with experimental satellites designed by ESA, NASA, and JAXA have shown that truly "smart" and virtually totally automated satellites are still some years away.

Another serious operational concern that is common to all application satellites is the possibility of hackers or even terrorist organizations obtaining unauthorized access to an operational commercial satellite and commanding it to move out of the correct orientation or position, discontinue service, fire its jets to put the satellite into a descent trajectory that would remove it from orbit, or to otherwise send spurious commands that would disable the satellite. There are several instances of attempts by hackers to gain access to satellites and send spurious commands. Most operational satellite systems have sophisticated "firewalls" and command controls to prevent such occurrences. Most large-scale systems with global operations have sophisticated codes that must accompany commands. In some cases, there is an additional "failsafe" requirement that another TT&C station must send a confirmation code.

There is an ongoing effort within the satellite applications industries to use artificial intelligence and expert system software to continuously monitor the health and performance of each satellite. The number of lines of code associated with the TTC&M functions (i.e., tracking, telemetry, command, and monitoring) can now run to millions of lines of code for the most sophisticated satellites. The trend for the field is thus toward greater onboard intelligence, autonomous operations, and operational security.

The regulatory aspects of application satellites are an area where there may be more dissimilarities than there are common elements. The field of telecommunications satellites has often served as the major shaper of international, regional, or national regulatory policies. This logically follows from the following key points: (1) Satellite communications represent the earliest major satellite applications and thus first shaped regulatory policy and decisions and process related to radio frequency allocations; (2) Satellite communications services represent by far the largest market for commercial satellite applications; (3) Satellite communications services (to a much greater extent than other satellite applications) tend to represent a competitive service with respect to terrestrial communications networks, domestic industries, and program service providers; and (4) Trade regulations, landing licenses, intellectual property right protections, competitive access to local markets, particularly those as addressed by the World Trade Organization have largely concentrated on satellite communications in rule making processes (largely for the reasons indicated in 1, 2, and 3 above (Pelton 2005).

A company or entity wishing to operate a satellite application business within a particular country will thus often find that the rules for being able to operate, sell services, have "landing rights," or protect their intellectual property will find a set of regulations largely defined within the context of satellite communications. The same is true with regard to how the World Trade Organization (WTO) addresses issues of opening of national markets to competition and what national regulatory practices are considered acceptable and valid. As the transition is made from "voluntarily declared" open competition policies for services to mandatory plans that are overseen by the WTO with the authority to impose fines for noncompliance, the rules for market entry and competitive access to commercial space markets will likely become more contentious for all types of satellite applications.

Common Market and Business Considerations in Application Satellite Programs

The global commercial markets for satellite applications, as noted above, are becoming increasingly competitive on a global basis. This may not only make the regulatory aspects of the business even more challenging in terms of national landing rights, requirements to maintain local offices equipped with marketing, regulatory, and technical staff, but will also tend to make market competitiveness even more important. Satellite operators that have the advantage of economies of scale will likely be able to negotiate reduced prices or obtain price advantage through competitive contract awards for satellite purchases, launch services, and launch insurance arrangements. These larger organizations will also be able to obtain TTC&M services at lower net costs simply because they have more satellites in orbit and they can spread these costs for these services more efficiently in terms of net operational cost per satellite. The same type of economy of scale also applies to global marketing efforts as well. For those commercial satellite organizations that are selling their products globally, the market and business elements will likely see:

- A very highly competitive market
- Economies of scale continuing to be extremely important
- The challenges of being compliant with all regulatory requirements ever more difficult at the national, regional, and international level
- Issues related to regulatory licensing and landing rights at the national level more laborious with increasing likelihood of these requirements serving as "nontariff barriers to competitive entry"

- Access to necessary frequency allocations more challenging and intersystem coordination more difficult – in large part due to the expanding number of commercial satellites
- Market domination by the largest of the suppliers

Certain innovations can aid commercial business planning. One of the most important innovations can be lifetime extension that allows operators to do more without needing to make large capital investments simply because their satellites can last longer in orbit. Innovation in autonomous operation can serve to reduce operating costs. One of the biggest challenges could be the development of alternative technology. There is continuing R&D to develop high-altitude platform systems that might be able to provide wide areas of telecommunications, broadcasting, remote sensing, and surveillance as well as even meteorological services. There could also be development of multipurpose, large-scale spacecraft buses that could offer a variety of services that might cut across lines of the traditional commercial satellite application industry. Satellite with multi-payloads could offer telecommunications, broadcasting, and/or remote sensing services. At this point such developments can and should be considered long shot possibilities, but one should always be attuned to the fact that technologies, system economics, and markets often change when innovations emerge.

The New Small Satellite Constellations

The latest development that is impacting the design and deployment of satellite applications systems involve the deployment of small satellite constellations, most typically in low earth orbit between the altitudes of 600–1200 km altitude. This approach started with communications satellite networks such as the Iridium and Orbcom constellation. These networks were manufactured by Motorola and Orbital Sciences (now Orbital ATK), respectively, and were seen as outliers to the mainstream deployment of sophisticated, large, and high-capacity networks in GEO orbit. More recently, however, the development of advanced manufacturing processes involving 3D printing and new lower-cost launching options seems to have opened up a new and significant approach to cost-efficient and highly capable satellite applications systems.

The deployment of the Skybox (now owned by Google) and Planet Labs small satellite constellations for remote sensing has suggested a new pathway forward for future commercial Earth observation networks. On the communications satellite concepts, the development of O3b (a medium earth orbit constellation) followed by the OneWeb large-scale small satellite network with a network of 768 satellites plus spares has suggested that there might be a new paradigm for satellite networks optimized for Internet services.

This new approach clearly has pluses and minuses that will be affected by many variables such as new lower-cost launch options, advanced manufacturing processes involving 3D printing and more, frequency and orbital positioning allocations and interference concerns, concerns over orbital debris reduction, and new earth station designs, manufacturing, and electronic beam formation. Today the deployment of small satellite constellation seems to be a key development in the design arsenal for at least telecommunications and remote sensing networks, and it may prove a useful future tool for meteorological satellite networks and satellite navigation networks of the future as well. Currently, the long-term implications remain to be seen as to whether these types of networks will develop as yet another option or an even more powerful trend that defines the future of satellite applications (Pelton and Jakhu 2013).

Dissimilar and Unique Requirements in Application Satellites

The analysis up to this point has largely focused on commonalities among the various application satellite technologies, the various operating systems, and satellite sparing practices as well as the TT&C services. It has also been noted that regulatory and business practices and considerations are often common as well. Although there are many common trends and parallel patterns for the various types of application satellites, it is perhaps equally important to note the important dissimilarities that also do exist.

Differences That Stem from the Differences in the Various Satellite Markets

Clearly there are different service and availability requirements that impact the technical design of satellites and also their operational management. All the three big types of telecommunication services (broadcasting, fixed, and mobile) involve real time communications and thus exacting availability standards. The Integrated Services Digital Network (ISDN) standards only allow for about an hour of outage in an entire year. Data relay (or machine to machine (M2M)) satellites can perform to a lesser standard but there are also high availability expectations for this type of satellite as well. Likewise, space navigation satellites that are now use to support aircraft takeoff and landing must have the highest availability service of any type of application satellites. This is why there is extensive deployment of spare satellites in both telecommunications and space navigation satellite systems. Likewise, there is often redundancy of critical components within these satellites.

This is not to say that in the case of meteorological and remote sensing satellites, there is not a need for reliability and continuous system availability. Nevertheless, the "sparing philosophy" is not as exacting for these systems. There can be, for instance, a reliance on the meteorological and remote sensing satellite systems of other countries in the case of unexpected outages.

Since space navigation satellites are used for strategic and defense-related purposes and since these satellites are also use for a number of applications that are human life dependent, these satellites are designed to the highest standards in terms of radiation hardening, redundancy, and system sparing. This is closely followed by telecommunications satellites, with defense-related satellites designed to the highest system availability standards and telecommunications satellites designed to achieve 99.98 % availability or better in most instances. This is a particular challenge for satellites that operate in the Ka-band frequencies and higher due to loss of signal from high rates of rain attenuation. This requires a high level of link margins for satellite beams that experience intensive rainfall.

The remote sensing and meteorological satellites are also designed to the highest standards with redundancy of critical components. Nevertheless because of sharing of data and other precautions, service availability at such a high rate as 99.98 % is not an engineering requirement. In some instance, spare satellites are maintained on the ground and not launched until needed. The same approach can also be used in some cases by national communications satellite operators, particularly when there is spare capacity that might be obtained from other satellite or terrestrial sources. Finally the tracking, telemetry, and command and monitoring systems for space navigation and satellite communications must be engineered in a more exacting fashion to be able to monitor operations on a 24/7 basis and to be able to respond to anomalies or interference on a virtually instantaneous basis. In the case of meteorological systems, the same degree of global interconnectivity for TTC&M systems is not seen to be as critical.

Another very obvious difference is that the commercial satellite telecommunications markets worldwide are significantly larger than other commercial satellite applications. This difference does not have a technical impact, but it does have a business and operational impact.

The commercial satellite market offers the largest operators opportunities to achieve economies of scale in many elements of their procurements and operations. At least as far as spacecraft procurement goes for the rest of the industry they can benefit from the use of similar spacecraft platforms, solar arrays, battery systems, and perhaps TT&C systems, but the rest of the spacecraft is likely to be highly specialized and with high engineering and nonrecurring costs.

The commercial satellite industry is also the largest by far of all the commercial application satellite services. The business models and the nature of “customer relations” are, not too surprisingly, different from those of the entities that own and operate space navigation, remote sensing, and especially meteorological satellite systems. The satellite communications markets, especially for direct broadcast satellite services and broadband Internet access, have made the greatest transition toward selling “retail services” directly to consumers. In the case of space navigation, remote sensing, and meteorological satellites, the “marketplace” for their services remains predominately at the “wholesale” level in terms of the size of the revenues that support the operation of the satellite networks. Individuals may pursue space navigation units and farmers may purchase remote sensing data, but these “retail purchases” provide modest or no support for the operation of these satellite networks.

It is largely governmental agencies, large corporations, or international organizations that make the major purchases that support the operation of the “other” application satellite networks. It was only when direct broadcast satellite companies were able to sell on a retail basis their services to consumers that these industries’ revenues increased sharply in value. The same is likely true for the other parts of the commercial satellite industry. Thus it is only when successful business models that allow application satellite providers to sell their services directly to individuals on a "retail basis" that these industries can be expected to experience exponential growth of revenues. There are currently plans for the deployment of new types of application satellites to beam back electrical energy from space-based solar power satellites. Most of the business models for these new projects involve selling their solar-derived energy directly to energy companies. Only if these new companies can find a business model whereby they could sell their services directly to consumers, can they expect to achieve true commercial success and high levels of profitability.

Design and Engineering of Satellite Payloads: Communications Antenna Systems, Multispectral Sensors, and Radar Systems

The companies that contract for the integration and delivery of application satellites listed earlier such as Alcatel, Alenia Thales, Astra, BAE, Boeing, EADS, Lockheed Martin, Mitsubishi Electric Company, Motorola, Northrop Grumann, Orbital Sciences, and so on essentially design and manufacture the main elements of the spacecraft bus but rely on specialized companies to design and build specialized hardware such as an 18-m deployable antenna for mobile satellite communications, an active radar system or multispectral sensor for a remote sensing satellite, or an atomic clock for a space navigation satellite. It is the "payload" of an application satellite that defines its purpose and may represent a third or more of the cost of the entire satellite project. While there is considerable synergy in the design, engineering, test, and operation of application satellites, there is often little to none when it comes to the actual payload of a particular satellite.

It is the objective of commercial satellite manufacturers to make the design, engineering, manufacture, and test of the building blocks of an application satellite (i.e., bus structure, power system, TTC&M, stabilization and orientation system, thermal control system, etc.) as common as possible for various platforms that could be used for communications, remote sensing, space navigation, and meteorological observation. This, as previously observed, helps reduce costs, aids reliability, and speeds production and testing. The "non-common elements" of the payload represent the largest technical challenge and this highly specialized part of the spacecraft is typically designed and built by other contracts that specialize in particular payload that is required.

Conclusion

Application satellites are now a key part of the world economic structure. Billions of consumers rely on satellites to receive their daily news, see sporting events, to access the Internet, to know of current weather conditions, and especially to learn of threatening storms. Passengers on airplanes, on ships, and in cars and buses rely on space navigation systems for safe passage every day. The aerospace industry has learned to standardize many technologies related to the "spacecraft platforms" that are fundamental to the manufacturing of reliable and cost-effective application satellites. There are many technologies still to be explored to design better and lower-cost application satellite platforms, but progress continues to be made. Promising areas of future research and development (R&D) that can extend performance, increase reliability, and/or reduce costs are discussed within this chapter. One of the areas where there is considerable interest is finding even more reliable and cost-effective ways to deploy application satellites to orbit and make the launch services more competitive. One of the biggest differences among various satellite systems in terms of cost, operations, TT&C, and launch cost is whether the network employs LEO, MEO, and/or GEO satellites. Here there may or may not be commonality between various application satellite systems, but innovations do transfer from one market to another as better designs are found or improved ways are found to operate networks. This is particularly true with regard to LEO and MEO constellations. This is, however, largely due to the fact that there is now almost 50 years of experience with GEO satellite operations and the learning curve for these satellites is quite simply much longer.

Cross-References

- ▶ Lifetime Testing, Redundancy, Reliability, and Mean Time to Failure
- ▶ Overview of the Spacecraft Bus
- ▶ Telemetry, Tracking, and Command (TT&C)

References

L. Ippolito, J.N. Pelton, Satellite technology: the evolution of satellite systems, in *Communications Satellites: Global Change Agents* (Lawrence Erlbaum, Mahwah, 2004), pp. 33–54

J.N. Pelton, *Future Trends in Satellite Communications: Markets and Services* (International Engineering Consortium, Chicago, 2005), pp. 73–95. Chapter 6

J.N. Pelton, *The Basic of Satellite Communications*, 2nd edn. (The International Engineering Consortium, Chicago, 2006), pp. 119–140. Chapter 6

J.N. Pelton, R. Jakhu, *Small Satellites and their Regulation* (Springer, New York, 2013)

D.K. Sachdev, Three growth engines for satellite communications, in *AIAA 20th International Communications Satellite Systems Conference* (Montreal, 2002)

Part VII

Launch Systems and Launch-Related Issues

Launch Vehicles and Launch Sites

Joseph N. Pelton

Contents

Abstract

The Handbook of Satellite Applications focuses on the practical applications of satellites. This means that the handbook addresses the many uses that are made of communications, remote sensing, satellite navigation, and meteorological systems as well as the spacecraft, the ground systems, and tracking, telemetry, and command systems that make these networks possible. There are also chapters that address regulatory issues, economic and insurance issues, and even threats to the future operation of application satellites. This chapter addresses the remaining critical areas that are critical to the successful operation of application satellite systems.

All types of applications satellites could not carry out their function unless they were first launched into the right orbit. Even after successful launch they

---

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_72

must also be properly maintained there through necessary station-keeping operations. This chapter addresses the history of rocket and launch vehicle development and explains the basic technical capabilities that allow applications satellites to be placed into orbit with greater and greater reliability. This chapter also briefly addresses in-orbit operations that allow spacecraft to be maintained in orbit and to operate over increasingly long practical lifetimes. Over the past 60 years of the space age, an expanding variety of different propulsion systems and launch systems have been developed to carry out the important tasks of launch, station-keeping, and deorbit or removal of spacecraft to a graveyard orbital location.

One of the key elements of success for applications satellites of all types is the fact that gradually the reliability and the lift capability of launch vehicles have improved over time. It has been hoped for many years that new technology could allow the cost of launches to be significantly reduced, but to date such breakthroughs in the economics of launch systems have not yet been achieved. The precision thruster systems that allow spacecraft to be pointed with ever greater precision and to maintain crucial station-keeping have quite successfully continued to evolve. This has allowed application satellites to operate for much longer lifetimes and with greater pointing accuracy that has increased their functionality. Further developments that have most recently occurred in launch systems and new commercial launch sites are addressed in the following chapter.

Keywords

Avionics • Celestial mechanics • Computer guidance • Fairing • Guidance • International Telecommunication Union (ITU) • Ion engine • Launch vehicle • Launch sites • Liquid fuel • Propulsion • Rocket stages • Solid fuel • Station-keeping • Thrusters

Introduction

The history of rocketry and propulsion is actually quite long and complex, but it has only been in the last 60 years that practical launch systems have been developed and implemented. The development of rocket systems has, in many ways, been a dual pathway forward. One pathway has been the improvement of missile systems developed by military organizations as weapons for the delivery of bombs, and this approach has largely focused on solid rocket systems that use the controlled force of solid explosives to deliver a payload to a particular location against an adversary. The other pathway has been that of civilian space programs that have tended to focus more on liquid propellant systems to launch applications or scientific satellites to orbit or even to launch payloads with humans aboard into space. Liquid-fueled rockets cannot be fired instantaneously like solid-fueled rockets that, in effect, can be launched essentially by the push of button that is akin to lighting the fuse on a bomb. Both types of launch systems have paid an important part in space activities over the past 60 years.

The purpose of this chapter is to explore what types of launch capabilities exist to place application satellites into orbit and the newly evolving capabilities that allow this to be done more effectively, more reliably, and at lower cost. This chapter will also address the importance of launch sites and launch range safety in the successful launch of applications satellites. Finally the future of launch systems will be briefly explored. There are many interesting and in-depth books on launch systems that provide greater detail on the rocketry and launch sites that will be indicated in the endnotes for those who wish more information on this subject (Lennick 2006; Taylor 2007). One of the most important features of the Handbook on Application Satellites are the two appendices provided at the end of this book that report in detail about all of the many launcher systems that are available from around the world today as well as the many launch sites. The new growth and development of commercial launch systems that characterizes the world of rocketry today will undoubtedly bring a wider range of capabilities to the owners and operators of application satellites in the decades ahead. The latest innovations in launch systems are addressed in the following chapter.

Early History of Rocket Technology and Systems

The concept of rocket propulsion is actually thousands of years old. Archytas of Tarentum (428–347 BC), who was friend of Plato and an outstanding mathematician and scientist, discovered the principles of propulsion almost 2,500 years ago. He devised a wooden pigeon that used steam-powered jet propulsion to fly around Archytas' home in ancient Greece. Ancient Chinese speculated about the use of gunpowder-powered rockets to fly into space. Through the ages, writers of science fiction suggested the use of rockets to fly to the Moon and beyond. Everett Edward Hale in the nineteenth century actually wrote of launching an application satellite – a Brick Moon – into polar orbit for the purpose of navigation and communications (Pelton 1981).

With Newton's discovery of the gravitational effects of the Earth and his explicit description as to how a projectile fired with enough speed could escape the Earth's "gravity well," however, eventually gave rise to serious thought about rockets and space travel. The Russian scientist Tsiolkowsky conceived of rocket designs that could actually carry people into space and others such as Willy Lev began a systematic study of space and rocketry. The American Robert Goddard is considered by many to be the modern-time father of rocketry. Beginning in the 1920s, he developed a series of increasingly powerful prototype rockets with liquid fuel propulsion and elaborate stabilization systems. Although mocked for his experiments at the time by the *New York Times* as the "Moon Man," in part because his rockets had limited range of his experiments, he persevered and developed more and more capable launchers. Goddard's work was increasingly taken seriously and his efforts rather directly led Wehrner von Braun and his team in Germany to develop the V-2 rockets as bomb delivery systems during the Second World War.

After the Second World War, Soviet scientists and engineers continued work in rocketry to develop missile weapons systems as well as rockets that could achieve orbit as did the United States but with much less funding and concerted governmental support behind the American efforts. This all changed in October 1957 when the Soviet Union launched the Sputnik I satellite into Earth orbit. The United States formed the National Aeronautics and Space Administration (NASA) in 1958 and gave significant new funding to develop missile technology within the US military as well as support to Dr. von Braun's team in Huntsville, Alabama to develop civilian rocketry capability.

The years that followed gave rise to a major "missile race" between the United States and the Soviet Union. The US election for President in 1960 won by President John F. Kennedy was largely focused on what has called the "missile gap" between the United States and the Soviet Union. The United States invested heavily in developing civilian rocket technology and even more money was invested in military rocket and missile systems. The 1960s culminated with the Apollo 11 mission that sent three astronauts on a lunar exploration mission that allowed two astronauts to land and explore the Moon's surface in July 1969. Since the start of the Space Age over 500 people have gone into space via rocket launches and a number of space stations have been launched to sustain astronauts and cosmonauts in orbit for sustained periods of time.

In the new age of commercial space travel, it is possible that a much larger number of people will be able to travel into space. Today not only are commercial spaceplanes under serious development for flights starting in 2013 but private space stations are planned for launch as well. The idea of people being able to travel into space on commercial spaceplanes was quite vividly envisioned in the Stanley Kubrick and Arthur C. Clarke movie *2001: A Space Odyssey* and today that prospect seems to transcend the jump from science fiction to science fact.

Development of Solid-Fueled Missile Systems

As noted earlier, there are two basic kinds of chemical propellants for missiles and rocketry systems. These are liquid- and solid-fueled launch systems. In both cases, the fuel is "oxidized" or ignited to create a powerful discharge of gas in order to provide the needed chemical propulsion or "thrust" to lift the rocket skyward. More recently, electrical propulsion systems have also been developed. These systems (i.e., ion thrusters) have much lower thrust but can operate for much longer periods of time and provide more net thrust over time. In the case of electrical propulsion, ions are expelled at very high velocities for guidance and station-keeping systems. These systems can be more reliable and provide higher thrust to mass ratios than chemical propulsion systems but do not have sufficient thrust to provide lift off from the Earth's surface. These thrusters, however, can be used for station-keeping and final orbital positioning. In time, nuclear-fueled systems or other more exotic capabilities such as tether-lift systems, rail guns, or even the so-called space elevator may be used to provide access to Earth orbit, but today chemically powered rockets are still the exclusive way for application satellites to attain orbital access.

Solid- and liquid-fueled rockets both have advantages and disadvantages. Liquid propellants tend to require complicated piping and very high-performance pumps to feed the rocket engines with a large and steady stream of rocket fuel. Liquid-fueled rockets can provide greater net propulsive thrust over their full period of operation, but they require time to fuel and thus quick liftoff is not possible. Further elaborate storage, handling, and fueling systems are required for this type of rocket. Also from a safety perspective, it is possible to throttle and precisely control the flow of fuel to the combustion chamber in the case of the liquid-fueled rocket. This means that a liquid-fueled engine can be immediately shut down by closing a valve that shuts off the flow of propellant to the engine. A liquid-fueled rocket is more complicated in design and slower to fuel and launch. The liquid-fueled rocket is also slower to build up thrust because of the pumping of the fuel into the combustion chamber.

Solid rockets do not require complicated engines, pumps, or plumbing. Instead the thrust of these rockets depend on the explosive power of the solid-rocket fuel and require stronger casings to endure the great pressure that come from the exhaust thrusts. Since they are essentially a controlled bomb, they can be ignited much more rapidly and their initial acceleration at liftoff is nearly instantaneous. From a safety viewpoint, however, most solid fuel systems (except for the hybrid systems that use nitrous oxide as an oxidizer and neoprene rubber as the fuel) cannot be throttled or controlled once ignition has been achieved.

The above factors are some of the considerations that are taken into account in terms of turning to solid fuels for missile systems as well as decisions to utilize liquid-fueled systems for systems involving astronauts. Such flexibility as to what propulsion system to use was not readily available at the start of the space age. This was simply because liquid-fueled systems became available first – starting with the Robert Goddard developments and the V-2 which were liquid-fueled systems.

Thus the V-2 in Germany; the Atlas, the Thor, and the Jupiter in the United States; and the R-7 ICBM in the Soviet Union were all liquid-fueled systems. These systems took a fair amount of time to launch. They had to be loaded not only with a fuel but also with the oxidizer, and this loading process was not only time consuming but also quite dangerous because a spark could set off a very dangerous explosion (Spaceflight).

As noted above, the first impetus to design and build launchers was in the context of war and weapon systems. Both the United States and the Soviet Union in the post–Second World War time period proceeded to build missiles, equipped with atomic bombs and in time with hydrogen bombs. These weapons systems served as primary deterrence against attack during the Cold War that existed between the United States and the Soviet Union for the decades that followed.

For the reasons noted above, there was a concerted effort to develop solid-fueled systems for quick, easy to fuel, and less hazardous fueling and launching operations. In the United States, the limited capacity Scout rocket, the Polaris, and the Minuteman systems were developed using solid-fuel propellants. The Minuteman – named for its quite launch capability – was designed to be suitable for instantaneous launch from land-based missile silos and the Polaris was developed so that it could be launched from submarines. Clearly, a liquid-fueled system would be extremely difficult to deploy from a submarine, although the Soviet Union managed to equip their largest submarines with liquid-fueled systems for a number of years because of problems in developing solid-fueled systems (Spaceflight: rockets and missiles. http://centenialofflight.gov/essay/SPACEFLIGHT/solids/SP13.htm).

In the Soviet Union the liquid-fueled R-7 was used not only to launch Sputnik but also employed for the manned missions that followed – starting with Yuri Gagarin's flight. The USSR continued to rely on the R-7 for a decade even though there were serious problems with loading of fuel on this missile and it was very slow to prepare for launch. It was not until 1971 – almost a year after the United States began deploying solid-fueled systems that the Soviet Union deployed the first RT-2 system. In fact the liquid-fueled systems remained in service as Soviet weapons systems for another decade. The Soviet Union and now Russia continue to rely primarily on liquid-fueled systems for its civilian space program.

Today the number of countries with some launch capability continues to increase as can be seen in Appendix 3. These countries include China, France, Germany, India, Iran, Israel, Japan, the Republic of Korea, the People's Republic of Korea, Russia, the Ukraine, the United States plus the countries of the European Space Agency.

Liquid-Fueled Launchers

Although some small satellites have been launched by the solid-fueled Scout vehicle, most of the commercial applications today are launched by liquid-fueled vehicles. These liquid-fueled systems almost always fall into the three categories of petroleum-based fuels (most typically a highly refined kerosene known as RP-1), cryogenic fuels (most typically liquid hydrogen), or hypergolic fuels (most typically some form of hydrazine). The Chinese Long March vehicle uses a form of hydrazine (known as UDMH) in its first two stages with nitric acid acting as the oxidizer and some Russian vehicles use hydrazine as well, but virtually all other major launch vehicles rely on either liquid hydrogen (LH_2) and liquid oxygen (LO_2) or on RP-1 rocket fuel with an oxidizer (typically LO_2). In some instances, there are solid fuel strap-ons to supplement liftoff capabilities. The following table adapted from data prepared by Robert A. Braeunig indicates the specific impulse of various types of rocket fuels when working with their various oxidizers (Braeuni 2008) (Table 1).

Table 1 Rocket propellant performance (Adapted and simplified from a chart prepared by Robert Braeunig)

Combustion chamber pressure, $P_c = 68$ atm (1,000 PSI). Nozzle exit pressure, $P_e = 1$ atm				
Oxidizer	Fuel	Hypergolic	Mixture ratio	Specific impulse (s, sea level)
Liquid oxygen	Liquid hydrogen	No	5.00	381
	Liquid methane	No	2.77	299
	Ethanol + 25 % water	No	1.29	269
	Kerosene	No	2.29	289
	Hydrazine	No	0.74	303
	MMH	No	1.15	300
	UDMH	No	1.38	297
	50-50	No	1.06	300
Liquid fluorine	Liquid hydrogen	Yes	6.00	400
	Hydrazine	Yes	1.82	338
FLOX-70	Kerosene	Yes	3.80	320
Nitrogen tetroxide	Kerosene	No	3.53	267
	Hydrazine	Yes	1.08	286
	MMH	Yes	1.73	280
	UDMH	Yes	2.10	277
	50-50	Yes	1.59	280
Red-fuming nitric acid (14 % N_2O_4)	Kerosene	No	4.42	256
	Hydrazine	Yes	1.28	276
	MMH	Yes	2.13	269
	UDMH	Yes	2.60	266
	50-50	Yes	1.94	270
Hydrogen peroxide (85 % concentration)	Kerosene	No	7.84	258
	Hydrazine	Yes	2.15	269
Nitrous oxide	HTPB (solid)	No	6.48	248
Chlorine pentafluoride	Hydrazine	Yes	2.12	297
Ammonium perchlorate (solid)	Aluminum + HTPB (a)	No	2.12	266
	Aluminum + PBAN (b)	No	2.33	267

Notes:
Specific impulses are theoretical maximum assuming 100 % efficiency; actual performance will be less
All mixture ratios are optimum for the operating pressures indicated, unless otherwise noted
LO_2/LH_2 and LF_2/LH_2 mixture ratios are higher than optimum to improve density impulse
FLOX-70 is a mixture of 70 % liquid fluorine and 30 % liquid oxygen
Where kerosene is indicated, the calculations are based on n-dodecane
Solid propellant formulation (a): 68 % AP + 18 % Al + 14 % HTPB
Solid propellant formulation (b): 70 % AP + 16 % Al + 12 % PBAN + 2 % epoxy curing agent

Avionics and Guidance Systems

Many people think of rocket launcher systems as powerful explosive systems that power a rocket to orbit and do not necessarily consider that there must not only be great thrust achieved through the rocket engines but that there must also be a very accurate guidance systems that steers the rocket and its payload on the exact pathway needed to acquire the proper orbit. The celestial mechanics needed to calculate the proper trajectory to get to the desired orbit is quite challenging indeed. Without very fast computing and guidance systems known as the avionics system, the launch of modern applications satellites would not be possible. Gyroscopic systems onboard the launcher send back to command centers data as to the rocket's exact location to the Earth so that the thrust vectors of the rocket and the gimbals on the rocket motors can be corrected to keep the launcher on the exact orbital trajectory. Millions of data points and calculations are made at extremely rapid speeds to keep the rocket exactly on course in all three x, y, and z reference planes (i.e., of roll, pitch, and yaw).

Completely different trajectories must be calculated for launches to polar, LEO, MEO, or GEO orbits. There are even different calculations and trajectories created if a rocket is launching a single payload or multiple payloads. Most modern launchers are multiple-stage rockets that must separate between firings of the stages and these elements of the launch must be accommodated by the avionics system and the tracking data to make sure there is a successful launch. Even the separation of the rocket's nose fairings at high altitudes and the spinning of the spacecraft at the time of release from the rocket are parts of the launch operation that need to be carefully planned and monitored in real time via telemetry sent down to the launch command center and tracking stations around the world that monitor and control every aspect of the launch.

This constant telemetry monitoring operation is not only critical to the successful launch operation but has another important element as well. Especially trained range safety officers are also directly involved in the launch operations. In case the rocket should for some reason go off course, there could be a need for the errant rocket to be destroyed in order to avoid loss of human life and perhaps buildings or even cities on the ground. Launch insurance arrangements typically include liability protection against an unsuccessful launch and to protect in particular for a rocket that might go off course and thus lead to substantial damages or lead to causalities. The possibility that a rocket could actually land in a city – such as a launch from the Kennedy Space Center landing in Miami, Florida – could lead to an incredible level of destruction and loss of human life, and thus special liability arrangements have been made in such a case that when commercial insurance liability limits are reached, an additional layer of liability insurance is actually provided by the US Government – as do the governments of other countries that support launch operations. In this case, however, the security precautions that are in effect and the reliability of today's launch vehicles make such a possibility extremely remote. The insurance arrangements and costs, however, do strongly motivate the site location process for launch sites and space ports to be situated in remote areas with launch operations being conducted either over the ocean or in extremely sparsely populated desert locations.

Launch Options for Commercial Application Satellites

The dominant provider of commercial launches today is Arianespace with their Ariane 5 vehicle. The current dominance of Arianespace in providing launch services to commercial satellite application providers is based on a combination of reliability, launch efficiency (i.e., the Ariane 5 launch site in Kourou, Guyana is essentially right on the equator and thus provides the maximum Earth-assisted boost to GEO orbit), and effective pricing.

Certainly the Ariane 5′s large lift capability provides economies of scale. This allows the launch of very massive satellites or the joint launching of several satellites to GEO at the same time. Recently, a new launch facility has been constructed at the Kourou launch site for the launch of Russian designed and built Soyuz and Soyuz-2 launch vehicles. This new capability at the Guyana launch facility will provide a wider range of launch options that can now be available from this equatorial site (Fig. 1) (Soyuz launch).

There are sufficient launch options available in the global marketplace to ensure competitive pricing. There are over a dozen competitive launcher entities offering launch services today and literally hundreds of launch options to meet needs for launch to low earth, polar, medium earth, highly elliptical, or GEO orbits as is obvious by a review of the information provided in Appendix 2.

The launch options and launch site opportunities are constantly changing. It is now possible to arrange for a Soyuz-2 launch from the Kourou launch site. The Atlas V and Delta II now provide relatively high-lift capabilities and more attractive prices than in the past. It is possible that the Atlas, Delta, or other US vehicle might be upgraded to “man-rated capabilities” and the new NASA Space Launch System (SLS) that will launch the Orion capsule into deep space may give rise to new launch options in future years (NASA Space Launch). Currently, both the Orbital Sciences Corporation and SpaceX are seeking to develop commercial launch capability that could safely fly cargo to the International Space Station and perhaps in time even fly astronauts to orbit as well. These new capabilities could also ultimately help to serve future launch needs for application satellite service providers (Lindemoyer 2011).

Perhaps most significant in terms of changing launch options are the Pegasus and Taurus launch capabilities from Orbital Sciences and especially the new Falcon launch vehicle that are truly “commercial vehicles” that appear to be able to offer increasingly cost-effective new launch options.

Cost of Launches and New Commercial Options

It is not possible to cite a specific launch cost for a particular class of vehicle. First of all the cost of a low earth orbit launch is much less than that of a GEO launch or one can lift much more mass to LEO for the equivalent cost of a GEO launch. A larger satellite with greater mass can often attain a more cost effective rate on a per

Fig. 1 Ariane 5 launch from Kourou, Guyana launch site (Graphic courtesy of Ariane space)

kilogram basis. The interface requirements, nose fairing, and other special requirements can also affect cost. In today's market, a cost of $10,000–20,000/kg for a satellite launch to GEO orbit is not uncommon. In general, these costs are expected to come down, particularly driven by new commercial launch vehicles services that offer lower costs in coming years.

Fig. 2 The launch of a Falcon 1 vehicle from its Kwajalein launch site (Graphic courtesy of spaceX)

Also most organizations will take out launch insurance that is often equivalent to 15–20 % of the total mission cost, even though this also can vary to higher or lower levels depending on the mission. Some organizations choose to self-insure.

The series of launch vehicles developed as a total new launch vehicle and a strictly commercial venture are promising even more competitive pricing and more economical service without giving up reliability. The key to the SpaceX approach has been complete vertical integration and thus designing and manufacturing all elements of the launcher. A number of organizations, perhaps most notably the US Air Force, has signed up for a number of Falcon launches. If these launchers indeed prove to be reliable and the costs are significantly less expensive than other launch options, this then will clearly impact the global market for these services (Fig. 2) (Falcon launch vehicle).

Some governments, particularly the United States, have significant restrictions as to what technology can be shared or released to other countries. The so-called ITAR (International Trade in Arms Restrictions) prohibit or restrict the launch of some satellites with “sensitive technology” to be launched by certain other countries because of concerns about unauthorized transfer of technology.

Launch Sites

Appendix 1 provides a listing of available launch sites around the world. These are distributed among many countries around the world. Many of these sites are in reasonably close proximity to the Equator. This is because the Earth's rotation speed at the Equator which is in excess of 1,600 km/h provides significant assistance to a satellite launch into GEO orbit which is located in the equatorial plane. This means that launch sites essentially located at the Equator such as Kourou, Guyana, or the Sea Launch provides the optimum location for a GEO launch. The launch sites that are further away from the Equator such as the sites of China, Japan, and the United States thus can be on the order of a 20 % disadvantage in relation to the locations exactly on the equatorial plane.

There are many other sites that launch into a polar orbit, LEO, MEO, or highly elliptical orbit. These sites are much less constrained as to their geographic location. A prime consideration for a launch site in all cases is to have a location that is considered as safe as possible. This would involve the possibility of a flight path that would not cover populated areas and would allow for safe aborted launches in the case the range safety officer believed termination of flight vehicle was necessary. Today most of the launch sites are those designed for vertical liftoff of chemically fueled vehicles. Under the regulations of the United States and some other countries, there is a need for an environmental impact statement to be issued prior to each launch.

There are a number of commercial "spaceports" now in planning or approved. These sites vary a great deal in terms of accommodating different types of flight options. Some spaceports include provision for only horizontal takeoff and landing with winged vehicles and these facilities greatly resemble a conventional airport. Other commercial spaceports are in many cases collocated with a governmentally licensed or owned launch site and are designed to accommodate both vertically and horizontally launched vehicles. These facilities are designed not only with launch pads but with specially designed fuel storage facilities and specially designed buildings and observation towers for space range officers.

Station-Keeping, Spacecraft Operations, and End of Life

The chemical rockets that are used to launch a spacecraft into geosynchronous orbit or to position satellites within a constellation of satellites in medium or low earth orbit is actually just the start of a process of orbital operations that can last for many years. The lifetime of an application satellite varies for a variety of factors. Orbital altitude is one important factor. Low earth orbit satellites, in particular, because of their lower altitude are subject to atmospheric drag that lessens their useful lifetime. Most of these satellites at the end of life are subject to a controlled descent and splash down into one of the earth's oceans. Satellites launched into MEO have intermediate lifetimes but pose the largest challenge at end of life. These satellites are not high enough to be put into so-called graveyard orbits and not low enough to be easily deorbited in a controlled manner. Fully 40 % of the station-keeping-fuel needed for orbital operations must be devoted to proper end-of-life deorbit.

The Geosynchronous orbit, because it is one tenth of the way to the Moon, does not present an atmosphere drag issue. Since it is so far removed from the Earth's primary gravity well, GEO satellites present perhaps the easiest condition for in-orbit operations and end-of-life operation. At the end-of-life, if fuel remains, operators merely push the satellite into a higher orbit where it will remain for millions of years in a so-called graveyard orbit. Even for the GEO orbits, there are tradeoff considerations. It is at least ten times easier to maintain a GEO satellite within its "station-keeping box" in terms of East–West excursions as compared to North–South deviations above or below the GEO arc. As a satellite in GEO orbit nears its end of life, operators often relax their maintenance of North–South station-keeping so that it can drift North and South of the equatorial plane in an "figure 8" shaped orbit in order to save fuel. Under International Telecommunication Union (ITU) regulations, a satellite is considered in GEO orbit as long as it remains within 5° of the equatorial plane.

There are thrusters onboard application satellites that can be fired to either reposition a satellite from one orbital position to another or to maintain it in its registered location under the official filing with the ITU.

In the earlier days of application satellites, most station-keeping thruster systems used a hypergolic fuel system (and most typically hydrazine thrusters) to maintain satellites in the desired orbit. More recently, there has been more and more common use of ion engines to maintain application satellites in orbit. This is because electrical propulsion systems provide lower-impulse thrust levels than chemical systems in a single burn, but they nevertheless allow for a longer operational life since they provide higher net thrust capability per kilogram of fuel. Another consideration that is increasingly coming into consideration is that hydrazine fuel is quite noxious and decaying satellites with fuel can explode and create orbital debris. Thus, even though ion-engine control systems are more expensive than hydrazine or bi-propellant systems that have been used in the past, these thruster systems are becoming increasingly common for virtually all forms of application satellites.

One of the latest developments in the operation of application satellites is the idea of creating a space tug and refueling and maintenance system in orbit. Such a device could possibly add new batteries and refuel the tanks of application satellites so that they could have an extended "second life" of perhaps many more years. Although this concept is still at an early stage, there are active experimental programs underway involving MacDonald Dettwiler Aerospace (MDA) (the designer and builder of the robotic Canadarm device) and Intelsat to see if a satellite could be captured by a robot arm and refueled (MDA).

Conclusion

Rocket propulsion has evolved a long way in the last 60 years. Launchers that are 98 % reliable in terms of successful lift to orbit have now been achieved. New commercial space ventures are seeking to develop systems that could be far more reliable. Efforts to create launch systems that are dramatically more cost effective have eluded rocket developers to date. Launcher systems, however, have become increasingly dependable and certainly have increased their launch to orbit capabilities by orders of magnitude. Thus, while launchers are not greatly more cost effective in terms of the cost of a lifting a kilogram to orbit, their expanded ability to lift larger satellites to orbit has allowed the satellites themselves to become more capable and cost effective due to economies of scale. More and more countries and companies have developed launch capabilities. The most dynamic new element in this regard is the Falcon class launchers developed by the Space eXploration Technologies Corporation known as SpaceX that is developing not only lower-cost launch capabilities for the orbiting of application satellites but perhaps ultimately seeking to develop a commercial option to lift astronauts to the International Space Station (ISS). Today as indicated in Appendix 2 (Launch Sites) and Appendix 3 (Major Launch Systems) there are a wide number of options available to the satellite applications industries on a global basis.

Research continues to develop better launch capabilities, new technologies to lift satellites to earth orbit, new capabilities to maintain satellites in orbit, carry out extended station-keeping, and even in time possibly to allow the refuelling of satellites in orbit. Ultimately, new capabilities to place satellites in orbit that are even more reliable and cost effective will evolve. There are yet other challenges to be faced such as dealing with orbital debris and new launch and robotic capabilities may possibly be able to address these problems as well.

Cross-References

- ▶ Major Launch Systems Available Globally
- ▶ The World’s Launch Sites
- ▶ Trends and Developments in Launch Systems

References

Braeuni, R.A.: Basics of spaceflight: rocket propellants. http://www.braeunig.us/space/propel.htm (2008)

Falcon launch vehicle. http://www.spacex.com/falcon1.php. Last accessed 14 Jan 2016

M. Lennick, *Launch Vehicles Pocket Space Guide: Heritage of the Space Race Pocket Space Guides* (Space Pocket Guides, New York, 2006)

Lindemoyer, A.: COTS status: NASA advisory council. http://www.nasa.gov/pdf/580727main_4%20-%20Lindenmoyer%20COTS%20Status_508.pdf (2011). Last Accessed 14 Jan 2016

MDA to provide in-orbit operations and maintenance support. http://sm.mdacorporation.com/. Last accessed 14 Jan 2016

J.N. Pelton, *Global Talk* (Sitjhoff and Noordhoff International, Alphen aan den Rijn, 1981), pp. 13–20

Soyuz launch complex in Kourou Guyana. www.russianspaceweb.com/kourou_els.html. Last accessed 14 Jan 2016

Spaceflight: rockets and missiles. http://centenialofflight.gov/essay/SPACEFLIGHT/solids/SP13.htm. Last accessed 14 Jan 2016

T.S. Taylor, *Introduction to Rocket Science and Engineering* (CRC Books, New York, 2007)

The NASA space launch system. http://www.nasa.gov/exploration/systems/sls/sls1.html. Last accessed 15 Jan 2016

Trends and Developments in Launch Systems

Joseph N. Pelton

Contents

Abstract

The development of more reliable and lower cost launch capabilities has been the steady and consistent objective for a rocket scientist for many decades. Indeed, ever since the first rocket launcher capabilities were proven by Goddard, von Braun, and other early pioneers, the goal has been to create a better launch vehicle. For the past 50 years, however, the prime development work has largely been defined by developing improved chemically powered launch vehicles. The prime development efforts for solid missile systems have been driven by military programs, while liquid-fueled rockets were largely spear-headed by aerospace companies supplying launchers for civil aviation programs. During the past 2 decades, a growing capability to deliver reliable launch services has grown up

J.N. Pelton (✉)
International Space University Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_102

in the world with Japan, China, India, and the Ukraine offering capabilities that rival the governmental programs of the USA, Europe, and Russia.

The biggest change of all has been the advent of “new space” entrepreneurial companies seeking to develop new lower cost and human-rated spaceplanes and highly competitive launch vehicles. These new commercial initiatives have served to alter the course of launch vehicle development in a variety of ways. New cost models and new commercial applications have driven thought in new ways. Areas of focus now include consideration of new ways to use electric ion propulsion, nuclear ion propulsion, and the development of hybrid systems that combine solid fuels with an oxidizer in such a way to allow hybrid propulsion systems to be turned on and off. Other new concepts include more efficient ways to launch from higher and more efficient altitudes by using balloons, carrier vehicles, or even towing launch systems to airborne launch sites. Another key area of research involves the ability of launch systems to be reused so that the rocket launcher can return to a launch site to be used over and over again. This is in addition to spaceplanes that can be used for multiple missions.

At the research level, there are in fact over a dozen innovative ways that spacecraft and payloads could be placed into earth orbit. These range from concepts that have been actively researched by space agencies such as using nuclear heat to create ionic propulsion to exotic ideas for the future such as mass drivers, tether sky hooks, and even space elevators.

And in addition to plans to make launchers more cost-efficient, reusable, and reliable, there are also new concerns about the environmental effects of rocket launchers on the fragile upper atmosphere where the density of molecules is perhaps a 100 times less than at sea level. This has given rise to particular concerns about solid fuel rockets that emit particulates and are perhaps 100 times more polluting than liquid-fuelled systems.

New entrants such as Swiss Space Systems (S-3), Virgin Galactic, Sierra Nevada, SpaceX, and External Engines, Firefly, Blue Origin, Copenhagen Suborbital, Kelly Space & Technology, inc., Myasishchev Design Bureau, Interorbital Systems, Armadillo (now Exos Aerospace), Masten, Planespace, Scaled Composites, Rocketplane Kistler, Stratolaunch, XCOR, t/Space, Space Transport Company, Zero2Infinity, and Starchaser Industries have all contributed to a wealth of ideas about new, lower cost, safer, and more reliable ways to launch to orbit. Some of these start-ups have now failed and are defunct but their innovative concepts live on (J.N. Pelton, P. Marshall, *Launching into Commercial Space*. AIAA, Reston, 2015).

In this era of rapid innovation and change, established aerospace companies such as Boeing, Orbital ATK, Northrop Grumman, Arianespace, Astrium-Air Bus, Lockheed Martin, Raytheon, SeaLaunch, the United Launch Alliance, the Great Wall Company of China, and others are also seeking to innovate and create newer and better launch systems that keep current with the latest in launch technology and systems. In particular, they have been driven to find ways to cut cost in the face of new commercial space launch systems that their vehicles must

compete (J.N. Pelton, P. Marshall, *Launching into Commercial Space*. AIAA, Reston, 2015).

This chapter provides the latest updates on new launch systems and ends with a brief update about the impact of new commercial spaceports and launch sites.

Keywords

Blue Origin • Commercial spaceports • Electric ion propulsion • Firefly Space Systems • High-altitude launch • Hybrid propulsion • Launcher One • Nuclear ion propulsion • Reusable launch vehicles • Orbital ATK • Scaled composites • Sierra Nevada • SpaceX • Swiss Space Systems

Introduction

For many decades, the development of newer and better launch systems was led by civilian space agencies and civil space agencies and defense ministries. These governmental space organizations consistently worked together with large and established aerospace companies that have had long-term working relationships with governmental entities. Beginning about 20 years ago, this basic scenario of how launcher systems were developed began to change.

On one hand, a number of new countries joined the ranks of those with significant launch capabilities. Japan, China, India, and the Ukraine emerged as highly capable launch providers as noted in Appendices 1 and 2. But even more importantly, in terms of new trends in launcher development has been the rise of so-called new space commercial initiatives to create entirely new models for technology innovation in the space launch industry.

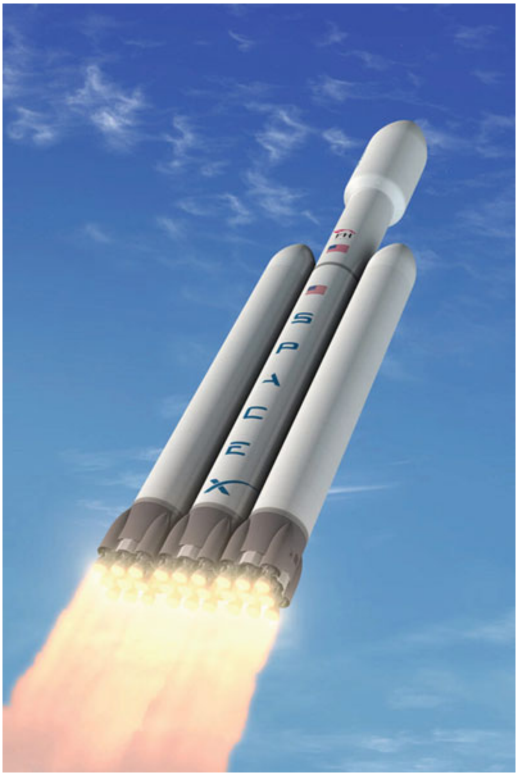
Exactly when this “new space” commercial revolution began is hard to say exactly, but many relate it directly to the start of the XPrize competition.

The initial XPrize initiative was created and announced at the US Smithsonian National Air and Space Museum by Peter Diamandis, a new era in commercial space launch technology and systems began. Instead of development of new launch systems as billion dollar programs by governmental agencies and major aerospace companies, this new era was driven by small start-ups that “thought outside the box” and conceived of entirely new ways of doing things. When the $10 million Ansari XPrize was firmly established in 2004, this so-called new space activity, especially focused on developing new suborbital flight spaceplanes began in earnest.

NASA also responded by creating its Commercial Orbital Transportation System (COTS) that encouraged new commercial capabilities to resupply the International Space Station. In “Phase D” of the program the option of having new commercial space transportation systems fly astronauts to orbit is now being pursued. Currently Boeing, SpaceX as well as Sierra Nevada are developing this capability under NASA contract.

In addition, a widened range of launcher capabilities have evolved to provide launch capability. These options range from the giant Ariane 6 launcher, now under development by Ariane Space, The Atlas V, Delta 4 Heavy, and the Falcon 9 Heavy

Fig. 1 The Falcon 9 heavy launch vehicle capable of lifting 5400 kg to LEO (Graphic courtesy of SpaceX)



down to the modest and small satellite launcher capabilities now being developed by Swiss Space Systems (S-3), Firefly Space Systems, IOS, and Launcher One. There are others that are developing much more capable electric ion propulsion systems, those that are developing huge carrier vehicles such as the Stratolaunch, and entirely new companies that have entered the launch services companies. Two of these new entrants, namely Elon Musk's SpaceX Company with his Falcon 9 launcher and Dragon Capsule, and Sierra Nevada with its Dreamchaser, have gone from start-up to a significant player in the launch industry in just a decade. Most recently NASA, in January 2016, has awarded some $14 billion in contracts to three launch service providers for 18 resupply missions to the International Space Station (i.e., six missions each). These suppliers are Space X (Falcon 9 launcher and Dragon capsule), Orbital ATK (Antares launcher with Cygnus capsule), and Sierra Nevada with the Dreamchaser reusable spaceplane. This is a remarkable change from who the suppliers would have been a decade ago (Davenport and Fung 2016).

The SpaceX Falcon 9 and Falcon 9 Heavy launch vehicles have been perhaps the most significant game changers in the entire history of launch vehicle development. These vehicles starting with the Falcon 1 just a decade ago came from virtually "nowhere" have now allowed SpaceX to become one of the predominant launch service providers (Falcon Heavy 2014) (see Fig. 1).

This chapter examines the rapid changes that have come to the launch industry, especially in the past year, and what some of the most important vectors of change have been in terms of propulsion systems and fuels, new approaches to launching systems, carrier vehicles for high-altitude launches, reusable launchers, environmental issues and concerns, and new concepts for the future.

Chemical Propulsion

The idea of a rocket launcher that uses chemical explosives as a means of propelling a rocket is quite simple. One simply channels the explosive reaction out a jet that thrusts a vehicle forward. In the fourth century BCE, Archytas of Tarentum mastered this technology with steam exhaust to get a wooden pigeon to fly tethered to a string. The trick was to get enough propulsive power, to channel it safely without blowing up, and to steer it in the right direction. It took quite a few centuries to figure how to do this safely and reliably. Solid-fueled rockets ultimately became missiles that could be sent off at short notice as weapons, but once fired they did not shut off until all the propellant has burned. Liquid-fueled rockets with a liquid fuel that is mixed with an oxidizer can be controlled and shut on and off and can be easily staged so the weight of the used up fuel tank can be jettisoned to make the rocket more effective.

The problem with chemical propulsion, however, is twofold. First, it really is a bomb that is dangerous to control and safety is thus always an issue. Second, the conversion of chemical explosions to propulsive thrust is really not all that efficient. The electric super heating of a fuel such as Xenon to create propulsive ions creates a lot less thrust, but this is a process that can continue for a long time rather than just a few minutes in a chemical explosion. The net thrust from an ion thruster over time turns out to be at least two to three times greater than a chemical explosion that just goes bang. The latest results with the NASA's NEXT (i.e., NASA Evolutionary Xenon Thruster) electric ion thrusters have proved to be 10–12 times more efficient in terms of a total thrust to mass ratio. It is also cleaner (Redd 2013).

Today's satellites rather than using hypergolic fuels like hydrazine gas (i.e., a chemical burn) for station-keeping and orbital redeployments tend to use electric ion thrusters instead. This is because a small amount of thrust is needed to keep the satellite on station and the net weight of the thruster system is less and the overall thrust capabilities are greatly increased for the lifetime of the satellite. The reason that electric ion thrusters are not used for spacecraft launches from the ground is that the thrust is much too weak. But more powerful electric ion thrusters are being developed with the NASA NEXT thrusters now producing over 12 times the total thrust of chemical thrusters over the lifetime of a satellite.

If spacecraft could be raised high enough on balloons, dirigibles, or carrier vehicles, it might be possible at a future time for an electric ion thruster system to lift small payloads to orbit. Organizations such as JP Aerospace have sought to develop a "dark sky station" at super high altitudes. Their engineers claim that eventually they could "fly" small payloads to low earth orbit. Certainly it is true that in the future, ion thrusters could raise the orbit of a satellite from low earth orbit out to GEO by flying a slow outward spiral. This could be a cost-effective deployment method for the future. The problem is that the increasing build-up of orbital debris in low earth orbit would entail some considerable risk of a collision between the payload and the space debris.

In time, other techniques that are safer, more cost-effective, and environmentally friendly may develop that are superior to chemical rocket propulsion. These might involve mass driver systems, tether sky hooks, electromagnetic systems, rail guns, or space elevators, but for the foreseeable future, chemical propulsion systems will remain the key technology for commercial satellite deployment. Even within the constraint of conventional chemically fueled rockets, high-altitude carrier vehicles, reusable launch vehicles, and more efficient design and manufacturing can produce significant new efficiencies. Innovations such as 3D printers, advanced manufacturing and automated testing, and other new ways of doing business has seemingly changed the space launcher business forever. Chemically fueled rockets systems that were seen as a “mature” technology have thus experienced a renaissance in the past two decades.

The new space start-up firms such as Scaled Composites, Virgin Galactic, Sierra Nevada, SpaceDev, and especially SpaceX have clearly shown this to be the case.

Electric Ion Propulsion

As noted in the previous section, electric ion propulsion has been the subject of concentrated research in order to develop a superior form of propulsion to chemical propulsion for some time. NASA has tested the performance of its NEXT xenon thruster for over 5 years and demonstrated an order of magnitude net increase in total thrust over chemical propulsion based on this record-length test period. This is to say that this NEXT thruster would over its lifetime be up to 12 times as efficient as a chemical rocket in terms of total thrust produced.

Thus, there are ever increasing applications for these small thrusters. The first application was for satellite station-keeping and employed thrusters developed in Russia. Today most advanced satellites employ xenon thrusters for orbital maneuvers and station-keeping of spacecraft in their proper location in GEO orbit or for maintaining the correct orbital position in a constellation of satellites (NASA NEXT ion thrusters).

The operation of the thruster involves the inert Xenon gas being squirted into a chamber. A continuous firing electron gun shoots electrons at the xenon atoms and this creates a plasma of negative and positive ions. The positive ions diffuse to the back of the chamber, where high-charged accelerator grids seizes these ions and propel them out creating thrust. The energy to power the electron gun can either come from conventional batteries (with modest lifetime), solar panels, or a radioisotope thermoelectric generator which would be like the nuclear battery, which might be like that employed on the Mars Rover *Curiosity.*

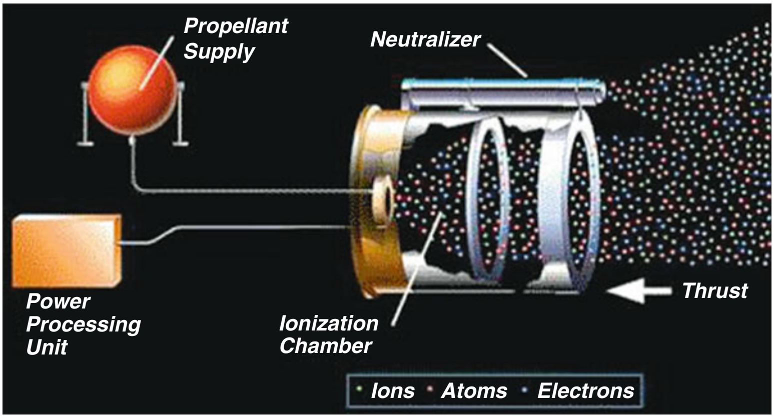


Fig. 2 Xenon ion thruster for sustained propulsion that can maintain continuous thrust for years at a time (Graphic Courtesy of NASA)

These types of thrusters might also be deployed in the future on space debris elements to achieve active deorbit over time. The low thrust levels make it difficult to use these systems for lift off from the ground. In fact, friction on the ground makes ion thrusters on earth impractical. Nevertheless, it might be possible to deploy satellites from dark sky stations, high-altitude balloons, or high-altitude carrier vehicles that could fly quite small, low mass payloads to high enough altitudes that the thrusters could overcome the initial gravitational pull. These possible deployments could only be for quite small payloads (Fig. 2).

Nuclear Ion Propulsion

The idea of using ion thrusters for very sustained propulsion for many years at a time for missions like interplanetary travel or for large-scale constellations that are intended to last for decades at a time might also consider the super heating of the Xenon fuel, not by solar arrays and batteries but by nuclear isotope heating systems. The energy to power the electron gun that creates the plasma ions in the future could come from a radioisotope thermoelectric generator. In this case, it would most likely be a nuclear battery, just like that was used on the Mars Rover *Curiosity* for such extended periods of operation. Radioisotope power sources will not make the ion thrusters more powerful, but simply longer lived.

The clear downside of ion thrusters, though, is that the amount of thrust produced is still quite small. Ion thrusters currently can deliver only about 0.5 N of thrust. This is in contrast to a chemical thruster, such as a hydrazine jet, which can produce hundreds and even thousands of newtons for short bursts of thrust. An ion thruster in the frictionless environment of space would take about an hour to accelerate a spacecraft to about 100 km/h. But over the course of years, a spacecraft could build up a velocity to as high a speed as 500,000 km/h.

The prime future application could thus include interplanetary missions. One might also create a grid of thrusters on future shorter missions to create a larger net thrust over time. Certainly it is a possibility to use ion thrusters to redeploy a spacecraft from low earth orbit to GEO orbit. There is a catch in that this very slow spiral deployment would take weeks and perhaps months to complete. In light of the increasing amount of debris in low earth orbit, this could involve a higher risk element of possible collision as this slow deployment takes place.

High-Altitude Launch Systems

The initial launch part of a launch operation when the rocket must accelerate all of the stored fuel plus the rocket itself is the most difficult. Also the higher the elevation of the rocket the more the pull of the Earth's gravity decreases. As one flies out of the Earth's gravity well, the pull is significantly less. When a spacecraft arrives at GEO orbit, the pull is 1/50th of that experienced as ground level. Further the higher one goes, the thinner the atmosphere, and thus there is lesser drag that slows the rocket's acceleration.

This means that a rocket launched from a balloon, dirigible, or flying rocket launcher has a significant advantage over a static ground launch. Orbital Sciences used a converted B52 Stratofortress aircraft as part of its launch operation for the Pegasus launcher that first flew operationally in 1990 over a quarter of a century ago. Currently, a Lockheed L 1011 Tri-star Stargazer carries the Pegasus underneath for release at 40,000 ft (or about 12 km) of altitude. This has the advantage of no launch site operations costs, the air speed, reduced gravitational pull, and thinner atmosphere so that the Pegasus can launch more payload than if the launch had occurred on the ground.

Today there are many launcher concepts that involve the advantages of not having a ground launch. JP Aerospace envisions future ion-engine launches from dark sky stations that could fly very modest payloads to low earth orbit. Kelly Space and Technologies Corporation envisions the possibility of a rocket that could be towed to high altitude. Planetspace (now defunct) and others have envisioned a launch from a balloon.

Currently there are two projects that have received a great amount of publicity with regard to creating a carrier plane that is optimized for the launch of rockets or spaceplanes at high altitude. Burt Rutan, of Space Composites, received a great deal of publicity when his specially designed White Knight carrier plane twice flew the SpaceShipOne up high into the sky so that this innovative spaceplane could claim the XPrize in 2004. Today the White Knight 2 is the carrier plane that has been developed to support the ongoing launch of SpaceShip2 for space adventure missions for "space tourists" (Fig. 3).

The White Knight 2 is now also designed for the "Launcher One" deployment at 50,000 ft (or 14 km). Thus the Launcher One is designed to carry small spacecraft to low earth orbit. The Launcher One has now been contracted to launch some of the

Fig. 3 White Knight 2 with SpaceShip2 aboard for test flight (Graphic courtesy of Virgin Galactic)

Fig. 4 Conceptual image of Dreamchaser spaceplane with carrier aircraft (Image courtesy of Sierra Nevada)

last spacecraft to orbit for the OneWeb Constellation that is to be deployed in 2017 and 2018.

The inspiration represented by the White Knight 1 and 2 has led to further developments. The Dreamchaser spaceplane by Sierra Nevada is also designed to be launched from a carrier plane (See Fig. 4) (White Knight 2).

Even more ambitious is the idea of creating a "super carrier" that is known as "Stratolaunch". This project began as a joint venture among Paul Allen (of Microsoft and Space Ship 1 fame), Elon Musk (of Space X), and Burt Rutan. At this stage, Elon Musk and Space X have dropped out of the venture. In 2015, Stratolaunch Systems was placed under the supervision of Paul Allen's new company Vulcan Aerospace, a subsidiary of Vulcan, Inc. At one stage, Orbital ATK joined the venture when Space X left, but it is not clear as to where the Antares launcher is a candidate for deployment via Stratolaunch.

The project involves a completely mobile launch system. The stratolaunch has three primary components. These are a huge carrier aircraft, with six 747 aircraft engines and a giant wingspan that is being built by Scaled Composites, a multistage payload "launch vehicle" which would be launched at high altitude into space from under the carrier aircraft, and a mating and integration system that is being developed by Dynetics. As a result of no launch site costs and the advantage of deployment at 14 km, this could be a very highly cost-effective system if this development proves

Fig. 5 The Stratolaunch System when complete will represent the world's largest aircraft (Photo courtesy of Vulcan Aerospace)

out in practice. Currently, the first launch is scheduled for late 2016 or 2017. The ambition is to make the system reliable enough that it might be used to send astronauts to the International Space Station (Zolfagharifard 2015).

The feasibility of high-altitude spacecraft launches has been demonstrated many dozens of times with the Pegasus launch system. Today the initiatives that include the Dreamchaser, SpaceShip2, LauncherOne, and most ambitiously Stratolaunch by Vulcan Aerospace suggest that high-altitude rocket launches have become a significant new focus of launch systems for the future. The next few years should reveal just how effective this new approach has proven to be technically, operationally, and from a financial and business perspective (Fig. 5).

Reusable Launcher Systems

Another key new element of thought with regard to launch systems is the focus on developing reusable launcher systems that can land after accomplishing their mission and thus be reused again. At this point, the two organizations that are giving the most attention to this idea are Jeff Bezo's Blue Origin with its New Shephard launch system and Elon Musk with his Falcon 9 launcher system and smaller suborbital Grasshopper vehicle. John Carmack's Armadillo Aerospace has also worked in this area but he has now left the field and it is not clear whether the follow-on effort will be as productive as before without his financial backing.

The Blue Origin and Space X vehicles represent quite different systems. The "New Shephard" vehicle does not quite look like a rocket with its stubby and flat top appearance. This vehicle is intended only for suborbital flights by space tourists. Its

Fig. 6 Blue Origin's New Shephard suborbital launcher that successfully touched down after launch on November 24, 2015 (Graphic courtesy of Blue Origin)

maximum intended altitude is 100 km (or 62 miles). Since this vehicle flies at much lower speeds and can be designed with a much stubbier and fatter profile, i.e., more like a farm's silo than a pencil, it is easier for it to take off and then land vertically than the Falcon 9 (Grush 2015) (See Fig. 6).

The part of the Falcon 9 that SpaceX is trying to recover doesn't actually reach orbit either. This is because this rocket stage separates from the Dragon capsule that actually goes into orbit. Thus Space X in their recovery efforts are seeking only to land the first stage of the vehicle. This is the critical and most expensive part of the Falcon 9 launcher since it contains the main engines and also houses most of the fuel.

This first stage of the Falcon 9 reaches an elevation of about 200 km (or about 124 miles) or double the height and at least double the velocity and up to 15 times the level of thrust when compared to the New Shepard. In short, it is more technically difficult to make a perfect vertical landing of the Falcon 9 on a floating platform in the ocean than the vertical land of the New Shepard (Grush 2015).

The landing supports that deploy during the landing operation must aerodynamically conform to the body of the launcher and thus this is a quite challenging operation. Space X was able to land its Grasshopper vehicle, but the challenges associated with the much taller Falcon 9 are much more difficult (See Fig. 7).

There is good reason to believe, however, that both recoveries can ultimately be made and that this will serve to reduce the total cost of the launch services provided by either Blue Origin or SpaceX going forward. If indeed both firms are successful, this will likely impact the business models of other launcher operations. In short, more operators will likely seek to recover and reuse their launchers.

Fig. 7 The deployable landing legs for Space X Grasshopper launch vehicle (Graphics courtesy of Space X)

Environmental Issues and Concerns

At the beginning of the space age in the 1950s, there was little concern about environmental issues. The concerns about climate change had really not become wide spread. There were very few launches occurring. If there were concerns about air pollution, there were focused on coal-burning power plants, factories, oil refineries, and aircraft traffic. Today there is wide spread concern about air pollution. The launch vehicles that give off particulate emissions in the very top reaches of the stratosphere are highly polluting. The Space Dev company, founded by Jim Benson in 1997, developed a new and critical hybrid rocket motor technology. What made this technology was the fact that he had developed a "throttle-able" rocket engine design. This design was based on the innovative combination of laughing gas (nitrous oxide as the oxidizer) and neoprene rubber (as the fuel). It was such a hybrid rocket motor that powered the SpaceShipOne when it secured the $10 million Ansari X Prize in 2004. After Benson retired, Space Dev was acquired by Sierra Nevada and the Dream Chaser also uses the hybrid rocket system that is very reliable but spews out particulates that makes this type of rocket system much worse in terms of pollution than liquid-fueled system and especially much worse than liquid hydrogen and liquid oxygen type motors.

Space Ship Two initially had been designed to use neoprene as the fuel. This is specifically known as hydroxyl-terminated polybutadiene (HTPB) but then in 2014 moved to a polyamide that is similar to nylon as the solid fuel since this was thought to give higher performance. Subsequent to the accident on October 31, 2014, the decision has been made to return to the original HTPB solid fuel. The problem is that both the HTPB and polyimide fuel sources spew out particulate particles that create an upper atmosphere pollution issue. This is to say that SpaceShip2 and Dreamchaser both have reliable and throttle-able propulsion systems, but they definitely create stratospheric air pollution. The relatively few flights that are currently flown do not create major issues, but if this fuel source continues to be used for an increasing number of flights and this ultimately is considered for hypersonic transportation in operational spaceplanes, this would be considered a serious pollution problem going forward.

Spaceplanes that fly at supersonic speeds also give rise to noise pollution issues. The sonic boom associated with the take-off and landing of the Corcorde SST led to its being grounded in the USA. NASA and aerospace corporation research as carried out by Lockheed Martin, the QSST (Quiet Super Sonic Transportation) Corporation, and others has developed technology such as extendable needle noses that allows supersonic aircraft to create a series of smaller microsonic boom in place of a single tremendous boom.

The future of spaceplanes, rocketplanes, and hypersonic transport today is still a long ways from being established with a great deal of new technology to be developed. Nevertheless, the research and development agenda should spend efforts to develop cleaner fuel sources and less noisy craft now rather than finding that these issues block the deployment of these systems sometime in the future. Liquid-fueled systems are clearly cleaner and designs that create microsonic booms are clearly superior to hypersonic craft that create gigantic sonic booms over cities where they land (Foust 2015).

Advanced Launch Concepts

The main stream approach to deploying application satellites in earth orbit currently is focused to developing launcher systems that are more reliable and lower in cost. The development of reusable rocket systems and launchers that are modular and manufactured more efficiently using technologies such as 3D printing and are deployed at high altitude off of carrier vehicles are just some of the technologies that seem likely to bear fruit. There are new rocket launchers that have enormous capacities such as Ariane 6, Falcon 9 Heavy, Delta 4 Heavy, and Atlas 5 Heavy that will clearly achieve economies of scale and will accommodate very large spacecraft. At the opposite extreme, there are now small-scale and modular projects such as Firefly, Launcher One, S-3, and other systems that are designed to accommodate small spacecraft. Ion propulsion systems also are evolving rapidly and can support station-keeping and satellite repositioning, but in the future may also assist with orbit raising or even interplanetary missions.

But in parallel to these efforts to develop better rockets, there are ongoing efforts to create entirely new technology that are safer, cleaner, and lower in cost to get significant payloads into space and to return materials and even manufactured materials back to Earth.

There are a wide range of options that have been discussed. These include the use of tethers and a skyhook system that could lift payloads to successively higher orbits that might evolve over time to a true space elevator. There are concepts that involve mass drivers, electromagnetically accelerated "rail guns" that could send payloads to orbit, and even use of nuclear fusion power sources to create launchers for large space infrastructure or even star ships. There are many sources that discuss these advanced technologies (Non-Rocket Space Launch).

Today's launcher technology is quite sufficient to support a wide range of space applications, but the future needs associated with a true off-world economy and the creation of large-scale space infrastructure such as to protect the earth from the most violent solar storms and to combat climate change may require space deployment systems that are ten to one hundred times more cost-efficient than today's launchers (Pelton 2016, Planetary Defense).

New Commercial Launch Sites and Spaceports

Yet another key trend in terms of launch operations is the reduced cost of launch operations. There is a growing number of commercial launch sites and so-called spaceports around the world. These commercial operations tend to efficient and cost-effective places from which to carry out commercial launches. These sites keep multiplying in the USA and around the world. In addition to these land-based launch operations, there are also new options that are developing such as high-altitude carrier planes, and also ocean-based platforms that can launch from the equator in support of GEO orbit spacecraft (See Chapter ▶ "Major Launch Systems Available Globally").

The increase in launch facilities – especially new low-cost commercial spaceports – and the new high-altitude systems carrier planes can certainly help reduce the cost of launch services. As this occurs, however, there needs to be vigilance to ensure that the proliferation of these sites does not give rise to safety issues. The rapid proliferation of commercial spaceports and the need for FAA-AST safety inspectors to certify and recertify new commercial space launch centers could give rise to concerns that safety standards being consistently and rigorously applied.

Conclusion

There are many industries that are largely stable, predictable, and to which technical innovation comes slowly. This is certainly not the case with the development of new satellite systems, new launch vehicle design and capabilities, and launch center operations. Innovations are happening across the board and certainly across all aspects of the launch services industry.

Highlights that are occurring with regard to launch system capabilities, in-orbit station-keeping, as well as launch operations include the following:

- New advanced design, prototyping, manufacturing, 3D printing, and accelerated testing capabilities that have been particularly driven by “new space” launch service providers that include Space X, Sierra Nevada, External Engines, Firefly, and Virgin Galactic Launcher One.
- Advanced electric ion propulsion systems that can provide long duration station-keeping, orbital relocation and even orbit-raising. In the future, advanced ion systems with nuclear-powered electric guns in multiple grids might be used for more rapid orbit raising and might even operate from dark sky stations.
- Development of reusable launch vehicles that could add to reliability and reduce launch costs both for space tourism suborbital flights as well as launch to earth orbit.
- Development of high-altitude carrier vehicles, such as White Knight 2 and Stratolaunch, that can improve performance and reduce the cost of launch services.
- Reduced costs of launch operations due to in-air high-altitude launches and the competitive pressures that come from more commercial launch sites and spaceports.

In addition to these changes driven by the “new space” commercial innovators, new capabilities in launch services have also developed around the world in places such as Japan, China, India, and the Ukraine. Japan, China, and India are now able to send probes to the Moon and Mars. Even countries such as Israel, Iran, North Korea, Australia/New Zealand, and Brazil are developing rocket launcher capabilities.

As chemical launchers are experiencing a renaissance, however, there are continuing efforts to think outside the box to create entirely new launch capabilities using mass drivers, electromagnetic accelerators, rail guns, tether/skyhooks, and even space elevators to find totally new ways to lift payloads and astronauts to orbit. These efforts may eventually lead not only to safer and lower cost systems to deploy satellites of the future but also to clean up space debris and to avoid the stratospheric pollution that chemically fueled rocket launches now entail.

Cross-References

- ▶ The World's Launch Sites
- ▶ Major Launch Systems Available Globally
- ▶ Launch Vehicles and Launch Sites
- ▶ Satellite Orbits for Communications Satellites

References

"Falcon Heavy in Pictures: SpaceX's Huge Private Rocket Galley", Space.com May 22, 2014 http://www.space.com/25963-spacex-falcon-heavy-rocket-images.html. Last accessed 16 Jan 2016

C. Davenport, B. Fung, "Three Firms Win NASA Contract", Washington Post, 15 Jan 2016

J. Foust, "SpaceShipTwo Bounces Back to Rubber Fuel" Space News, October 14, 2015 http://spacenews.com/virgin-galactic-switching-back-to-rubber-fuel-for-spaceshiptwo/#sthash.Ft5zPdWo.dpuf. Last accessed 16 Jan 2016

L. Grush, Why you shouldn't compare Blue Origin's rocket landing to SpaceX, The Verge.com November 24, 2015 http://www.theverge.com/2015/11/24/9793220/blue-origin-vs-spacex-rocket-landing-jeff-bezos-elon-musk. Last accessed 16 Jan 2016

"NASA NEXT ion drive breaks world record" http://www.extremetech.com/extreme/144296-nasas-next-ion-drive-breaks-world-record-will-eventually-power-interplanetary-missions. Last accessed 16 Jan 2016

Non-Rocket Space Launch https://en.wikipedia.org/wiki/Non-rocket_spacelaunch. Last Accessed 16 Jan 2016

J.N. Pelton, "Planetary Defense: The Time Has Come" *Space Safety Magazine*, Jan 2016

N. Redd, "NASA's Innovative Ion Space Thruster Sets Endurance World Record" Space.com, September 24, 2013 http://www.space.com/22916-nasa-ion-thruster-world-record-test.html. Last accessed 16 Jan 2016

White Knight 2, http://www.scaled.com/projects/model_348_whiteknighttwo. Last accessed 16 Jan 2016

E. Zolfagharifard, Paul Allen launches 'Vulcan Aerospace': Daily Mail, (17 Apr 2015). London

Part VIII

Hazards to the Future Space Applications

Orbital Debris and Sustainability of Space Operations

Heiner Klinkrad

Contents

Abstract

The orbital particle environment around the Earth is dominated by man-made space objects, except for a limited particle size regime below 1 mm, where meteoroids provide a significant contribution, or may even prevail in some orbit regions. The mass of man-made objects in Earth orbits is on the order of 6,800 t, of which more than 99 % is concentrated in trackable, cataloged objects larger than typically 10 cm. The mass of meteoroids within the regime of Earth orbits is only on the order of 2–3 t, with most probable sizes around 200 μm. As a consequence of their size spectrum and associated mass, man-made space objects, in contrast with meteoroids, represent a considerable risk potential for space assets in Earth orbits. To assess related risk levels, a good understanding of the space debris environment is essential, both at catalog sizes and subcatalog sizes. The derivation process and the key elements of today's debris environment models will be outlined, and results in terms of spatial densities and impact flux levels will be sketched for those orbit regions that are most relevant for space applications.

H. Klinkrad (✉)
Institute of Space Systems, Braunschweig University of Technology (TU Braunschweig), Braunschweig, Germany
e-mail: H.Klinkrad@tu-braunschweig.de

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_77

To cope with the existing space debris environment, spacecraft can actively mitigate the risk of collisions with large-size, trackable space objects through evasive maneuvers. Alternatively, or in addition, the risk of mission-critical impacts by nontrackable objects can be reduced through shielding, in combination with protective arrangements of critical spacecraft subsystems. With a view on the future debris environment, international consensus has been reached on a core set of space debris mitigation measures. These measures, which will be explained in more detail hereafter, are suited to reduce the debris growth rate. However, even if they are rigorously applied, they are found to be inadequate to stabilize the debris environment. Long-term debris environment projections indicate that even a complete halt of launch activities cannot prevent the onset of a collisional runaway situation in some LEO altitude regimes. The only way of controlling this progressive increase of catastrophic collisions is through space debris environment remediation, with active mass removal, focused on retired spacecraft and spent orbital stages.

Keywords

Collision avoidance • Debris collision flux • Collision risk assessment • Debris environment models • Debris environment projections • Debris environment remediation • Debris mitigation • Evasive maneuvers • Impact protection • Inter-Agency Space Debris Coordination Committee (IADC) • Orbital debris • Space debris • Sustainability of space activities • US Space Surveillance Network (SSN)

Introduction

More than half a century of space flight activities since the launch of Sputnik-1, in 1957, has generated a significant man-made particle environment in Earth orbits that is referred to as “space debris.” According to a definition by the Inter-Agency Space Debris Coordination Committee (IADC), “space debris are all man-made objects including fragments and elements thereof, in Earth orbit or re-entering the atmosphere, which are non-functional.” This sizeable population of space debris must be considered in the payload and mission designs to ensure successful space operations with an acceptable, low risk of losing or degrading a mission, or of suffering casualties during human space flight. Likewise, payloads and orbital stages must be designed, operated, and disposed of such that they do not further deteriorate the space debris environment, or pose an unacceptable risk to the ground population or to air traffic during re-entries.

Throughout this chapter, a snapshot of the orbital population of space objects close to 2015 (±1 year) will serve as a reference. The orbital debris environment in January 2016 was the product of more than 5,150 launches and more than 250 on-orbit breakups that led to more than 17,700 objects which are accessible through the unclassified catalog of the US Space Surveillance Network (SSN)

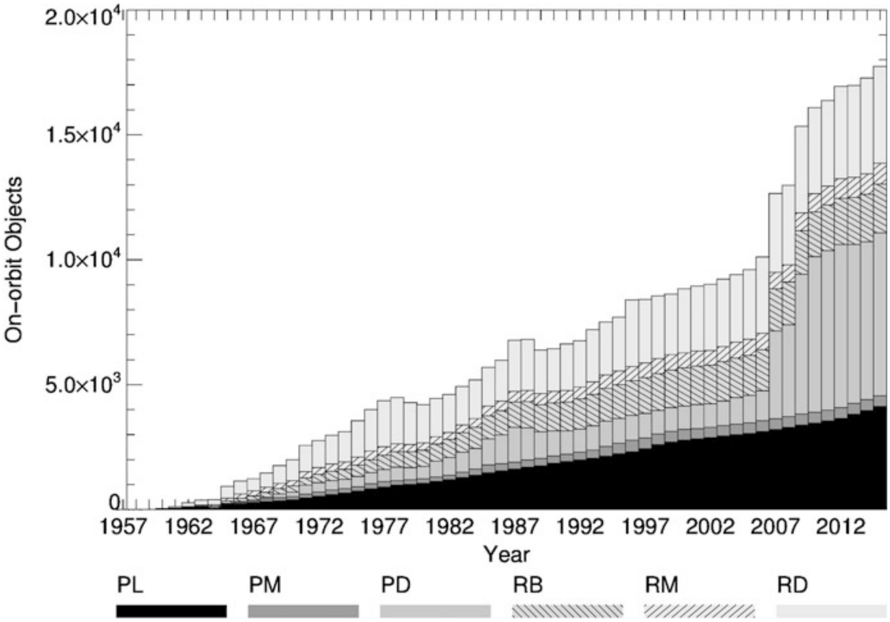


Fig. 1 Historic evolution of the US SSN catalog of trackable space objects through December 2015 (*PL* payloads, *PM* payload mission-related objects, *PD* payload debris, *RB* rocket bodies, *RM* rocket body mission-related objects, *RD* rocket body debris; Credit: ESA)

(see Fig. 1 and Table 1). Another ~6,000 objects were systematically tracked but were either classified, or they were not yet correlated with a launch or deployment event. All SSN catalog objects combined represent some 6,800 t of on-orbit mass. Several 10 t of further material from different sources are expected to exist at subcatalog sizes, below diameters of 10 cm. Only 6 % of the catalog entries are operational spacecraft (slightly more than 1,000), while 28 % are nonfunctional but intact objects, and 58 % are fragments, mainly resulting from explosions but also from recent in-orbit collisions. 71 % of the catalog objects are in low Earth orbits (LEO), 7 % are in or near geostationary orbits (GEO), and 22 % are in highly eccentric orbits (HEO), medium Earth orbits (MEO), or other orbit classes. Since 2007, the SSN catalog has experienced two significant step increases: on January 11, 2007, the Chinese Feng Yun 1C satellite was intercepted in an ASAT (Anti-Satellite) test, generating 3,428 catalog objects, of which 2,932 were still in orbit 9 years later, and on February 10, 2009, the first accidental hypervelocity collision between two intact catalog objects (Iridium 33 and Cosmos 2251) generated 2,296 cataloged fragments in two separate clouds, of which 1,555 were still in orbit 7 years later. Both of these events have produced a long-lasting increase in spatial object densities, and hence in collision risk, at altitudes between 750 and 900 km.

The risk of collision-induced catastrophic fragmentations or mission-terminating impacts is the highest in the low Earth orbit (LEO) regime. It exceeds the risks in

Table 1 Orbital distribution of US Space Surveillance Network catalog objects in January 2016 (*MRO* mission-related objects, *LEO* low earth orbits, *GEO* near-geostationary orbits, *NSO* navigation satellite orbits, *LMO* orbit passing through LEO and MEO, *GTO* GEO transfer orbits, *MGO* orbit passing through MEO and GEO, *MEO* medium earth orbits, *HEO* highly eccentric orbits; Credit: ESA)

	Payloads			Rocket bodies					
Orbit	Intact	MRO	Debris	Intact	MRO	Debris	unkn.	Total	
LEO	2,224	122	6,196	834	501	2,641	69	12,587	70.9 %
GEO	1,022	35	7	197	0	28	1	1,291	7.2 %
NSO	219	1	1	65	2	0	0	288	1.6 %
LMO	89	49	135	208	217	743	1	1,442	8.2 %
GTO	76	10	15	222	56	251	3	633	3.6 %
MGO	63	3	145	158	4	14	1	388	2.2 %
MEO	48	52	14	15	8	12	0	149	0.8 %
HEO	47	3	0	44	3	75	1	173	1.0 %
Others	331	156	39	198	36	38	6	803	4.5 %
Total	4,119	431	6,552	1,941	827	3,802	82	17,754	100 %
	23.2 %	2.4 %	36.9 %	10.9 %	4.7 %	21.4 %	0.5 %	100 %	

other orbit regions, including the geostationary orbit (GEO) by at least 3 orders of magnitude. As a consequence, the following analysis will concentrate on the collision risk levels for the International Space Station (ISS), as an example of a manned LEO platform, and on the collision risk levels for a typical remote sensing spacecraft, on a sun-synchronous orbit, as an example of a robotic LEO platform. The concepts of active protection (shielding) and passive protection measures (avoidance maneuvers), and their effectiveness as a function of debris size will be discussed as possible risk mitigation measures for the specific debris environment of given operational orbits at 360 km altitude and 51.1° inclination for the ISS, and at 780 km altitude and 98.5° inclination for an Earth observation mission.

Roughly 35 % of the entire mass in orbit is concentrated in the LEO regime, within just 0.3 % of the operationally used volume from LEO up to super-GEO altitudes. Debris risk mitigation through collision avoidance, passive protection, and end-of-mission disposal turns out to be a necessary but insufficient condition to maintain an acceptable space debris environment. Long-term projections indicate that even drastic mitigation measures, such as an immediate, complete halt of launch and release activities will not result in a stable LEO debris environment (see Liou and Johnson 2008a, b; Bastida and Krag 2009; Liou 2011; Klinkrad and Johnson 2009, 2013). Catastrophic collisions between existing space hardware of sufficient size will within a few decades start to dominate the debris population sources and lead to a net increase of the space debris population, also at sizes which may cause further catastrophic collisions. A self-contained collisional cascading process in the LEO regime may hence ultimately lead to a runaway situation (the so-called Kessler syndrome), with no further possibility of control through human intervention. The only way to prevent the on-set of collisional cascading is an active removal of mass from orbit. Since most of the LEO mass is concentrated in decommissioned though intact satellites and orbital stages, an effective mass removal operation must focus on this class of objects and on preferred orbit classes for their mission deployments. Several operational concepts and physical principles have been explored to enable a space debris environment remediation through mass removal. Some of the most promising of these concepts suggest the use of electrodynamic or momentum-exchange tethers, space tugs, the deployment of drag augmentation devices or solar sails, or the release of large momentum-retarding surfaces. Such options will be reviewed in the following.

Apart from the systematically trackable catalog population of space objects, there is a much larger population of subcatalog debris objects than can disable or seriously degrade a space mission. The related objects can only be observed in a statistical manner, by means of research radars, telescopes, and in situ detectors. Based on orbital and physical characteristics of the observed debris, and based on ground test benchmark data, debris environment models can be established that compose an image of the current environment from a replicate of historic launch, release, and breakup events. One of the leading debris models, ESA's MASTER software (Meteoroid and Space Debris Terrestrial Environment Reference; see Oswald et al. 2005; Flegel 2010), will be used in the following risk assessments. An in-depth technical discussion of underlying theories and analysis techniques is provided in Klinkrad (2006) and will not be repeated here.

Space Debris and Their Effect on Space Applications

The resident mass in operationally used orbit regions around the Earth is to 99.95 % dominated by man-made space debris, totaling approximately 6,800 metric tons in the year 2016. Only a few tons of additional material within the same reference volume originates from natural meteorites, with most probable sizes of about 200 μm. As a consequence, space debris dominate the risk for operational space missions and will be the focus of the following discussion.

Within one decade after the first space launch, the annual launch rates reached a level of 120 at the end of the 1960s and a peak of almost 130 by the mid 1980s. As a consequence of reduced Russian space activities at the end of the 1980s, annual launch rates dropped to about 50 by 2005. By 2015, they reached again a level above 80. By January 2016, there were some 5,166 successful launches (out of 5,521 launch attempts) that deployed 4,119 payloads, 1,941 rocket stages, and 1,253 mission-related objects (MRO) into orbit (see Table 2). These intact objects account for most of the in-orbit mass of about 6,800 t. However, they only account for 41.2 % of the space object population that can be routinely tracked by operational surveillance networks. Out of 17,754 objects of the US Space Surveillance Network (SSN) catalog in January 2016, the dominant space debris population contributed 10,354 trackable objects (58.8 %). With 12,587 objects (70.9 %), the vast majority of the SSN catalog resides in low Earth orbits (LEO), below altitudes of 2,000 km, another 1,291 objects (7.2 %) are in the vicinity of the geostationary ring (GEO), mainly at altitudes of 35,786 ± 2,800 km and inclinations of $0° \leq i \leq 15°$, and the remaining 3,876 objects (21.9 %) are distributed across medium Earth orbits (MEO), semisynchronous orbits of navigation constellations (NSO), GEO transfer orbits (GTO), highly eccentric orbits (HEO), orbits that pass through LEO and MEO (LMO), orbits that pass through MEO and GEO (MGO), and high-altitude orbits beyond the GEO regime. Table 1 shows the catalog composition; Table 2 shows individual contributions to the SSN catalog according to launch nation; and Fig. 1 shows the historic evolution of the catalog population.

The US Space Surveillance Network has a cataloging size threshold that ranges from about 10 cm in the LEO regime to about 1 m in the GEO ring. Related routine observations are performed by a network of radars for LEO and low MEO altitudes and by globally distributed electro-optical telescopes for the remaining part of MEO up to GEO altitudes. For the dominant LEO catalog population, Fig. 2 shows the altitude distribution of objects, with a main maximum close to 800 km and a secondary maximum slightly below 1,500 km. Since the vast majority of catalog objects are on near-circular orbits (with more than 50 % of the eccentricities smaller than 0.01), the depicted, resident probability weighted, mean altitude distribution is very similar to the actual perigee and apogee altitude distributions. Figure 3 shows that the inclination distribution of LEO orbits is driven by mission and launch

Table 2 Status of the US Space Surveillance Network catalog in 2015 according to the NASA Satellite Situation Report (Credit: JSpOC)

Launch nation/ organization	Code	Objects in orbit			Objects decayed		
		PL/RB	Debris	Total	PL/RB	Debris	Total
Arab Sat. Com. Org.	AB	13	0	13	1	0	1
Asiasat Corp.	AC	7	0	7	0	0	0
Algeria	ALG	2	0	2	0	0	0
Argentina	ARGN	14	0	14	2	0	2
Austria	ASRA	2	0	2	0	0	0
Australia	AUS	16	0	16	0	0	0
Azerbaijan	AZER	1	0	1	0	0	0
Belgium	BEL	2	0	2	0	0	0
Belarus	BELA	2	0	2	0	0	0
Bolivia	BOL	1	0	1	0	0	0
Brazil	BRAZ	16	0	16	1	0	1
Canada	CA	43	5	48	1	2	3
Chile	CHLE	2	0	2	0	0	0
China (P.R.)/Brazil	CHBZ	3	57	60	0	30	30
China (P.R.)	PRC	311	3,489	3,803	187	1,001	4,988
China (Rep.)	ROC	9	0	9	0	0	0
CIS (Russia)	CIS	2,531	3,792	6,323	4,812	10,169	14,981
Colombia	COL	1	0	1	0	0	0
Czechoslovakia	CZCH	4	0	4	2	0	2
Denmark	DEN	9	0	9	0	0	0
Equador	ECU	2	0	2	0	0	0
Egypt	EGYP	5	0	5	0	0	0
ESA	ESA	73	47	118	20	19	39
ESRO	ESRO	0	0	0	7	3	10
Estonia	EST	1	0	1	0	0	0
Eumetsat	EUME	8	8	16	0	0	0
Eutelsat	EUTE	50	0	50	0	0	0
France	FR	202	321	523	82	631	713
France/Germany	FGER	2	0	2	0	0	0
France/Italy	FRIT	2	0	2	0	0	0
Germany	GER	49	1	50	15	1	16
Globalstar	GLOB	84	1	85	0	1	1
Greece	GREC	2	0	2	0	0	0
Hungary	HUN	0	0	0	1	0	1
India	IND	92	84	176	23	301	324
Indonesia	INDO	13	0	13	1	0	1
Inmarsat	IM	16	0	16	0	0	0
Intelsat	ITSO	82	0	82	1	0	1
Iran	IRAN	1	0	1	8	0	8
Iraq	IRAQ	1	0	1	0	0	0
Israel	ISRA	14	0	14	10	0	10

(*continued*)

Table 2 (continued)

Launch nation/ organization		Objects in orbit			Objects decayed		
	Code	PL/RB	Debris	Total	PL/RB	Debris	Total
ISS (Space Station)	ISS	5	0	5	1	89	90
Italy	IT	24	0	24	11	1	12
Japan	JPN	195	34	229	113	163	276
Kazakhstan	KAZ	5	0	5	0	0	0
Laos	LAOS	1	0	1	0	0	0
Lithuania	LTU	0	0	0	2	0	2
Luxemburg	LUXE	2	0	2	0	0	0
Malaysia	MALA	7	0	7	0	0	0
Mexico	MEX	9	0	9	0	0	0
NATO	NATO	8	0	8	0	0	0
Netherlands	NETH	5	0	5	1	0	1
New ICO	NICO	1	0	1	0	0	0
Nigeria	NIG	5	0	5	0	0	0
North Korea	NKOR	2	2	4	0	0	0
Norway	NOR	9	0	9	0	0	0
O3B Networks	O3B	12	0	12	0	0	0
Orb. Telecom Sat.	ORB	41	0	41	0	0	0
Pakistan	PAKI	4	0	4	1	0	1
Peru	PER	1	0	1	2	0	2
Philippines	RP	1	0	1	0	0	0
Poland	POL	2	0	2	1	0	1
Portugal	POR	1	0	1	0	0	0
Reg. African SatCom	RASC	2	0	2	0	0	0
Romania	ROM	0	0	0	1	0	1
Saudi Arabia	SAUD	13	0	13	0	0	0
Singapore	SING	10	0	10	0	0	0
Singapore/Taiwan	STCT	2	0	2	0	0	0
Soc. Europ. de Sat.	SES	54	0	54	1	0	1
South Africa	SAFR	4	0	4	0	0	0
South Korea	SKOR	18	0	18	1	0	1
Spain	SPN	18	0	18	2	0	2
Sweden	SWED	11	0	11	0	0	0
Switzerland	SWTZ	2	0	2	0	0	0
Thailand	TH	8	0	8	0	0	0
Turkmen./Monaco	TMMC	1	0	1	0	0	0
Turkey	TURK	11	0	11	0	0	0
Ukraine	UKR	1	0	1	0	0	0
Uruguay	URY	1	0	1	0	0	0
United Emirates	UAE	7	0	7	0	0	0
United Kingdom	UK	41	0	41	9	4	13
Unitied States	US	1,944	3,417	5,361	1,575	4,444	6,019
United States/Brazil	USBZ	1	0	1	0	0	0

(*continued*)

Table 2 (continued)

Launch nation/ organization	Code	Objects in orbit			Objects decayed		
		PL/RB	Debris	Total	PL/RB	Debris	Total
Venezuela	VENZ	2	0	2	0	0	0
Vietnam	VTNM	3	0	3	2	0	2
Column totals		6,211	11,263	17,475	6,902	16,859	23,761
Overall total	41,236			17,475			23,761

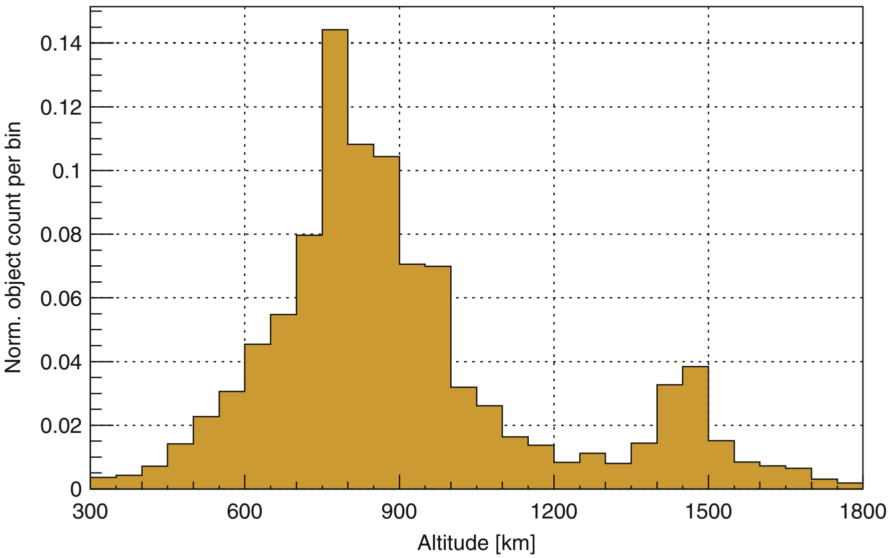


Fig. 2 Altitude distribution of catalog-size objects (>10 cm) in low earth orbit (LEO) in 2010. The normalized count is in fractions per 50 km altitude bin for a total of 11,581 objects (Klinkrad and Johnson 2013)

constraints, with distinct, preferred inclination bands around 65°, 75°, 82°, 90°, and 98°. Figure 4 illustrates how the altitude and inclination distributions of catalog objects are correlated.

Space debris caused by fragmentation events are the most important source of catalog objects, with a contribution of 58.3 % to the trackable population in January 2016. In the course of space history more than 250 on-orbit fragmentation events were inferred from the detection of new objects and from the correlation of their determined orbits with a common source. The dominant breakup causes are believed to have been deliberate explosions or collisions (dominated by an ASAT test that destroyed Feng Yun 1C in January 2007), propulsion-related explosions, battery explosions, and four known accidental collisions (the Cosmos 1934 spacecraft with a

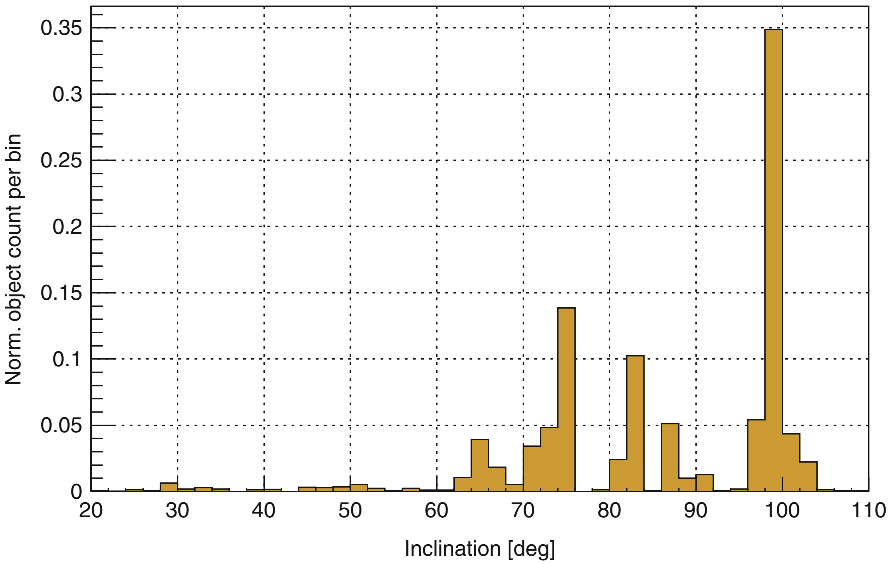


Fig. 3 Inclination distribution of catalog-size objects (>10 cm) in low earth orbit (LEO) in 2010. The normalized count is in fractions per 2° orbit inclination bin for a total of 11,581 objects (Klinkrad and Johnson 2013)

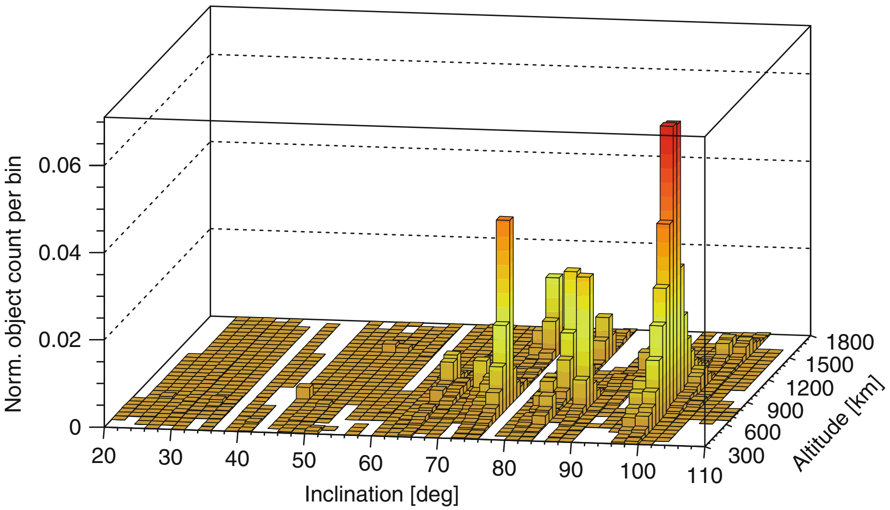


Fig. 4 Inclination and altitude distribution of catalog-size objects (>10 cm) in low earth orbit (LEO) in 2010. The normalized count is in fractions per bin of 2° × 50 km for a total of 11,581 objects (Klinkrad and Johnson 2013)

Cosmos 926 MRO in December 1991, the Cerise spacecraft with an Ariane H-10 fragment in July 1996, a Thor stage with a CZ-4B stage fragment in January 2005, and Cosmos 2251 with Iridium 33 in February 2009). About a third of all breakups were of an unknown cause. With the exception of three known GEO explosion events (an Ekran-2 satellite on June 22, 1978, a Titan III-C Transtage on February 8, 1994, and a Breeze-M stage on January 20, 2016), all known fragmentations occurred on orbits passing through LEO altitudes, with about 74 % of the orbits entirely within LEO and with 15 % on highly eccentric trajectories passing through LEO (Klinkrad 2006). Table 3 shows a list of the 10 most significant in-orbit breakups, sorted by the number of cataloged fragments. Nine of these top 10 events occurred on orbit inclinations of $90° \pm 10°$, mainly at altitudes of 800 ± 50 km. Since the orbit inclination is a very stable parameter, directly linked to the orbit momentum and only marginally affected by orbit perturbations, it strongly governs the latitude distribution of resulting spatial object densities. Figure 5 indicates that the highest concentration of catalog-size objects is at high latitudes δ, where $\delta \approx i$, with i being the inclinations of breakup orbits. As a consequence, catastrophic collisions between catalog objects are most likely at high latitudes in densely populated altitude bands. Fragmentation debris from in-orbit explosions and collisions dominate the space debris population down to the cm-size regime (see Table 4). The most significant breakup-related relative increase of the catalog population occurred in 1961, when the first accidental explosion in space of an Ablestar injection stage more than tripled the catalog population from 110 to almost 400. The most significant absolute growth of the catalog so far occurred in January 2007, when the Feng Yun 1C kinetic ASAT test produced some 3,400 trackable fragments (+33 %), and in February 2009, when the accidental collision between Cosmos 2251 and Iridium 33 generated another 2,300 fragments (+17 %).

At subcatalog sizes, residues from solid rocket motor (SRM) firings become important. The number of solid rocket motor firings up to 2015 was on the order of 1,200, with peak rates of up to 47 events per year, and a mean annual rate of 23.5. The injection orbits where SRMs were applied are up to 80 % associated with US missions. The size of the solid motors, in terms of propellant capacity, covers a wide range. The most frequently used SRMs are the Star 37 motors, with a propellant mass of 1,067 kg, used for instance as final stage of Delta launchers to deploy GPS/Navstar payloads, the Payload Assist Module PAM-D, with 2,011 kg, also used as Delta final stage for instance for GTO injections, and the Inert Upper Stage (IUS), deployed from Titan IV or Space Shuttle, for instance to inject payloads into GTO with a first stage of 9,709 kg, and subsequently deliver the payload into a circular GEO by a second stage of 2,722 kg propellant. Another powerful SRM engine, HS-601 with 4,267 kg, is used by Long March LM-2E launchers both for LEO and GTO payload injections.

SRM combustion residues are mainly composed of aluminum oxide and residues of motor liner material. Aluminum powder is added to most solid fuels, typically with a mass fraction of 18 %, to stabilize the combustion process and improve the motor performance. It is assumed that about 99 % thereof is continuously ejected with the exhaust stream during the main thrust phase in the form of Al_2O_3 dust of

Table 3 Leading ten on-orbit breakup events, sorted by highest counts of on-orbit fragments in January 2016 (Credit: JSpOC and ESA)

	Launch date	Max. count	COSPAR		SSN sat.no.	Assessed cause
Object name	Event date	Curr. count	H_p [km]	H_a [km]	i [deg]	Object type
Feng Yun 1C	1999/05/10	3,428	1999-025A		25730	Deliberate
	2007/01/11	2,932	843	863	98.64	Payload
Cosmos 2251	1993/06/16	1,668	1993-036A		22675	Collision
	2009/02/10	1,173	843	863	98.64	Payload
Iridium 33	1997/09/14	628	1997-051C		24946	Collision
	2009/02/10	382	776	791	86.39	Payload
Cosmos 1275	1981/06/04	346	1981-053A		12504	Battery
	1981/07/24	289	960	1,014	82.96	Payload
Thorad Agena D 2nd stage	1970/04/08	376	1970-025C		4367	Unknown
	1970/10/17	238	1,063	1,087	99.80	Rocket body
Zi Yuan 1 (CBERS 1)	1999/10/14	431	1999-057A		25940	Unknown
	2007/02/18	213	772	782	98.22	Payload
CZ 4B 3rd stage	1999/10/14	431	1999-057C		25942	Unknown
	2000/03/11	213	727	745	98.54	Rocket body
Zenit-2 second stage	1992/12/25	279	1992-093B		22285	Propulsion
	1992/12/26	200	845	848	71.02	Rocket body
Delta 2910 2nd stage	1975/06/12	274	1975-052B		7946	Propulsion
	1991/05/01	199	550	674	97.90	Rocket body
Delta 300 2nd stage	1973/11/06	201	1973-086B		6921	Propulsion
	1973/12/28	179	1,502	1,511	102.05	Rocket body

diameters largely within 1 μm $\leq$ d $\leq$50 μm. Due to design constraints, many solid motors have nozzles protruding into the burn chamber, causing cavities around the nozzle throats. During the burn phase, trapped Al_2O_3, molten aluminum droplets, and parts of released thermal insulation liner material can cumulate in this pool and form slag particles which can grow to sizes of typically 0.1 mm $\leq$ d $\leq$ 30 mm. These slag particles are released at the end of the main thrust phase, as the internal motor pressure decreases. It can be assumed that during more than 1,200 SRM firings, more than 1,000 t of propellant were released into space of which approximately 320 t were Al_2O_3 dust particles and 4 t were slag particles formed of Al_2O_3, metallic aluminum, and motor liner material. Due to orbital perturbations and their different effects on μm-size dust and cm-size slag, merely 1 t of Al_2O_3 dust and 3 t of SRM slag particles are believed to be still on orbit. Apart from more than 1,000 orbit insertion burns, there were also several hundred SRM burns to deorbit objects in a

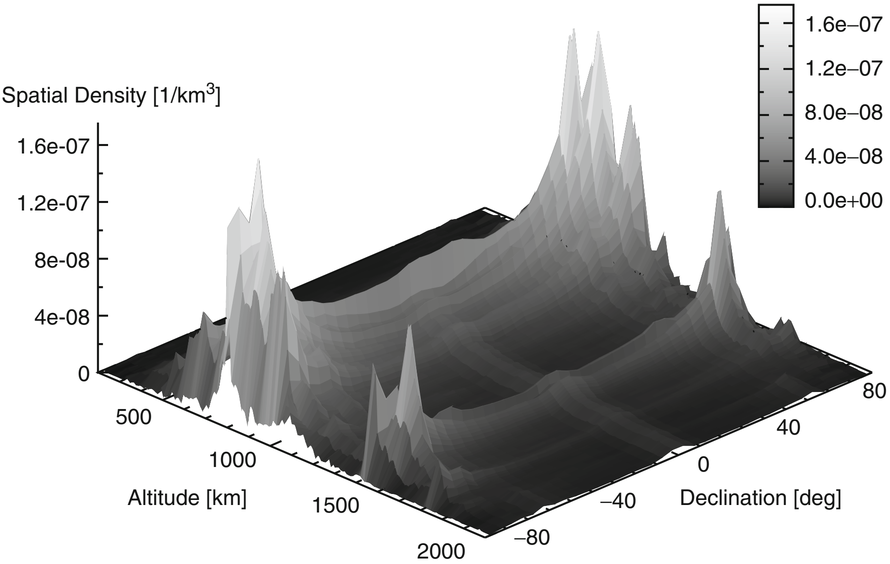


Fig. 5 Spatial density distribution of catalog-size objects (>10 cm) in low earth orbit (LEO) in 2010, as a function of altitude and declination (Klinkrad and Johnson 2013)

controlled fashion. These deorbit burns were almost exclusively performed for Russian reconnaissance satellites at very low altitudes, and the resulting SRM combustion products had a correspondingly low orbit lifetime. However, some in situ measurements (mainly from returned space hardware) show temporal increases in small-particle impact rates due to these events. At sizes of $1\ \mu m \leq d \leq 1$ cm, SRM combustion residues dominate the space debris environment (see Table 4).

Apart from intact objects, fragmentation debris, and SRM residues, there are other contributors to the space debris population: (1) sodium-potassium (NaK) coolant released from 16 Russian RORSATs as they ejected their reactor cores in the 1980s, (2) multilayer insulation (MLI) material that is unintentionally released by spacecraft or rocket stages, (3) ejecta material that is released by small-particle impacts on surfaces of spacecraft and orbital stages, and (4) degradation products that are released by aging surfaces of spacecraft and orbital stages. The debris mass contribution from these sources is much less than 1 % of the overall on-orbit mass, and they are either too small in numbers (NaK, MLI) or too small in size (surface ejecta and degradation products) to constitute a significant risk for space missions.

The population of trackable and nontrackable objects can be reproduced by space debris environment models, such as ESA's MASTER-2009 model (Flegel 2010). Such models consider historic launch and release events, known in-orbit fragmentations, known solid rocket motor firing events, intentional releases of NaK coolant liquid from Buk reactors of Russian RORSAT satellites, unintentional releases of surface degradation products (MLI and paint flakes), and the generation of ejecta and spall by surface impacts. Table 4 lists the resulting debris sources and their

Table 4 Sources and their contributions to ESA's MASTER 2009 space debris model in different size regimes for epoch May 1, 2009

Diameter	>1 μm	>10 μm	>100 μm	>1 mm	>1 cm	>10 cm	>1 m
LMRO	45,919	45,919	45,919	31,138	5,827	5,814	4,174
Expl.	5.64e + 9	4.12e + 9	3.84e + 8	1.53e + 7	433,466	14,719	432
Coll.	3.58e + 9	1.13e + 9	1.18e + 8	4.46e + 6	92,677	2,927	63
MLI	22,241	22,241	22,241	22,241	15,790	5,750	773
NaK	30,162	30,162	30,162	30,162	18,410	–	–
SRM slag	4.98e + 12	4.98e + 12	2.33e + 12	1.39e + 8	177,914	–	–
SRM dust	6.07e + 14	1.18e + 13	–	–	–	–	–
Paint	1.97e + 12	1.62e + 12	2.28e + 11	–	–	–	–
Ejecta	8.61e + 13	2.70e + 13	1.08e + 12	8.00e + 6	–	–	–
Total	6.99e + 14	4.53e + 13	3.64e + 12	1.66e + 8	744,084	29,210	5,442

contributions to the MASTER-2009 population at the reference epoch of May 2009 for the applicable size regime larger than 1 μm. From the risk point of view, the almost 170 million particles larger than 1 mm, at typical LEO collision velocities of 10–14 km/s, can disable sensitive satellite subsystems, the more than 740,000 particles larger than 1 cm can render a spacecraft dysfunctional, and the almost 30,000 objects larger than 10 cm are likely to cause a catastrophic breakup of a satellite or orbital stage.

Figure 6 shows the altitude distribution of MASTER-2009 objects larger than 10 cm in terms of resulting spatial densities (in objects/km^3). The contributing debris sources at these sizes are explosion and collision fragments, intact objects, and lightweight sheets of MLI. Highest concentrations are in the LEO regime, between 750 and 900 km, with almost equal contributions from explosion fragments, collision fragments, and intact objects. In general, however, explosion fragments dominate the LEO and GEO regions, with GEO object concentrations about three orders of magnitude below the LEO maximum. When going to a 1 cm size threshold, additional source terms come in, including NaK droplets and solid rocket motor slag, while launch and mission-related objects start playing a minor role. Figure 7 shows the individual contributions as a function of altitude. Reducing the size threshold further to 1 mm leads to the addition of ejecta particles, as shown in Fig. 8. With the decrease of the debris sizes from 10 cm to 1 mm, the enveloping curve of spatial densities tends to flatten, due to an increasing share of particles on eccentric orbits with a wider distribution over altitudes. As a consequence, the relative magnitude of the GEO concentration peak with respect to the LEO maximum reduces from 3 to less than 2 magnitudes. One cause of the increase of orbit eccentricities with decreasing object sizes lies in the area-to-mass ratio that drives solar radiation pressure and airdrag forces and is inversely proportional to the object diameter.

Spatial object densities are an essential input to debris collision risk assessments. The statistical behavior of the orbital debris population can be well represented by the laws of kinetic gas theory. Hence, the number of collisions c encountered by an object of collision cross section A_c, moving through a stationary debris medium of uniform particle density D, at a constant relative velocity Δv, during a propagation time interval Δt is given by

$$c = \Delta v D A_c \Delta t \tag{1}$$

where $F = \Delta v D$ is the impact flux (in units of $\mathrm{m}^{-2}\mathrm{s}^{-1}$) and $\Phi = F\,\Delta t$ is the impact fluence (in units of m^{-2}). The collision probability follows a binomial law which can be well approximated by a Poisson distribution, generating the following probability $P_{i=n}$ of n impacts, and $P_{i=0}$ of no impact.

$$P_{i=n} = \frac{c^n}{n!}\exp(-c) \qquad \mapsto \qquad P_{i=0} = \exp(-c) \tag{2}$$

The probability of one or more impacts is hence the complement of no impact, given by

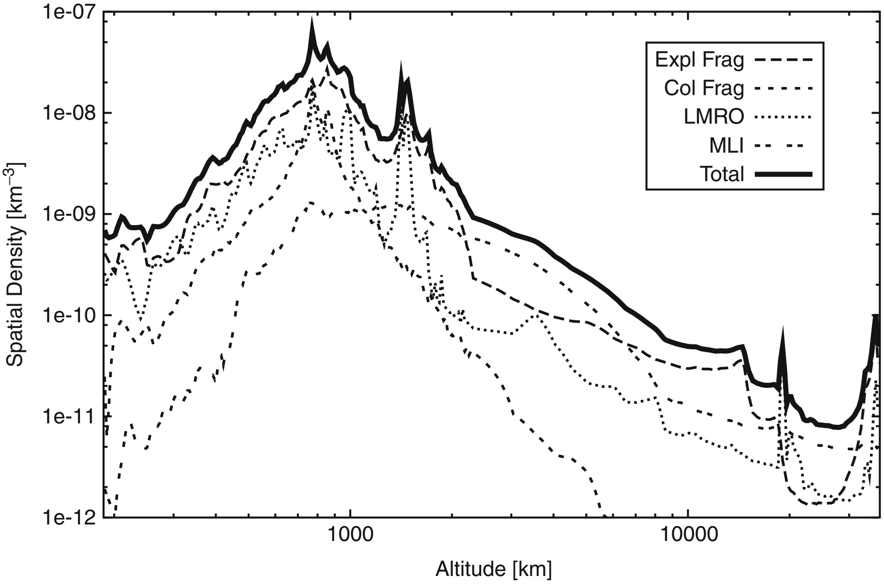


Fig. 6 Spatial density distribution of MASTER-2009 objects of d > 10 cm, in LEO to GEO altitudes, discriminated by sources

$$P_{i \geq n} = P = \quad 1 - \exp(-c) \ \approx \ c \qquad \mapsto \qquad P \approx \ \Delta v D A_c \Delta t \tag{3}$$

The challenging part in the evaluation of this equation is the particle flux $F = \Delta v D$. In the MASTER-2009 model, three-dimensional, time dependent spatial object density distributions are established for a grid of spherical volume elements covering the entire Earth environment from LEO to GEO altitudes. Contributions from each member of the orbital debris population go into this distribution. For each of these objects, the velocity magnitude and direction is retained for each volume element passage. This information is later retrieved to determine relative impact velocities with respect to a target object passing through individual cells of the volume grid (Klinkrad 2006). The resulting impact flux is then determined from a summation across all volume cells that are passed by the target object, with contributions from all debris objects that passed the individual cells.

When considering relative velocities between two objects on circular orbits at the same altitude, with the same orbital velocities v but on different inclinations, Eq. 4 yields the resulting collision velocity as a function of the impact azimuth A within the local horizontal plane (where $A = 0°$ denotes impacts from the flight direction).

$$\Delta v \approx 2v \cos(A) \tag{4}$$

Since near-circular orbits are dominant for debris of critical sizes, Eq. 4 provides a good approximation of the correlation of impact velocity with impact geometry. It also states that the maximum relative velocity can be twice the orbit velocity, for an

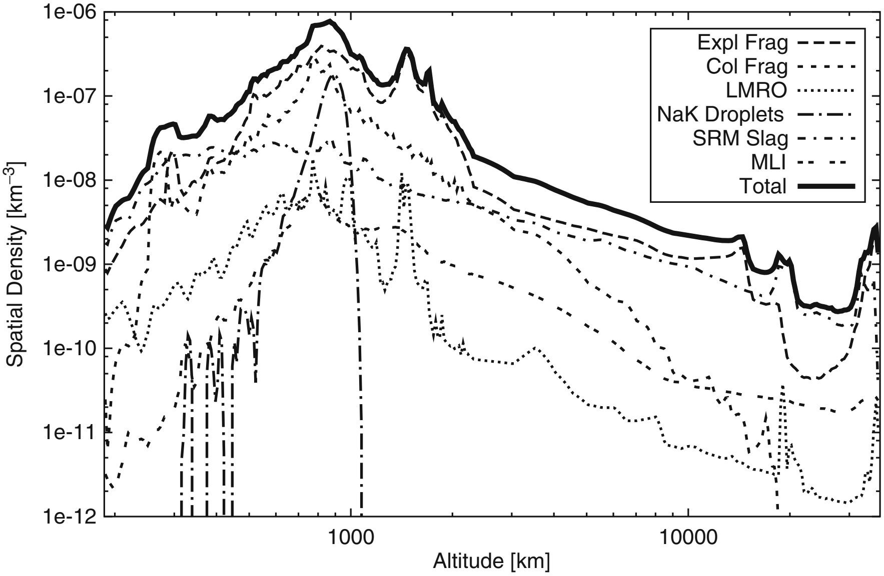


Fig. 7 Spatial density distribution of MASTER-2009 objects of d > 1 cm, in LEO to GEO altitudes, discriminated by sources

approach from the flight direction, and that the minimum relative velocity can be close to zero, for a sideways approach from ±90°. Impacts from the rear quadrants can only occur for impactors that travel on the eccentric orbits, during their perigee passes. Likewise, impacts from 0° can only occur, if the impactor has an orbit with a "complementary inclination" of 180° minus the inclination of the target object. Only in that case can both objects be in the same orbit plane, on counter-rotating orbits, if their ascending orbit nodes are separated by 180°.

For typical target orbits defined in Table 5, the mean times between impacts by orbital debris of different sizes are listed in Table 6 for a common reference cross section of 1 m^2, assuming a spherical target object, and a debris environment according to MASTER-2009 (Flegel 2010). In accordance with spatial densities shown in Figs. 6, 7, and 8, the highest collision risk for any of the selected sample orbits is encountered for ERS-2 on a sun-synchronous orbit of 774 × 789 km at an inclination of 98.5°. Apart from the debris concentration at this altitude, the collision frequency is also driven by the collision velocity (see Eq. 1). For ERS-2, it attains a most probable value of about 14 km/s, which is close to the maximum for two circular orbits at this altitude. Objects that could impact at such velocities are originating from the complimentary inclination band close to 81.4° (=180°−98.6°, see Figs. 3 and 4). Since all major flux contributions are from inclinations $i \geq 65°$, the resulting collision velocities are mostly within 14 ± 2 km/s at impact azimuth angles $-30° \leq A \leq +30°$ (see Eq. 4), with particles mainly originating from breakup events for the size regime larger than 1 cm. The

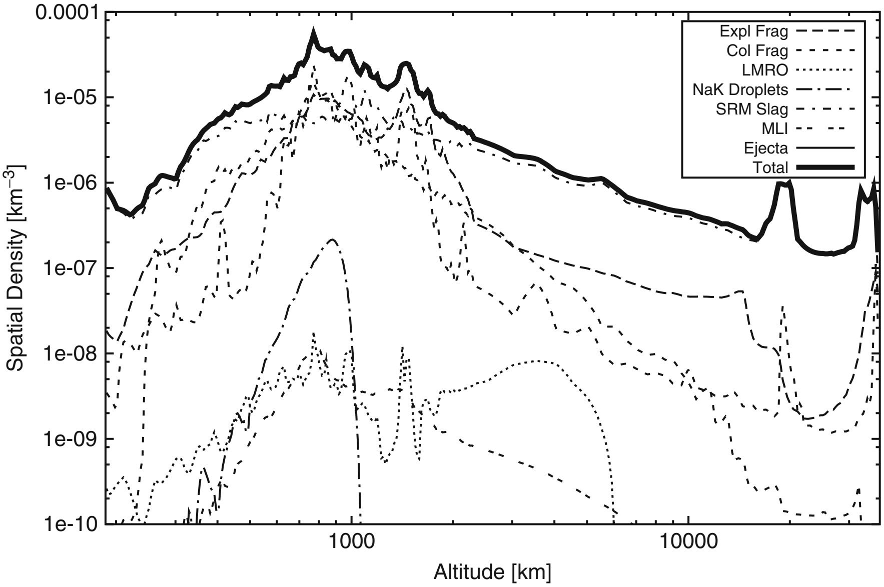


Fig. 8 Spatial density distribution of MASTER-2009 objects of d > 1 mm, in LEO to GEO altitudes, discriminated by sources

Table 5 Sample orbits for analyzing space debris collision flux

	H_p [km]	H_a [km]	i [deg]	a [km]	e [−]	ω [deg]
ISS	356	364	51.6	6,738	0.000601	0
ERS-2	774	789	98.6	7,159	0.001096	90
Globalstar	1,399	1,401	52	7,778	0.0001	0
GPS	19,997	20,003	55	26,378	0.0001	0
GTO	560	35,786	7	24,551	0.717405	178
GEO	35,782	35,790	0.1	42,164	0.0001	0

Table 6 Mean time between impacts of a given debris size for a spherical target of 1 m^2 cross section on sample orbits as defined in Table 4, according to ESA's MASTER 2009 space debris model

Diameter	>0.1 mm (days)	>1 mm (years)	>1 cm (years)	>10 cm (years)
ISS	9.0	636	41,102	942,507
ERS-2	0.7	42.5	1,252	43,783
Globalstar	1.7	102	9,208	126,550
GPS	244.8	10,794	1.1e + 7	7.2e + 8
GTO	36.8	2,627	241,546	4.4e + 6
GEO	676.3	18,674	6.5e + 6	1.4e + 8

situation changes for the ISS orbit. Its lower altitude goes along with a reduction of the debris flux by about 1 order of magnitude, and its lower inclination of 51.5° results in a gap of complementary inclination bands at $180°-51.6° = 128.4°$. The populated inclination bands only start at about 100°. As a consequence, there are no impacts from azimuth angles $-15° \leq A \leq +15°$, and most probable collision velocities are at 10 ± 1 km/s, resulting in approximately 50 % of the impact energy as compared to ERS-2. In contrast to ERS-2, slag residues from SRM firings are dominating the 1 cm debris population for ISS. They mostly reside on highly eccentric orbits of low inclinations, with perigee velocities that allow low-velocity impacts also from rear quadrants of ISS azimuth angles. When looking at a typical geostationary target orbit, the spatial density of the debris environment as compared to the LEO peak drops by about three orders of magnitude for the 10 cm population and by about two orders of magnitude for the 1 cm population. For the GEO orbit velocity of about 3 km/s, the predicted collision velocities are in the range of $0 \leq v \leq 1.5$ km/s, with a most probable value of 0.8 km/s, caused by old GEO objects that reached a maximum inclination excursion of 15° due long-periodic orbit perturbations with a period of 53 years. Due to the low relative velocities, the impact azimuth angles are mostly at ±80°. There are minor flux contributions from objects on GEO transfer orbits (GTO) and on 12 h Molniya orbits. They have apogee velocities of about 1.5 km/s, causing frontal impacts at 1.5 km/s on the faster GEO objects.

There are different ways to mitigate the risk and/or consequences of a collision of an operational spacecraft with a space debris object. For large-size catalog objects, the concept of conjunction event analysis and collision avoidance can be pursued. For subcatalog debris that cannot be tracked, passive protection measures can be taken.

To avoid catastrophic collisions with catalog-size objects of $d \geq 10$ cm, the ISS operators perform a conjunction event screening on the basis of the US Space Surveillance Network (SSN) catalog. This screening is performed at least three times a day, for 72 h ahead, in five steps:

1. Identification of approaches that fall within a 60 km radius, centered on the ISS (using US SSN orbit data in Two-Line Element (TLE) format)
2. Use of more accurate, osculating orbital elements, if the approach falls within $\pm 10 \times \pm 40 \times \pm 40$ km (radial × along-track × out-of-plane)
3. Consideration of orbit uncertainties, if the approach falls within $\pm 2 \times \pm 25 \times \pm 25$ km
4. Determination of collision probabilities, if approach falls within $\pm 0.75 \times \pm 25 \times \pm 25$ km
5. Decision on an evasive maneuver, if an accepted risk threshold is exceeded (e.g., 1 in 10,000)

In the first 4.5 years of operation, the ISS performed 7 debris avoidance maneuvers, with 3 of them executed by the visiting Space Shuttle. Due to improved procedures, based on more reliable orbit data, the subsequent avoidance maneuver was only 5.5 years later, executed by the attached ATV-1 on August 27, 2008, to avoid a fragment of Cosmos 2421. This fragment was 1 of 500 cataloged objects generated during three main breakup events in early 2008, just 60 km above the ISS altitude (Johnson 2009). By the year 2015, the ISS had performed 21 avoidance maneuvers since 1999, with 5 of these in 2014.

As is done by NASA for the ISS, ESA maintains a conjunction event analysis service for their operational LEO satellites. Once a day, the entire TLE catalog of the US SSN is screened for close conjunctions with the accurately known ESA spacecraft orbits for 7 days ahead. If the predicted collision probability exceeds a level of 1 in 3,000, then more precise orbit data are obtained for the conjunctor object, either through the processing of radar data from tasked observations or through conjunction characterization data obtained from JSpOC (US Joint Space Operations Center). In most cases, the more accurately know conjunctor orbit with its much reduced error dispersion leads to a maneuver suppression, even if the flyby geometry is unchanged. If, however, the collision probability remains at a level above 1 in 1,000, then a collision avoidance maneuver is initiated by the relevant project team.

Envisat, launched in 2002, had to perform five avoidance maneuvers up to December 2009. Due to the Chinese FengYun 1C ASAT test in January 2007 and as a result of the collision between Cosmos 2251 and Iridium 33 in February 2009, the debris environment at the Envisat and ERS-2 orbit altitude significantly deteriorated. As a consequence, the overall avoidance maneuver frequency in the year 2010 increased to 9 (4 each for Envisat and ERS-2, and 1 for Cryosat-2). The risk of catastrophic collisions of Envisat with a 10 cm fragment from the FengYun 1C and Cosmos 2251/Iridium 33 breakup events alone increased by +58 % as compared to the rest of the US SSN catalog. The risk of a mission terminating impact by a 1 cm class debris object even grew by +86 %, as compared with a modeled space debris population prior to these events. By 2015, ESA monitored close conjunctions for 6 of their operational LEO spacecraft. In 2014, they performed 12 evasive maneuvers for their satellite fleet.

To protect against nontrackable debris and meteoroids, the ISS has its manned modules covered by stuffed Whipple shields. For ESA's Columbus module, for instance, they consist of a 2.5 mm bumper and a 4.8 mm back wall, separated by an 11 cm standoff distance. Between the bumper and the back wall fabric layers of 4 mm Kevlar and 6 mm Nextel sheets are embedded as a "bullet-proof vest." The shields of the ISS manned modules can withstand impacts by objects up to 1.4 cm in size, at velocities on the order of 10 km/s. The related kinetic energy corresponds to a 1.5 t mid-size car hitting at 50 km/h, or to the energy released by an exploding hand grenade. An ISS module of 100 m^2 cross section is expected to have impacts from debris objects of $d \geq 1$ cm at a rate of 1 in 410 years. Meteoroid impacts are negligible in this size regime. For the same module cross section, impacts from objects of $d \geq 1$ mm will occur at a rate of 1 in 6 months, with a 90 % probability that they originate from meteoroids. Whipple shields rely on impact velocities that are larger than about 7 km/s, in order to break up the impacting object into a cloud of solid, liquid, and gaseous matter that can more easily be withheld by the back wall and intermediate fabrics, due to a wider spreading and time-distributed arrival of the fragment cloud, with a resulting reduction of the pressure peak. While the volume and mass requirements of such shields are prohibitive for normal spacecraft, there are still ways of reducing their impact risk. The Canadian Radarsat, for instance, used light-weight Nextel fabric covers as external protection and used rearrangements of sensitive spacecraft subsystems to improve the survivability of their 5-year mission by up to 89 %. This gain was achieved for a mass penalty of 0.6 % (+17 kg).

In order to increase the safety of US space assets, the US Space Command is upgrading its operational surveillance network. In particular, the replacement of the UHF-based surveillance fence that extends along the 33rd parallel across the United States by an S-band system is expected to allow catalog maintenance down to 2 cm sizes at the ISS altitude. This could increase the SSN catalog size to more than 100,000 objects. With the full orbit knowledge of these objects, one would be in a position to almost close the gap between avoidable and shieldable objects for ISS and hence significantly improve the on-orbit safety for manned space flight.

The space debris environment at critical sizes above 10 cm has in the past been dominated by explosion fragments and by dysfunctional but intact remnants of previous missions. Collisions played a minor role until the FengYun 1C ASAT test in 2007 and the accidental collision between Cosmos 2251 and Iridium 33 in 2009. By 2010, these two events alone accounted for almost 40 % of the US SSN catalog. In order to curtail the growth rate of hazardous space debris, particularly in the LEO regime, the international space community has identified and adopted a set of space debris mitigation measures. The main categories of recommendations can be summarized as follows:

- Reduction of mission-related objects
- Prevention of on-orbit explosions (passivation)
- Limitation of nonexplosive release events
- Collision avoidance between trackable objects and operational assets
- Postmission disposal of space systems

Mission-related objects (MROs) contribute 7 % of the trackable catalog population, with 66 % of these related to launch systems and 34 % related to payloads. MROs, also referred to as operational debris, are defined as objects released during nominal operations by both spacecraft and rocket bodies. This includes debris from launcher staging and payload separation (such as adapters, shrouds, and clamp bands) and objects released during spacecraft deployment and commissioning (such as parts of explosive bolts, solar array latches, and lens covers). Most of these objects are released with low relative velocities, and so they remain in close proximity to the operational orbit of the source object.

The release of MROs can be limited by system design. The best method of reducing the population of MROs is not to produce the objects in the first instance. This is reflected in most debris mitigation standards through recommendations to minimize or to avoid the use of debris-generating systems (e.g., yo-yo de-spinners, nozzle closures of propulsion systems, protective lens covers, etc.). System design is also encouraged to ensure that released parts (e.g., antenna deployment mechanisms, protective covers, explosive bolts, ullage motors, heat shields, etc.) are retained with the primary object. This can be achieved through the use of lanyards, sliding or hinged covers, and special catchment devices. Moreover, materials and basic system technologies (e.g., tanks, surface materials, structures, etc.) should be selected such that they are resistant to environmental degradation (e.g., aging by radiation, atomic oxygen and micro-particle impact erosion, and thermal cycling).

Explosions of spacecraft and upper stages in orbit have been the major source of debris in the past, more than 250 such events up to 2015, at a mean annual rate of about 5. These failures, which caused more than 10,000 on-orbit, cataloged fragments by 2015, might have been avoided, if on-board passivation techniques had been employed. Such procedures are a standard on many of today's launchers, and so far there are no recorded explosions of successfully passivated orbital stages. End-of-life (EOL) passivation was first considered as a design requirement at the beginning of the 1980s. All upper stages and spacecraft which were launched before then, and which are still in orbit, continue to pose an explosion hazard (note that a Titan III-C transtage launched in 1967 suffered an on-orbit explosion 27 years later). Hence, there are a significant number of latent explosion sources still on orbit.

Space debris mitigation standards recommend that all on-board reservoirs of stored energy (e.g., propellants, pressurants, batteries, momentum control gyros) should be permanently depleted when they are no longer required for any nominal or postmission operations. The following passivation aspects should be considered:

- Idle burn or venting of residual propellants, with valves left open
- Venting of all pressure systems and/or activation of pressure relief mechanisms to avoid explosions due to external heating
- Discharge of batteries, shut down of charging lines, and maintenance of a permanent discharge state
- Deactivation of range safety systems
- Dissipation of energy contained in momentum control gyros

Fuel depletion or “idle” burns of orbital stages may be performed such that the resulting thrust leads to a braking maneuver, leaving the stage in a reduced-lifetime orbit. The residual lifetime should be less than 25 years to be compliant with international recommendations for space debris mitigation.

The class of non-breakup release events includes residues from SRM firings (slag and dust), sodium-potassium droplets that were generated during RORSAT reactor core ejections, or surface degradations products that are caused by aging paint coatings or multilayer insulations (MLI). All of these debris sources can be reduced or even suppressed in total through design measures.

Collision avoidance, as another debris mitigation measure, is nowadays implemented by many space operators for their operational payloads. This concept, however, can only be applied to about 5 % of the catalog population, assuming that less than 1,100 of the on-orbit payloads in 2015 were operational, of which about 80 % could be maneuvered. Hence, future collisions will most often occur between uncontrollable debris objects. To reduce the number of catastrophic collisions between large, intact but nonoperational objects, the use of ground-based lasers is investigated. If a close conjunction is predicted, then a radar-guided laser beam (see Fig. 10) could ablate material from one of the objects or use the impact from photons on the target to impart a momentum that could sufficiently alter the flyby distance to a safe level.

In 2015, the mean time between two catastrophic collisions in the LEO region was on the order of 5 years. One way of reducing future collision rates is through postmission disposal measures, i.e., through mass removal of (still) active space assets. International guidelines recommend removing spacecraft and orbital stages after their mission completion, in particular from the densely populated LEO regime and from the unique GEO ring. For GEO spacecraft disposals, an orbit raise to a graveyard region at approximately 300 km above GEO is recommended. The magnitude of the altitude raise to a near-circular disposal orbit is determined by the area-to-mass ratio of the spacecraft. It is defined such that long-term orbit perturbation effects will not lead to a return of the orbit into a “GEO protected region” that extends ±200 km around the GEO ring (which is at 35,786 km altitude). Table 7 shows a summary of GEO postmission disposals over an 11-year time span. It is evident that the degree of compliance with international guidelines has gradually improved and has reached a level of about 70 % in 2014.

The end-of-life mitigation measures for the “LEO protected zone” (that is below 2,000 km altitude) recommend an active deorbiting or a natural decay of payloads and orbital stages into a destructive reentry within 25 years after mission completion. For typical area-to-mass ratios of such objects, a timely natural decay requires an end-of-mission altitude below 600 km. Alternatively, chemical or combined chemical/electrical propulsion can be used to induce a direct reentry. A monopropellant hydrazine system would need about 8.8 % of the spacecraft mass for a controlled deorbit from 800 km (6.3 % for a bi-propellant system). Electrical propulsion systems, due to their higher ejection velocities, can be more mass efficient by a factor of about 10. Their lower thrust levels, however, will lead to an uncontrolled reentry. An accelerated uncontrolled reentry can also be induced by thin, conductive tethers of several kilometers length that orientate themselves along the local vertical through gravity gradient forces. As they cut through the magnetic field lines, they induce a tether current that is closed through the ambient plasma and that leads to a retarding Lorentz force, acting opposite to the direction of motion of the spacecraft, with best performance for low-inclination orbits. For a mass penalty of less than 3 %, such systems are able to reduce orbital lifetimes of Globalstar satellites (at 1,400 km and 52° inclination) from 9,000 years to less than 2 months, and they can reduce orbital lifetimes of Iridium satellites (at 780 km and 86° inclination) from 100 years to less than 8 months.

Space debris mitigation guidelines, standards, and requirements have been developed by several space agencies since the early 1990s. In parallel, the knowledge on space debris sources increased, and the understanding of effective remedial actions improved.

Table 7 Eleven-year history of post-mission disposal activities of geostationary spacecraft through 2014 ($L_1 = 75°E$, $L_2 = 105°W$, "too low," and "compliant" refer to the IADC orbit raising recommendation in the IADC Space Debris Mitigation Guidelines (Anonymous 2002); Credit: ESA)

EoL disposal	'04	'05	'06	'07	'08	'09	'10	'11	'12	'13	'14	11-Year total	
Left at L_1	2	1	2	1	2	3	1	–	1	–	2	15	(8.3 %)
Left at L_2	1	1	1	–	1	–	–	–	–	–	–	4	(2.2 %)
Left at L_1/L_2	–	1	–	–	1	–	–	–	–	–	–	2	(1.1 %)
Drift orbit (too low)	5	5	7	1	1	6	4	3	4	5	3	44	(24.5 %)
Drift orbit (compliant)	5	11	9	11	6	12	11	12	10	15	13	115	(63.9 %)
Annual total	13	19	19	13	11	21	16	15	15	20	18	180	(100 %)

A first step to a wider, international application of debris mitigation measures was taken by the Inter-Agency Space Debris Coordination Committee (IADC) in 2002, with the publication of their Space Debris Mitigation Guidelines (Anonymous 2002). This document, which was first presented at the UNCOPUOS Scientific and Technical Subcommittee in 2003, serves as a basis for the development of space debris mitigation principles in two directions: towards a nonbinding policy document and towards applicable implementation standards. The former route was followed by a UNCOPUOS working group, while the latter direction was pursued by an Orbital Debris Coordination Working Group (ODCWG) within the Technical Committee 20 and its Subcommittee 14 of the International Organization for Standardization (ISO TC20/SC14). To a large extent, these UN and ISO working groups recruit their experts from IADC member organizations.

International space debris mitigation policies and standards, based on the consensus of the IADC guidelines, could in the future facilitate and harmonize the implementation of space debris mitigation measures at a global scale. Internationally agreed standards could enforce appropriate debris mitigation measures on spacecraft operators and launch service providers through the mechanisms of conditional launch license issuance and insurance coverage, depending on the acceptance of a space debris mitigation plan by the launch authority. More than 50 years after the beginning of space flight, the voluntary implementation of debris mitigation and disposal measures by many space operators has become common practice. For several launching nations, the compliance with national regulations or with a national space law makes debris mitigation measures even mandatory.

While debris mitigation is a necessary condition to maintain an orbital environment with a tolerable risk level for space missions, long-term forecasts of the debris environment indicate that some orbit regions may still become unstable within a few decades. Figure 9 illustrates the evolution of the LEO debris population larger than 10 cm for a hypothetical case of no future launches. The case corresponds to an extreme, hypothetical mitigation scenario, with immediate deorbiting of payload (s) and insertion stage(s) after orbit injection, and with no intermediate release of mission-related objects. Predictions with NASA's LEGEND model (Liou and Johnson 2008a, b) demonstrate that even for such optimistic assumptions, the LEO environment will become unstable. Within 20 years, collision fragments will start to outnumber explosion fragments, and within 70 years, an initially stabilizing effect from naturally reentering objects will be superseded, and the 10 cm population growth will follow the slope of the collision-induced fragment increase. In the course of the 200 year projection, more and more collision fragments will collide with other collision fragments. This so-called Kessler syndrome is a self-maintained collisional cascading process that is fed by the LEO mass reservoir of 2,350 t in 2015. Its natural termination would only be reached in the very far future when all LEO crossing objects are ground to subcritical sizes that can no more reach the specific impact energy threshold of ≥40 kJ/kg for causing a catastrophic breakup. As a consequence, space debris mitigation is a necessary but insufficient condition to maintain a stable orbital environment. This is even more noteworthy in the light of current compliance rates with recommended end-of-mission disposal procedures. These compliance

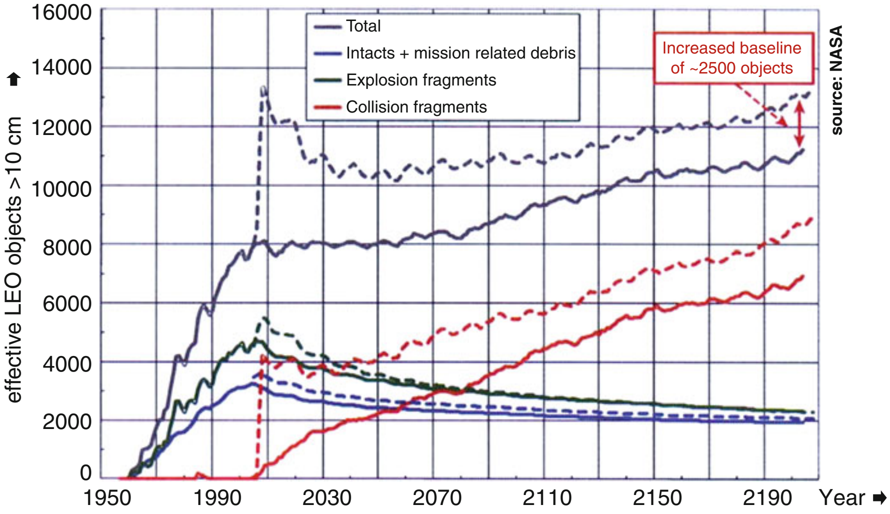


Fig. 9 Long-term prediction of the LEO debris environment of critical-size objects of d > 10 cm, discriminated by source terms, for a "no future launch" scenario (Liou and Johnson 2008a)

rates are on the order of 60 % overall, and only 30 % for maneuverable spacecraft (Krag et al. 2015).

In order to sustain an acceptable debris risk level for future space missions, debris mitigation measures must be augmented by space debris environment remediation measures that actively remove mass from orbit, with priority on the LEO regime (Klinkrad and Johnson 2009, 2013). The effectiveness of space debris environment remediation measures is governed by their capability to reduce the short- and long-term risk of catastrophic collisions. An initial indicator of the debris environment deterioration is the concentration of critical-size objects of 10 cm and larger that have the capability to cause catastrophic breakups. Figure 4 shows the distribution of the catalog objects in LEO. Highest concentrations are at 800 ± 200 km, spread over inclination bands at 65 ± 2°, 72 ± 2°, 82 ± 1°, and 97 ± 3°, with an almost equal share of intact objects, explosion fragments, and collision fragments. There is a lower, secondary LEO peak at 1,400 ± 100 km and minor local peaks for MEO navigation satellite constellations and for GEO objects, both of which are about three orders of magnitude lower.

In order to rank priorities for mass removal from orbit, it is important to determine a risk metric. Three parts of such a metric can be defined:

- metric #1: [catastrophic collision rate] = [10 cm collision flux] × [mean target cross section]
- metric #2: [short-term risk due to a catastrophic collision] = [10 cm collision flux] × [mean target cross section] × [target mass]
- metric #3: [long-term risk due to a catastrophic collision] = [10 cm collision flux] × [mean target cross section] × [target mass] × [target orbit lifetime]

In these metrics, as a simplification, it shall be assumed that the target cross section and the target mass are significantly larger than those of the impactor and that the orbit lifetime of resulting collision fragments is similar to the orbit lifetime of the intact target object. The following results are extracted from (Klinkrad and Johnson 2013) for an analysis epoch in 2010.

In 2010, the orbit environment consisted of more than 12,000 cataloged LEO objects, larger than 10 cm, of a total mass of almost 2,300 metric tons. The corresponding rate of catastrophic collisions was 0.19 per year, resulting in one such event every 5 years. About 45 % of these collisions would have a rocket body, while 55 % would have a spacecraft as their main object. For metric #1 (catastrophic collision rates), as much as 22 % can be attributed to a single 2° × 50 km bin at 87 ± 1° inclination and 775 ± 25 km altitude, covering 80 intact objects, of which 73 are satellites of the Iridium constellations, each with 660 kg mass and 22 m^2 cross section. These intact objects are facing fragments from the Iridium 33/Cosmos 2251 collision of February 10, 2009, and from the Chinese ASAT test of January 11, 2007, as the main causes of their 10 cm collision flux. A secondary maximum of catastrophic collision rates at 11 % is due to a cluster of Cosmos satellites at a bin of 83° ± 1° inclination and 975 ± 25 km altitude.

The short-term risk to the orbital debris environment can be expressed by the metric #2, where the dominant target object masses drive the number of critical-size collision fragments, which determine the short-term level of debris environment deterioration. Using the same assumptions for determining catastrophic collision rates as above, the mass-weighted short-term environment risk is governed to 61 % by rocket bodies and to 39 % by spacecraft. Approximately 28 % of the overall short-term risk is due to objects in a single bin of 2° × 50 km, centered at 71 ± 1° inclination and 825 ± 25 km altitude. Most of the corresponding mass is related to Russian Zenit-2 second stages with dry masses of 8,900 kg and cross sections of 33 m^2 each, and to 15 Cosmos spacecraft of 3,200 km mass and 6 m^2 cross section each. Of the 20 top-ranking objects according to metric #2, 19 are Zenit 2 rocket bodies, 16 of which are located in the above-defined bin.

The long-term risk to the orbital debris environment can be expressed by the metric #3, where the on-orbit residence time of resulting collision fragments is applied as a weighting factor to the metric #2 contributions. As a simplifying, conservative assumption, the same average orbital lifetimes shall be considered for the target object and its resulting fragments. The resulting aggregate of the individual products of catastrophic collision rate, target mass, and target orbit lifetime, over all intact LEO objects below 1,300 km, leads to a long-term debris environment risk indicator that is governed to 72 % by rocket bodies and to 28 % by spacecraft. Approximately 42 % of the overall long-term risk is due to the same objects that dominate the risk metric #2, stemming from a single bin of 2° × 50 km, centered at 71 ± 1° inclination and 825 ± 25 km altitude. Again, most of the related mass is due to Russian Zenit-2 second stages, each with an empty weight of 8,900 kg, with a cross section of 33 m^2, and with an orbit lifetime on the order of 700 years. Of the 23 top-ranking objects according to metric #3, 19 are Zenit-2 second stages, with 16 thereof from a single 2° × 50 km bin. By removing these stages, the long-term risk metric #3 could be reduced by about 24 %.

Long-term debris environment projections (see Liou and Johnson 2008a, b; Bastida and Krag 2009; Liou 2011; Klinkrad and Johnson 2013) based on an extreme scenario with no future launches and 90 % success rates of LEO postmission disposals indicate that the current environment will lead to a net increase of the long-lived 10 cm debris population by about 30 % in the next 200 years (see Fig. 9). This result confirms the onset of collisional cascading in some LEO orbit regions, also known as the Kessler syndrome. In the case of continued launch activities at today's rates, the 10 cm debris population will even grow by 60 %, fueled by 24 catastrophic collisions (Flohrer et al. 2011). These collisions will almost exclusively occur between members of the previously identified, densely populated LEO inclination bands and between orbits of low to moderate eccentricities. Further parametric studies of the long-term debris environment evolution predict that active mass removal, focusing on inclination and altitude bands with high mass concentrations in a few large objects, can reduce the number of catastrophic collisions to 14 within 200 years and lead to a stable 10 cm object population, if 5–10 removals per year are performed (Bastida and Krag 2009; Liou 2011; Klinkrad and Johnson 2013).

Several research groups, with different backgrounds and application targets, have devised techniques that could be used for the removal of mass from orbit. Table 8 shows an overview of methods that could be within technological reach (Klinkrad and Johnson 2013). With the exception of ground- and air-based directed energy methods (mainly lasers, see Fig. 10), all techniques are space based, and all of them are suited for the most critical LEO regime (with some also applicable for MEO or GEO mass removals). All methods in Table 8 that are restricted to debris sizes below 10 cm can contribute to space environment remediation but at a size regime that normally does not lead to catastrophic collisions and that hence does not fuel the collisional cascading process. The focus shall thus be on techniques that can effectively reduce the orbit lifetime of intact objects and fragments that are larger than 10 cm, including full-size satellites and orbital stages. In order to qualify as a remediation measure (as opposed to a mitigation measure), all techniques must be applicable to dysfunctional target objects, for instance, with the assistance of a remover spacecraft or through the attachment of external deorbiting devices in a rendezvous mission.

Solar sails can be used to increase the eccentricity of a target orbit. The periodic changes in perigee altitude, in combination with the increased, nonconservative drag perturbation at the perigee passes, lead to a secular decrease of the orbit energy and hence accelerates the orbit decay. The decay rate is directly proportional to the area-to-mass ratio of the solar sail/spacecraft compound. Solar sails could be inflated spheres or arrangements of flat surfaces. They should be metalized to increase the photon-surface momentum exchange (see Fig. 11). Drag augmentation devices directly affect the area-to-mass ratio of an object and hence increase the air drag that leads to an orbit lifetime reduction. For both, solar sails and drag augmentation devices, the benefit of reducing the orbit lifetime of the target object should outweigh the drawback of an increased collision cross section.

Table 8 Debris removal techniques and their applicability with respect to orbit regime and debris size (Klinkrad and Johnson 2013)

Debris removal technique	Altitude regime	Debris size
Solar sail	LEO, MEO, GEO	>1 m
Magnetic sail	LEO, MEO, GEO	>1 m
Attachable deorbit/reorbit module	LEO, MEO, GEO	>1 m
Capture/orbital transfer vehicle	LEO, MEO, GEO	>1 m
Drag augmentation device	LEO	>10 cm
Momentum tethers	LEO, GEO	>10 cm
Electrodynamic tethers	LEO	>10 cm
Airborne laser/directed energy	LEO	<10 cm
Space-based laser/directed energy	LEO, MEO, GEO	<10 cm
Ground-based laser/directed energy	LEO	<10 cm
Space-based magnetic field generator	LEO	<10 cm
Sweeping/retarding surface (balloon, film, foam ball, etc.)	LEO	<10 cm

The magnetic sail concept is using a magnetic field to deflect the plasma of the solar wind in order to accelerate or decelerate a spacecraft. The magnetic sail utilizes a loop of superconducting cable to which an electrical current is applied. The magnetic field created by the current in the loop stiffens the cable into a rigid circular shape. Charged particles encountering the magnetic field are deflected, and momentum is imparted on the loop. In the solar wind, the magnetic sail creates drag and accelerates the spacecraft in the direction of the wind. Employing the magnetic sail in nonaxial configurations produces a force perpendicular to the solar wind that can be used for maneuvers. However, the technical implementation of the concept is not yet mature, and it would be vulnerable to small particle impacts from debris and meteoroids.

Tethers can be applied in mass removal systems in two different ways: as conductive electrodynamic tethers or as momentum exchange tethers. When two sizeable objects are connected by a momentum exchange tether, and if this tether is reeled out along the local vertical, then different orbit velocities and perturbing accelerations cause a swinging motion, primarily within the common orbital plane. If the tether is then cut at the time of its highest retro-grade ΔV, then the lower object will obtain a lower perigee (e.g., in LEO for direct deorbit, or for release into a reduced-lifetime orbit) and the upper object will obtain a higher apogee. Likewise, in a reverse mode, such a system can be used to reorbit MEO or GEO objects. The related tether loads can be significant, and the tether design is technologically demanding. For a net gain, the active remover satellite would have to deorbit more than one large object, and also dispose of itself.

In the case of an electrodynamic tether (EDT), an electromotive force is generated within a conductive wire that is attached to a space vehicle as it moves through the Earth's magnetic field. If a pair of plasma contactors at either end of the tether emits and collects electrons, an electric current flows through the tether by closing the circuit via the ambient plasma. The tether then generates a Lorentz force via

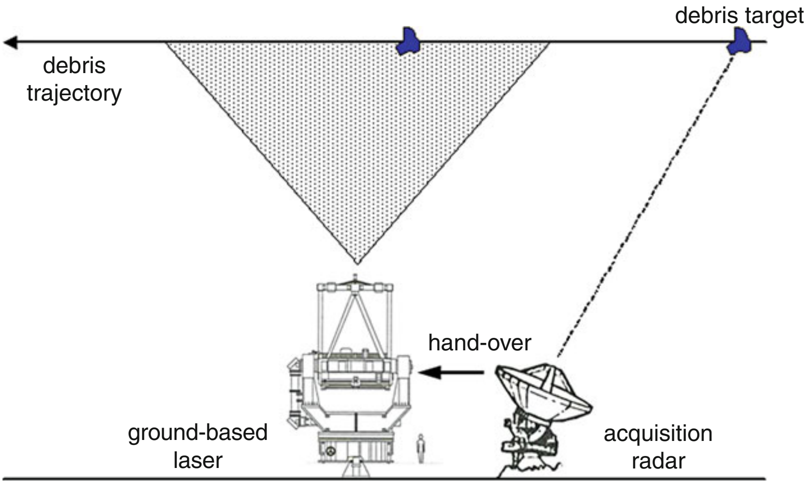


Fig. 10 Debris removal and/or debris orbit changes induced by a ground-based laser, tasked by a colocated acquisition and tracking radar (Credit: NASA)

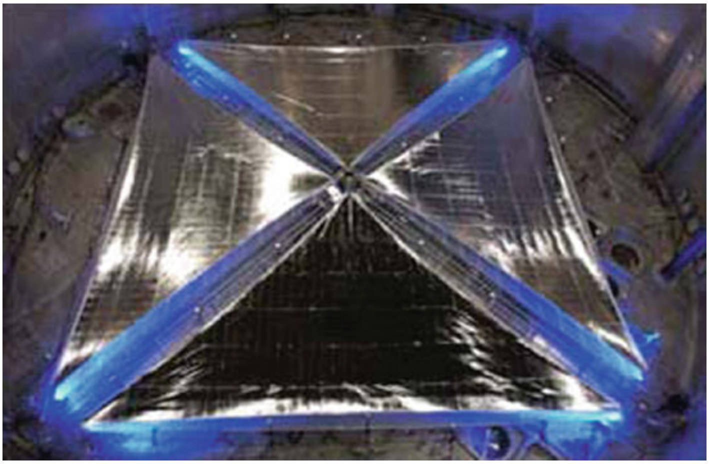


Fig. 11 Use of solar radiation pressure and/or drag augmentation for LEO orbit lifetime reduction

interaction between the tether current and the geomagnetic field. This force acts as a deceleration, opposite to the direction of flight, and it hence reduces the orbit lifetime by dissipating orbital energy. The efficiency of this method depends on the average magnetic induction, and it thus decreases with $1/r^3$ and with cos i_m, where r is the geocentric radius and i_m is the mean geomagnetic inclination of the orbit. The resulting, reduced orbit lifetime is proportional to $1/L^2$, where L is the tether length. Figure 12 shows how a servicing satellite attaches an EDT to a dysfunctional payload to accelerate its decay. An electrodynamic tether is a promising deorbit

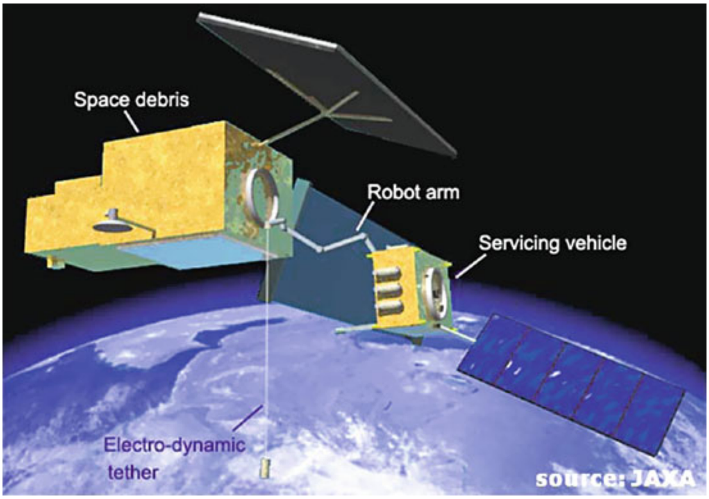


Fig. 12 Use of conductive tethers for LEO orbit lifetime reduction through electrodynamic forces induced by the geomagnetic field (Credit: JAXA)

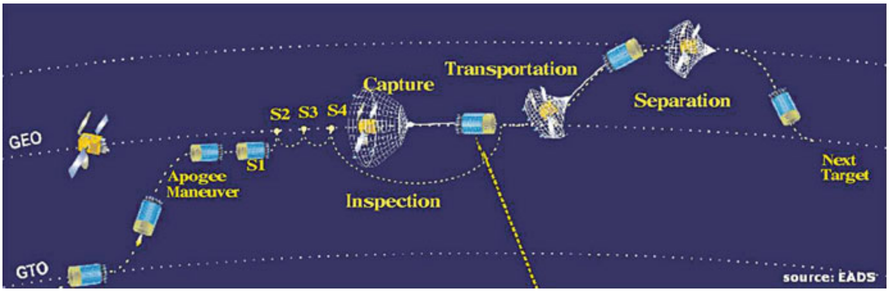


Fig. 13 Use of a space tug to reorbit objects from the GEO region (Credit: EADS Astrium)

concept due to its relatively simple design, its low system mass, and its efficiency even at high LEO altitudes (Pardini et al. 2006; Klinkrad and Johnson 2013). A conductive aluminum tether with a system mass fraction of 2.5 % as compared to the client object can reduce the lifetime of a high-inclination, low LEO constellation at 780 km altitude from 100 years to less than 1 year (e.g., for Iridium). For a medium-inclination, high LEO constellation at 1,400 km, the orbital lifetime can be reduced from 9,000 years to less than 2 months (e.g., for Globalstar).

Technologically, the most mature solutions of orbit mass removal are the attachments of de- or reorbit propulsion modules or the capture of a target object by a space tug. Figure 13 shows the latter approach for a GEO tug satellite. The shown ROGER remover spacecraft of 3.5 t mass is intended to be launched as a secondary payload into a GTO orbit, from which it injects itself into the GEO ring. It then performs rendezvous operations with a preassigned client object, inspects it via video cameras, and casts a net over it. The net is then tightened via reels in the end masses while leaving it connected to the remover spacecraft by means of a tether. The whole compound is then tugged into the GEO graveyard orbit, where the tethered net enclosing the debris object is released. With an overall propellant mass fraction of 77 % and 20 disposable nets (or, alternatively, 10 disposable nets and 2 reusable, tethered gripper devices), the ROGER remover satellite is planned to perform up to 20 GEO disposal missions. While this concept is particularly attractive for GEO, it could also be employed in densely populated LEO regions. Several similar concepts, mostly based on prior studies of servicing spacecraft, are under investigation in space industry.

Mass removal from orbit has a technical, a financial, and a legal dimension. As of today, many of the solution concepts listed in Table 8 are not yet sufficiently advanced in their technology readiness, and even the most mature concepts would incur significant costs if they were realized. Moreover, the removal of on-orbit mass that belongs to another launch authority and/or space operator requires mutual agreement on the procedure, on the cost sharing, and on possible liabilities, particularly for an uncontrolled reentry.

Conclusion

Out of 17,754 objects that were contained in the US Space Surveillance Network catalog in January 2016, approximately 1,100 were operational spacecraft, of which roughly 80 % could be maneuvered, and of which 400 were in the GEO ring, while more than 500 were in the LEO regime. Since LEO and GEO are of particular interest for space operators, these orbit regimes were denoted at “protected regions” by IADC and UNCOPUOS. In order to safeguard a sustainable long-term usability of the LEO and GEO regions, space debris mitigation measures must be applied rigorously by all space faring nations and supernational organizations. The necessary mitigation measures have been identified, e.g., by the 13 IADC members, and cast into international guidelines and standards, into agency-specific sets of requirements and/or into national space laws. Analyzes of the long-term evolution of the space debris environment indicate that such agreed mitigation measures are a necessary but insufficient condition to maintain the space object population at a stable level. Even an extreme mitigation scenario with no future launches will result in a long-term collisional cascading (the so-called Kessler syndrome) at some LEO altitudes. This runaway process is fueled by existing mass on orbit, and the only way to stabilize the environment is through active mass removal from particularly densely populated altitude and inclination bands. This is a challenging task from a technical, economical, and legal point of view that can only be successfully implemented, if an international consensus is reached among space faring nations. In the past, the Scientific and Technical Subcommittee (STSC) of UNCOPUOS, with guidance from IADC members and contributions from COPUOS members, installed a working group that developed the UNCOPUOS Debris Mitigation Guidelines (Anonymous 2009) in the course of a multiyear work plan. Likewise, in 2010, UNCOPUOS

STSC established a working group on the "sustainable use of outer space." This initiative could be a starting point for the development of an international framework that could include space debris environment remediation as one of its main objectives.

Following the publication of previous reports on "space traffic management" and "space debris mitigation," the International Academy of Astronautics (IAA) has published a report on "space debris environment remediation" in 2013 (Klinkrad and Johnson 2013). Its authorship, with more than 20 contributors from 11 different countries and many different disciplines, could further consolidate the basis for international deliberations on the technical, economical, and legal aspects of mass removal from orbits with critical mass concentrations, to allow a continued and safe use of space also in the far future.

The overarching principle of a responsible and sustainable use of space was formulated back in the 1990s by the late Joseph P. Loftus, former assistant director of NASA/JSC: "Space operations should comply with a general rule of the National Park Service: 'What you take in you must take out'."

Cross-References

- ▶ The World's Launch Sites
- ▶ Major Launch Systems Available Globally
- ▶ Space Weather and Hazards to Application Satellites

References

Anonymous, IADC space debris mitigation guidelines. IADC-02-01, rev.1 (2002)

Anonymous, UNCOPUOS space debris mitigation guidelines. A/RES/62/217, UNCOPUOS Scientific & Technical Sub-Committee, Vienna (2009)

B. Bastida, H. Krag, Strategies for active removal of space debris, in *Proceedings of the 5th European Conference on Space Debris*, ESA-SP-672, Darmstadt (2009)

S. Flegel, *Maintenance of the ESA MASTER Model.* Final report of ESA contract 21705/08/D/HK (2010)

T. Flohrer, R. Choc, B. Bastida, Classification of geosynchronous orbits. Issue 13. European Space Agency, GEN-DB-LOG-00074-OPS-GR (2011)

N.L. Johnson, The International Space Station and the space debris environment – 10 years on, in *Proceedings of the 5th European Conference on Space Debris*, ESA-SP-672, Darmstadt (2009)

H. Klinkrad, *Space Debris – Models and Risk Analysis* (Springer-Praxis, Berlin/Heidelberg/New York, 2006)

H. Klinkrad, N.L. Johnson, Space debris environment remediation concepts, in *Proceedings of the 5th European Conference on Space Debris*, ESA-SP-672, Darmstadt (2009)

H. Klinkrad, N.L. Johnson (eds.), *Space Debris Environment Remediation* (Study by the International Academy of Astronautics (IAA), Paris 2013). ISBN 978-2-917761-30-4

H. Krag, S. Lemmens, S. Frey, B. Bastida Virgili, Current practices in implementing mitigation measures for LEO missions, in *66th International Astronautical Congress*, Paper IAC-15-A6.6.04, Jerusalem (2015)

J.C. Liou, N.L. Johnson, Instability of the present LEO satellite populations. Adv. Space Res. **41**, 1046–1053 (2008a)

J.C. Liou, N.L. Johnson, A sensitivity study of the effectiveness of active debris removal in LEO. Acta Astronaut. **64**, 236–243 (2008b)

J.C. Liou, An active debris removal parametric study for LEO environment remediation. Adv. Space Res. **47**, pp.1865–1876 (2011)

M. Oswald, P. Wegener, S. Stabroth, C. Wiedemann, J. Rosebrock, C. Martin, H. Klinkrad, P. Vörsmann, The MASTER 2005 model, in *Proceedings of the 4th European Conference on Space Debris*, ESA-SP-587, Darmstadt (2005)

C. Pardini, T. Hanada, P.H. Krisko, Benefits and risks of using electro-dynamic tethers to de-orbit spacecraft, in *57th International Astronautical Congress*, IAC-06-B6.2.10, Valencia (2006)

Coping with the Hazards of Space Debris

Joseph N. Pelton

Contents

Abstract

The issue of space debris has become one of increasing concern as the amount of debris has become more severe, especially in low Earth orbit and polar orbits used for communications, remote sensing, and meteorological sensing and forecasting. The Chinese missile shootdown of the defunct Fengyun (FY-1C) weather satellite in 2007 and the collision of the Iridium and Cosmos satellites in 2009 have greatly heightened this concern. Increasingly sophisticated tracking systems have been implemented by the US Air Force, the European Space Agency, and the several affiliated national tracking system, plus tracking systems in Australia and other parts of the world, and more radar and optical tracking systems are planned. The new S-band space fence system, in particular, will allow an increase of tracking of space debris from about 23,000 elements that are 10 cm or larger (i.e., about the size of a baseball) in low Earth orbit to well over 200,000 elements

---

J.N. Pelton (✉)
International Space University Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_99

that are greater than 1 cm in diameter (i.e., about the size of a marble) in low Earth orbit. The Space Data Association that has been formed by commercial satellite operators is increasingly able to share information among themselves to minimize the possibility of collisions and to be aware of close conjunctions in a timely manner.

In addition, new laws and national regulations as well as guidelines adopted by the Inter-Agency Space Debris Committee (IADC) and the UN Committee on the Peaceful Uses of Outer Space (COPUOS) to ensure that all satellites are deorbited within 25 years of the end of spacecraft life represent key steps forward. There are clearly more steps that need to be taken to move toward better collision avoidance systems plus active deorbit and debris mitigation, especially of the largest debris elements from low Earth orbit. It is also key to ensure that the deployment of new large-scale constellations in low Earth orbit is accomplished with strict controls to minimize any new collisions that might occur within these constellations themselves or to avoid collision with defunct debris elements. The addition of constellations with a thousand spacecraft or more in just one constellation has given rise to particular concerns in this regard.

In addition, there needs to be (i) new and better international collaboration to strengthen all elements associated with the more precise tracking of debris in all Earth orbits; (ii) more control processes to prevent debris increase and avoid the formation of new debris elements, including the active deorbit of all launch systems after they have inserted spacecraft into orbit; (iii) better coordination of information among satellite system operators through such mechanisms as the Space Data Association as its membership and participation levels grow; and (iv) new technology and international agreements and perhaps commercial arrangements to incentivize the active deorbit of space debris in future years consistent with existing space treaties and international agreements.

This chapter addresses in some detail the various tracking capabilities that exist or are planned around the world to monitor the orbits of space debris and to provide alerts so as to avert possible conjunctions. It provides information about how these systems are being upgraded, and space situational awareness is being coordinated over time. It notes how governmental systems are being augmented by private capabilities that are able to augment space situational awareness and to assist with avoidance of collision. These systems and processes will perhaps assist with future space debris mitigation and active removal. All of these increasing space situational capabilities are crucial to the future successful operation of application satellites in the twenty-first century.

Keywords

Active space debris removal • ESA • EISCAT • Image Information Processing Center and Supercomputing Facility (IIPCSF), NASA • Optical tracking, Space Data Association • S-band radar space fence • Satellite conjunctions • SPACETRACK • TIRA • UN COPUOS • US Strategic Command (USSTRATCOM) • US Space Surveillance Network • Working Group on the Long-Term Sustainability of Space Operations

Introduction

At the dawn of the space age, a half century ago, the idea that human-manufactured orbital space debris would be a major concern to commercial organizations operating networks of communication satellites, remote sensing networks, navigation satellites, and meteorological satellites was almost unthinkable. Few of these space systems even existed, and the vast reach of outer space was truly enormous. Just the volume of space that surrounds the Earth out to geosynchronous orbit represents an astonishingly large 300,000,000,000,000,000 (3×10^{17} Km^3). The space around our planet is a very large neighborhood. And at the outset, satellites were quite small and compact – the size of beach balls. But spacecraft and rockets became larger and larger, and more and more satellites were launched.

A lack of care was taken about explosive bolts, upper stage rockets left in orbit, and satellites were launched and then deserted year after year. There were no rules about deorbiting defunct satellites at the end of life. Each year the amount of orbital debris increased, and the situation that Dr. Donald Kessler of NASA warned about back in the 1980s – that of substantial debris buildup – has come to pass.

On January 11, 2007, China conducted a now widely publicized antisatellite missile test to shoot down the defunct Chinese weather satellite, the FY-1C polar-orbiting satellite of the Fengyun series. This 750 kg satellite was hit at an altitude of 865 km (537 mi) and was instantly splintered into over 2000 trackable space debris elements. The so-called kill missile was traveling in the opposite direction of the satellite at a speed of 8 km/s (or 28,800 km/h).

Then 2 years later, there was a collision of the Iridium 33 and defunct Kosmos 2251 satellite at a relative velocity of 42,000 km/h that occurred on February 20, 2009, at 16:56 Universal Time Coordinates (UTC). This violent intersection also created over 2000 new debris elements. Today, the space debris problem continues to increase. According to Dr. Kessler who first predicted the “Kessler syndrome” and the possibility of an ever-increasing cascade of space junk, there is now a “likelihood” of a major space collision every 10 years. He also explains that the cascade effect that is now occurring will create more and more debris elements even with no additional launches – and of course we are planning an ever-increasing amount. So what are we to do to preserve space operations for the future?

These two major collision events have underlined the dangers and demonstrated the increasing difficulty of controlling the space debris problem particularly in low Earth orbit and the congested polar orbits. The nearly 3000 metric tons of debris in low Earth orbit is thus of increasing concern (See Fig. 1).

US Space Surveillance Network detects, tracks, catalogs, and identifies artificial objects orbiting the Earth, i.e., both active and inactive satellites, spent rocket bodies, or fragment debris. The system is the responsibility of the Joint Functional Component Command for Space, part of the US Strategic Command (USSTRATCOM).

The current US standard for protecting astronaut-occupied spacecraft is to maneuver to avoid an object if it is calculated to have a higher than 1:10,000 chance of hitting the asset. The objective is to create a 200 km buffer zone or “bubble of protection” around the International Space Station for instance. This provides less

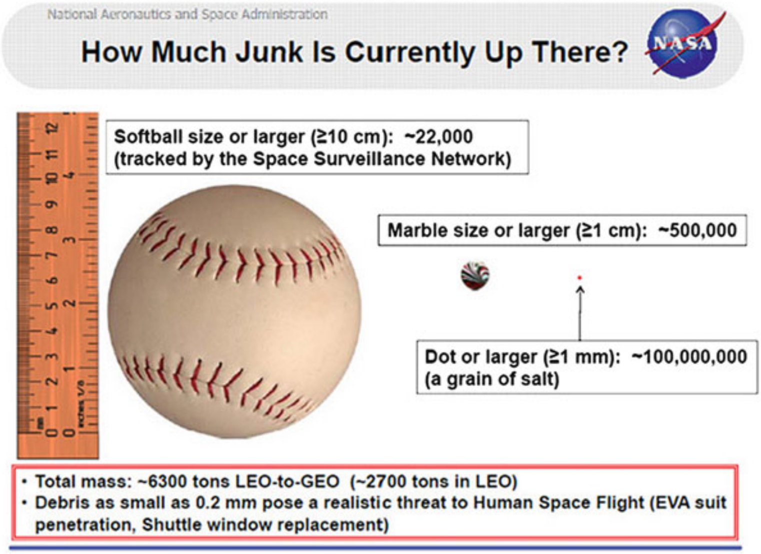


Fig. 1 The NASA score sheet statistics on space debris (Graphic courtesy of NASA)

than 30 s of separation and reaction time between the ISS and crossing orbital debris (Cooney).

The US Space Surveillance Network (SSN) has a specified set of strategic objectives as defined by the US Congress. These explicit duties include:

- Predict when and where a decaying space object will reenter the Earth's atmosphere
- Prevent a returning space object, which to radar looks like a missile, from triggering a false alarm in missile-attack warning sensors of the USA and other countries
- Chart the present position of space objects and plot their anticipated orbital paths
- Detect new man-made objects in space
- Correctly map objects traveling in the Earth's orbit
- Produce a running catalog of man-made space objects
- Determine which country owns a reentering space object
- Inform NASA whether or not objects may interfere with satellites and International Space Station orbits

The SPACETRACK program represents a worldwide Space Surveillance Network (SSN). This is a complex network of sensing devices that now includes electro-optical, passive radio frequency (RF), and radar sensors. The SSN is largely composed of elements owned and operated by governmental agencies, but increasingly its capabilities are augmented by instruments owned and operated by private commercial firms as well. The SSN is tasked to provide space object cataloging and identification, satellite attack warning, timely notification to US forces of satellite flyover, space treaty monitoring, and scientific and technical intelligence gathering. This is a unique combination of tasks that are civilian space activities and US defense and military duties (US Space Surveillance Network).

The continued increase in satellite and orbital debris populations, as well as the increasing diversity in launch trajectories, nonstandard orbits, and more and more satellites – including small, micro, CubeSats, and so-called Femto (or microsatellites) – has made the task of monitoring the skies more and more difficult. This has led to the need to upgrade the SSN to meet existing and future requirements. It has also led to efforts to ensure the cost-effective operation of the SSN through automation where possible.

The prime area of upgrade for the SSN is the near-term creation of the new S-band space fence. On June 3, 2014, Lockheed Martin division won a $914 million contract to build this new space fence radar system in the Pacific Ocean area along the Kwajalein Atoll in the Marshall Islands. It will serve as the next-generation space surveillance radar system. This nearly $1 billion contract covers the cost of the engineering, manufacturing and development, production, and deployment for the S-band space fence. The currently planned initial operational capability date for the installation is in 2018 (Lockheed Martin Lands 2014).

When deployed the S-band ground-based radars will be designed to detect, track, and measure objects in space, most especially in low Earth orbit, although it will be able to track larger objects in higher orbits as well. The new radars for the space fence program will be able to detect much smaller microsatellites and debris than current systems and speed up detection of possible threats to GPS satellites or the International Space Station or communications and surveillance satellites.

The geographic separation along the Kwajalein Atoll plus the higher wave frequency of the new radar system will allow for the detection of much smaller microsatellites and debris than current systems – down to the size of a marble. It will significantly improve the timeliness with which operators can detect potential threats to GPS satellites or the International Space Station or other space assets and infrastructure (Space Fence 2015).

SPACETRACK, in addition to its space situational awareness and debris tracking and collision avoidance capabilities, has a clear military-related function. In particular, SPACETRACK has been assigned the responsibility to develop the systems interfaces necessary for the command and control, targeting, and damage assessment associated with any potential future US antisatellite weapon (ASAT) system capability. Part of SPACETRACK's capabilities is an Image Information Processing Center and Supercomputing Facility at the Air Force Maui Optical Station (AMOS).

Currently, information from the SSN is provided to governmental operators under a memorandum of understanding (MOU) to aid with the avoidance of orbital collision. The sensitivity of the SSN and SPACETRACK's military- and defense-related function was one of the factors that led to the creation of the Space Data Association (SDA) that functions from the Isle of Man and supports satellite operators from around the world. The prime capability of the SDA is to anticipate and help prevent conjunctions of GEO-based operational satellites by operators sharing data about the orbital locations of their satellites. Nevertheless, it is also increasing its capabilities to anticipate potential conjunctions in other orbits as well.

The Space Data Association and the Analytic Graphics, Inc. Tracking Network

The SDA was formed in 2009 by Inmarsat, Intelsat, and SES to share data. In April 2010, Analytical Graphics, Inc. (AGI) won the contract to design and operate the Space Data Center, SDA's automated space situational awareness system designed to reduce the risks of on-orbit collisions and radio-frequency interference. Initial Space Data Center operations began in July, and full capabilities were online April 2011. The current database is constructed using AGI's commercial software and is increasingly more capable with data being fed into the system by the member organizations of the Space Data Association (SDA). The data center then provides SDA members networked access to operational capabilities through a service-oriented architecture. The Space Data Center automatically ingests and processes operator orbital data, performs conjunction assessments and generates automated warning alerts, and supports avoidance maneuver planning and efficient RFI mitigation. It is an enhancement of AGI's SOCRATES-GEO/LEO system (Space Data Association).

Analytical Graphics, Inc. is now capable of tracking thousands of space objects with its Commercial Space Operations Center, or ComSpOC, which relies on optical and radio tracking assets and the company's own space surveillance software. AGI is building a catalog of space objects that it calls the SpaceBook (AGI-Lockheed):

Lockheed, in observing AGI's success in providing key services to SDA on an on-going commercial basis, has recently begun branching out to develop optical tracking capabilities in partnership with Australia' Electro Optic Systems.

Lockheed Martin and Optical Tracking Network

Lockheed Martin Space Systems, working with Australia's Electro Optic Systems, announced on August 25, 2014, that it is planning a new space object tracking site in Western Australia and hopes to sell the data to the US and Australian governments. http://spacenews.com/41727agi-lockheed-tout-commercial-space-surveillance-systems/#sthash.5CzBha2C.dpuf

In announcing the new agreement with Electro Optic Systems of Australia, Lockheed Martin clearly indicated that it was, in fact, entering the international space situational business. The official announcement said: "Through this agreement with Electro Optic Systems, we'll offer customers a clearer picture of the objects that could endanger their satellites, and do so with great precision and cost-effectiveness." The announced specific objective in using the EOS capabilities in Australia will be to zoom in on specific pieces of debris and determine their content, spin direction, and orbital speed. The data will be used to determine how much of a threat a given piece of debris poses to operating satellites.

Fig. 2 US Air Force space situational tracking satellite (Graphic courtesy USAF)

The use of optical sensors to track space debris is becoming more and more common. Currently, the ComSpOC has 20 optical sensors and three radio-frequency sensors in operation. ExoAnalytic Solutions of Mission Viejo, California, and the Las Cumbres Observatory Global Telescope Network of Goleta, California, are providing optical sensors. Rincon Research Corp. of Tucson, Arizona, provides data from radio-frequency radar (AGI-Lockheed).

This optical capability is currently seen by the US Space Surveillance Network officials as a useful complement to the S-band space tracking radar being located on the Kwajalein Atoll in the Pacific Ocean, near the equator. It has further been indicated the future planning might lead to a possible second space fence site to be located in Western Australia.

Nor is the use of ground-based radar and optical telescopes the only available tools. The US Air Force has also launched tracking satellites into orbit to augment its capabilities on the ground. One of these satellites used for space situational awareness, including the tracking of potential missile threats, is the US Air Force Satellite shown in Fig. 2.

Thus, to summarize the situation about current and future space debris tracking capabilities, the following is generally the case. Space Surveillance Networks are largely limited to larger objects, typically greater than 10 cm in low Earth orbits and greater than 1 m at geosynchronous altitudes. These sensitivity thresholds are by and large a compromise between system cost and performance.

Knowledge of the meteoroid and space debris environment at sub-catalog sizes has up to this point been “calculated” in a statistical manner through experimental sensors with higher sensitivities. This will change when the S-band radar space fence becomes operational in 2018. This new capability is expected to be able to track and catalog perhaps 250,000 elements of space debris in low Earth orbit down to 1 cm – or about the size of a marble. Ground-based optical telescopes can generally detect GEO debris down to 10 cm in size, while in situ impact detectors

(detectors flying onboard spacecraft) can sense objects down to a few micrometers in size. And while telescopes are perhaps best suited for GEO and high-altitude debris observations, radars are advantageous in the low Earth orbit (LEO) regime, below 2000 km.

Next Steps in the US Laser Ranging Capabilities for Debris Tracking

Laser researchers at NASA's Goddard Space Flight Center in Greenbelt, Maryland, are currently developing a precise "laser ranging" method to define and track orbital debris more accurately. This high-resolution telescope has the ability to actively determine high-velocity debris orbits by constant ranging. This activity is parallel to the tracking carried out by the smaller Electro Optic Systems telescope in Australia. This would overcome the current difficulties associated with passive optical and radar techniques and provide much greater accuracy information about orbital speeds and other similar data.

The current research is being carried out using the Goddard's Geophysical and Astronomical Observatory (GGAO) which is a 1.2 m (48 in.) with very high resolution.

This device that can transmit outgoing and receive incoming laser beams has been used to provide on-orbit calibration of some of Goddard's spacecraft. This device that has in the past actually obtained the precise distance and velocity and orbits of satellites as far away as an exploratory missions to Mercury can provide much greater accuracy in tracking Earth debris (Goddard NASA Team).

ESA and German and European Tracking Activities EISCAT

http://www.esa.int/Our_Activities/Operations/Space_Debris/Scanning_observing

There are other capabilities for tracking and cataloging space debris around the world. Some of the key capabilities are those operated via the European Space Agency. ESA radar tracking capability monitors debris increased after the Chinese ASAT test discussed above. These capabilities include the Tracking and Imaging Radar (TIRA) system near Bonn, Germany, operated by the Institute for High Frequency Physics and Radar Techniques that utilizes bistatic radar scanning.

TIRA debris detection system uses a 34 m dish antenna operating in L-band (i.e., 1.333 GHz) with a 0.45° beam width, at 1 MW peak power. Apart from tracking campaigns, the radar also conducts regular "beam park" experiments, where the radar beam is pointed in a fixed direction for 24 h, so that the beam scans 360° in a narrow strip on the celestial sphere, during a full Earth rotation. With the 24 h data collection, TIRA can detect debris data and determine coarse orbit information for objects of diameters down to 2 cm at 1000 km range. In a bistatic mode, together with the 100 m receiver antenna of the nearby Effelsberg radio telescope, the overall sensitivity increases down to almost 1 cm objects. A special seven-horn receiver, developed for the Effelsberg radio telescope, allows better resolution of object passages, permitting a reliable assessment of the object's radar cross section.

In addition to the German facility near Bonn, there is also the facility in Tromsø, Norway, known as the EISCAT Scientific Association (European Incoherent Scatter Radar). This facility operates a 930-MHz UHF radar and a 225-MHz VHF radar. In addition, there is a 500-MHz radar system that consists of a steerable 32 m dish and a fixed 42 m dish nearby.

The primary mission of the EISCAT network is to perform ionospheric measurements. However, following the development of a dedicated space debris computer to run at the back end of the processing units, these radars are now capable of statistical observations of LEO debris down to 2 cm that can be accomplished without diminishing the main EISCAT objectives.

European studies of orbital debris as addressed in the previous chapter suggest that Dr. Kessler's assessment of a debris collision every 10 years may be optimistic, and perhaps the result is more likely to be a collision every 5 years. The European Commission and ESA research also emphasize that about 10–15 large objects or about 7 t of debris need to be removed from space a year to reduce the risk of major collisions and damage to other spacecraft. An object larger than 1 cm hitting a satellite with sufficient relative velocity would likely damage or destroy key subsystems or instruments on board, and a collision with an object larger than 10 cm would destroy the satellite, according to European Commission study findings. The degree of the problem and its urgency is that the study estimates that objects larger than 1 cm are now calculated to reach around one million in 2020 (Sixth European Conference on Space Debris 2014).

Space debris consists of human-made objects in Earth's orbit that no longer has a useful purpose, such as pieces of launched spacecraft. It is estimated that up to 600,000 objects larger than 1 cm and at least 16,000 larger than 10 cm orbit the Earth.

Japanese Initiatives

Japan has also developed tracking capabilities using radar and optical technology but has recently developed, at the RIKEN research institute, a detailed proposal to use lasers in orbit. This would allow not only to track debris with some precision but also to use higher-power laser systems to help delete space debris from orbit. In the past it has been suggested that land-based lasers could do this, but Japan has suggested that they might place a test-case laser experiment on the International Space Station to examine how well this works from space.

The concept would be to combine a superwide field-of-view optical telescope to detect objects plus a recently developed high-efficiency laser system, known as CAN, that could track space debris and remove it from orbit.

The RIKEN-developed telescope could be used to find debris with high precision. It was originally planned to detect ultraviolet light emitted from air showers produced by ultrahigh energy cosmic rays entering the atmosphere at night.

The proposal is simply to adapt it to the new mission of detecting high-velocity debris in orbit near the ISS.

The CAN laser was originally developed to power particle accelerators. It consists of bundles of optical fibers that act in concert to efficiently produce powerful laser pulses. Combining these two instruments could, in theory, be used to locate and to help with deorbiting dangerous space debris in potentially destructive orbits. This system would work only for debris elements, around the size of one centimeter. The intense laser beam focused on the debris will produce ablations. The result would be to reduce the debris' orbital velocity, leading to its reentry into the Earth's atmosphere. The proof of concept proposal would involve a 20 cm telescope and a laser wrapped by 100 fibers. The full-scale operational model would have a 3 m telescope and be wrapped by10,000 optical fibers. The focus of the operational model would be on eliminating debris from the polar orbit at around 800 km altitude over about a 5-year period (Scientists want).

Active Removal of Orbital Debris

There are more than a dozen concepts about how space debris might either be actively removed from Earth orbit or its orbital affected so that the debris element would be removed more quickly from orbit due to accelerated decay. Increased solar activity during solar max also assists with debris removal. Satellites in orbit below 300 km tend to decay due to natural effects of gravity and atmospheric drag. The largest problems involve large defunct elements in low Earth orbit (especially around 800 km and especially in polar, sun-synchronous orbit).

There are many concepts that have been put forward about active debris removal, and these are discussed in several books, but today there are no systems actually carrying out such missions not only because of high cost and technical difficulty but also because of issues relating to the liability convention that actually creates disincentives to remove debris because of the most legal implications if the removal is unsuccessful (Pelton 2015).

Conclusion

The space situational tracking capabilities to determine space debris orbits have continued to be upgraded in the past 10 years. The S-band radar space fence when deployed in 2018 coupled with capabilities such as the Analytic Graphics, Inc./ Space Data Association tracking system, the new Lockheed Martin and Electro Optic Systems commercial tracking networks, and other capabilities around the world such as that of ESA, JAXA, and other space agencies brings a high degree of sophistication to being able to track space debris and detect collisions and potential satellite conjunctions in advance so that evasive action can be taken. This is the good news.

The less favorable news is the degree to which orbital debris continues to increase. The fact that even 1 cm debris elements can do enormous damage when collisions occur at very high relative velocities is quite sobering when it is realized that there are over 250 K of these elements today and that they could number a million by year-end 2020. The cascading effect is currently generating more and more debris elements that are quite hard to remove especially using the various techniques that have been suggested to help remove large debris elements. The proposal of the Japanese RIKIN Institute to use laser ablation may be the only cost-effective approach for addressing the smaller and very numerous debris elements. The low Earth orbit is truly becoming congested to the extent that it poses real problems with regard to the long-term sustainability of long-term use of outer space for all types of applications including telecommunications, space navigation, remote sensing, and meteorological satellites:

- There are clearly a number of actions that are needed to address this problem. The longer action is delayed, the more expensive, challenging, and difficult the problem becomes. The priority course of action would seem to include the following:
- Continuing efforts to deploy governmental and private tracking systems using radar, optical tracking, active optical ranging, and space-based tracking to be able to know the orbits of all elements in space – active and defunct. This is key to being able to undertake evasive actions.
- Undertake to remove as soon as possible the largest defunct debris elements in the critical low Earth orbits/polar orbits since this is the most dangerous situation that risks generating major showers of new debris.
- Make sure that all new satellites are equipped to meet the now well-established international guideline that all space objects are to be removed from orbit within 25 years of their end of life.
- Continue to carry out research as to new technology and systems that might be utilized to remove debris in the most efficient manner possible.
- The industry-created Space Data Association (SDA) should continue all of its various efforts to address this problem and how industry practices can help alleviate the space debris problem and operational vigilance that operating satellites do not collide.
- Consider various ideas and proposals that have been suggested to address the space debris problem. These include such ideas as (i) requiring a separate debris removal capability to go on all new satellites that would not be under the control of the satellite operator but a separate entity that would control ultimate deorbit and (ii) considering that all satellite launches, in addition to launch insurance, should pay into a debris removal fund so that resources would be created to clean up space over time.

The Inter-Agency Space Debris Committee (IADC) has proved a very key resource to address the space debris problem and greatly assisted the UN Committee on the Peaceful Uses of Outer Space (COPUOS) on this issue. It should continue to assist COPUOS and its Working Group on the Long-Term Sustainability of Outer Space Activities that continue to address this now chronic problem.

Cross-References

► Orbital Debris and Sustainability of Space Operations

References

AGI-Lockheed Tout Commmercial Space Surveillance Systems, http://spacenews.com/41727agi-lockheed-tout-commercial-space-surveillance-systems/#sthash.5CzBha2C.dpuf. Last Accessed 16 Jan 2016

M. Cooney, 10 crucial issues around controlling orbital debris, directing space traffic. Network World, 12 May 2014, http://www.networkworld.com/article/2226898/security/10-crucial-issues-around-controlling-orbital-debris-directing-space-traffic.html. Last Accessed 16 Jan 2016

Goddard NASA Team, http://www.nasa.gov/content/goddard/nasa-team-proposes-to-use-laser-to-track-orbital-debris/Last. Accessed 16 Jan 2016

L. Martin, Lands $914 Million Space Fence Contract (2014), http://spacenews.com/40776lockheed-martin-lands-914m-space-fence-contract/. Last Accessed 16 Jan 2016

J.N. Pelton, *New Solutions to Orbital Debris Problems* (Springer Press, New York, 2015)

Scientists want to blast space debris with a space station-mounted laser, http://www.networkworld.com/article/2914663/security0/scientists-want-to-blast-space-debris-with-a-space-station-mounted-laser.html. Last Accessed 23 Jan 2016

Space Data Association, http://www.space-data.org/sda/join-sda/. Last Accessed 16 Jan 2016

Space Fence (2015), http://www.lockheedmartin.com/us/products/space-fence.html. Last Accessed 16 Jan 2016

Sixth European Conference on Space Debris, European Satellite Agency (2014), http://www.congrexprojects.com/2013-events/13a09/introduction. Last Accessed 23 Jan 2015

U.S. Space Surveillance Network (SSN), https://en.wikipedia.org/wiki/United_States_Space_Surveillance_Network. Last Accessed 16 Jan 2016

Space Weather and Hazards to Application Satellites

Michael J. Rycroft

Contents

The answer my friend is blowin' in the wind. (A line from a song by Bob Dylan 1962)

M.J. Rycroft (✉)
Cambridge Atmospheric, Environmental and Space Activities and Research (CAESAR)
Consultancy, Cambridge CB3 9HW, UK
e-mail: michaelrycroft@btinternet.com

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_78

Abstract

Magnetic activity on the Sun causes disturbances moving out through interplanetary space as a stronger than usual solar wind. Enhanced solar activity can also produce large fluxes of penetrating energetic protons of great destructive potential. When either of these phenomena arrives near the Earth and strikes a satellite or spacecraft, damage is likely to be done, either directly and indirectly. Here, we give an overview of the several different effects which occur in low Earth orbit (LEO), in medium Earth orbit (MEO), and in geostationary orbit (GEO) as well as in high inclination elliptical Earth orbits where positioning satellites are located.

Keywords

Atomic oxygen • Coronal mass ejections (CMEs) • Galactic cosmic rays • Ionosphere • Ionospheric scintillations • “Killer” electrons • Magnetic storm • Plasma • Radio communications • Satellite drag • Satellite operations • Single event upsets • Solar activity • Solar cycle • Solar protons • Space debris • Space environment • Space weather • Sunspots • Van Allen radiation belts

Introduction

“Space weather” is defined by the US National Space Weather Program as referring to “conditions on the Sun and in the solar wind, magnetosphere, ionosphere, and thermosphere that can influence the performance and reliability of spaceborne and ground-based technological systems and can endanger human life or health.” From the viewpoint of an application satellite, the space environment is a hostile, harsh, and hazardous environment. It is thus a significant challenge to design application satellites so that they can operate reliably throughout a significant in-orbit lifetime, without the possibility of benefiting from physical repair.

Valuable background information on the science lying behind the subject is to be found in a chapter by Rycroft (2010), in the recent book of Schrijver and Siscoe (2009), and at the websites http://www.vsp.ucar.edu/Heliophysics/ and http://www-spof.gsfc.nasa.gov. There are several books at different levels on space weather, such as those by Freeman (2001), Song et al. (2001), Carlowicz and Lopez (2002), Daglis (2001, 2004), Bothmer and Daglis (2007), Lilenstein (2007), and Moldwin (2008). Singh et al. (2010) have recently written a useful overview paper; in addition, useful space weather websites are indicated in the endnotes.

First, we consider features on the Sun and solar activity as a driver of space weather phenomena. Then, we discuss the five key features of the space environment which impact the design of sensors, instruments, electronic, and all other equipment aboard all space vehicles, and which influence the performance of satellites or spacecraft in different orbits. Since the space age began, numerous measurements have been made of important physical parameters of the space environment; from these, useful empirical (numerical) models of variations of those parameters have been constructed. Such valuable models are complemented by theoretical models, that is to say models which are based upon our understanding of the physical processes operating. (For numerical values, we shall use here the International System of units [SI] based upon the meter [m], kilogram [kg], second [s], and Ampere [A]).

The Sun, Our Star

Energy from the Sun

Thermonuclear reactions occur within the Sun's core. The tremendous amount of energy released allows the Sun to provide essentially all the energy which the Earth receives from the cosmos. It is this energy which, directly or indirectly, powers all phenomena occurring on Earth and in its environs. It does this in the form of radiant energy (see chapter "▶ Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing" in this volume) or in a form of energy derived from this source. Although the flux of this radiant energy peaks in the visible part of the spectrum to which our eyes are tuned, much energy is emitted as ultraviolet, X-ray, and gamma-ray – i.e., high energy – photons, plus longer-wavelength radiation at infrared and radio wavelengths.

The Sun's spectrum is approximately that of a black body at 5,800 K; this is the temperature of the solar "surface," a plasma termed the "photosphere." Just above the photosphere is the chromosphere, a cooler region. The outer solar atmosphere, at heights exceeding a fraction of a solar radius, known as the corona, emits X-rays because its temperature is about one million K. As the solar wind, the Sun's outer atmosphere flows away from the Sun at speeds of several hundred kilometers per second.

Sunspots and the 11-Year Solar Cycle

From time to time, dark splodges appear on the photosphere; these are called sunspots. Two photographs of the Sun are shown in Fig. 1, taken on September 27, 2001 (on the right), and 2008 (on the left). The right hand image shows sunspots, generally occurring in pairs, at solar maximum conditions; the left hand image shows no visible sunspots during a very deep solar minimum. Sunspots are regions of intense magnetic fields, with complex (twisted) arched magnetic field lines above the surface connecting each pair of sunspots.

It is observed that as the Sun rotates (once in 27 days on average, but in approximately 25 days at the Sun's equator and about 32 days nearer the poles), the polarity of the leading sunspot is always in the same direction, for example outward, and that of the following sunspot is always in the other direction, or inward. This applies to all sunspots in one hemisphere of the Sun, with the situation being exactly the opposite for sunspots in the opposite hemisphere. All sunspots migrate

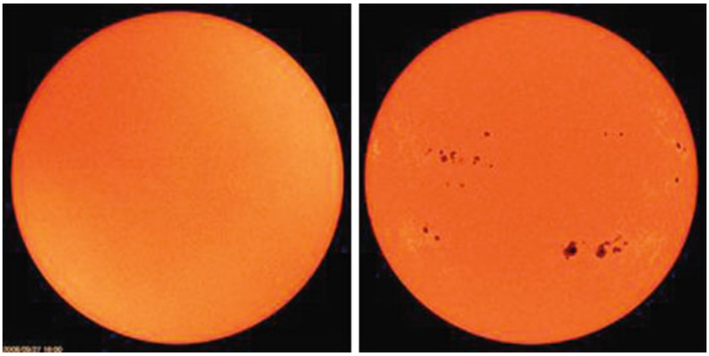


Fig. 1 Sometimes, the Sun has sunspots on its visible surface (Courtesy of NASA Goddard Space Flight Center (October 7, 2008). Spotless Sun: Blankest year of the space age. *Science Daily.* Retrieved June 28, 2011, from http://www.sciencedaily.com/releases/2008/10/081006184638.htm)

from high latitudes toward the Sun's equator on a timescale of 11 years. Then new sunspots appear at high latitudes, with the leading sunspots having the opposite polarity; this situation marks the beginning of each new 11-year solar cycle.

The positions of sunspots since the year 1874 are displayed in the so-called Maunder butterfly diagram shown in the top panel of Fig. 2. The lower panel gives the sunspot area as a percentage of the Sun's visible disk, with a typical value of up to 0.2 %. The numbers at the bottom give the number of the solar cycle; now the Sun is beginning the new solar cycle 24, after a very deep and long solar minimum in 2008. Figure 3 presents a graph of the yearly average number of sunspots over the last 400 years which also exhibits the 11 ± 1 year cycle since 1750; before that time the number of sunspots was very small, during the Maunder minimum. The sunspot cycle has a faster rise to solar maximum and a slower decline from it (Hathaway and Wilson 2004). The sunspot number predicted for the maximum of solar cycle 24 in 2013 is like that shown by the arrow. The maximum sunspot number predicted for solar cycle 25 is even smaller. Thus the Sun's activity also varies on a timescale of a hundred years and more. The intensity of radio emission from the Sun at a wavelength of 10.7 cm, which can be received at the Earth's surface by a radio telescope, also varies markedly over 11 years; this quantity (known as F10.7) is often used as a proxy measure of solar activity.

The Sun's activity varies over this 11-year cycle. The brightness of the Sun, its irradiance, is ~0.1 % larger at solar maximum than at solar minimum even though sunspots absorb some light emitted below them. The solar flux variation over the solar cycle is approximately 10 % in the ultraviolet but varies by 10–100 times at X-ray wavelengths. Between these bands, at a wavelength of 28.4 nm, radiation is emitted by multiply ionized iron atoms (Fe XV). Figure 4 shows a sequence of 11 images of the Sun taken from the SOHO satellite in the extreme UV part of the spectrum at yearly intervals from 1996 to 2006; this composite figure shows clear evidence of the 11-year cycle of solar activity, with solar activity maximizing around the year 2001.

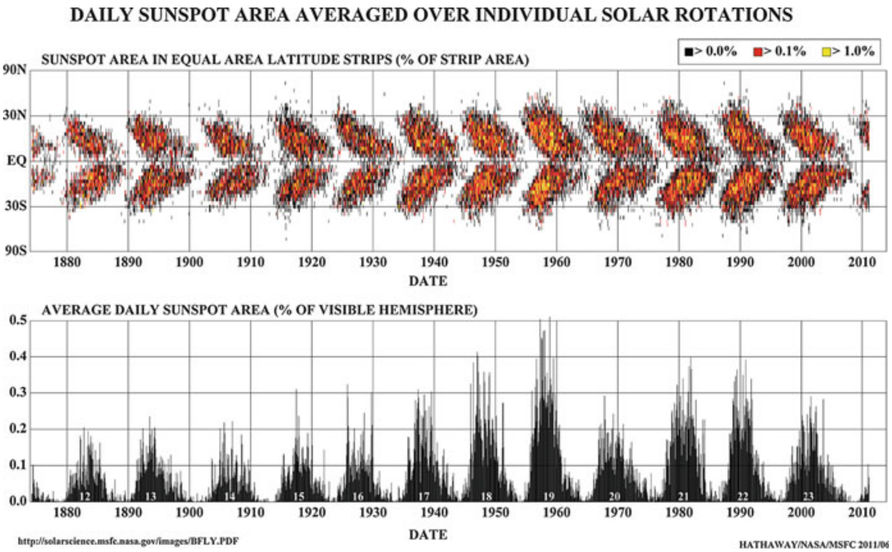


Fig. 2 The positions of sunspots (*upper panel*) and the area of the Sun covered by sunspots (*lower panel*) plotted since the year 1874 show the 11-year cycle of solar activity (Courtesy of NASA, http://solarscience.msfc.nasa.gov/images/BFLY.PDF)

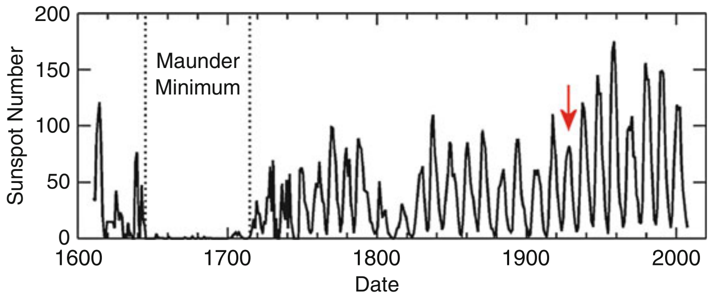


Fig. 3 The number of sunspots plotted since the year 1600; the 11-year solar cycle is clear since the early 1700s (Courtesy of NASA, http://science.nasa.gov/science-news/science-at-nasa/2009/29may_noaaprediction/)

On September 1, 1859, Carrington was observing sunspots when one sunspot region suddenly brightened noticeably; this was the very first observation of a solar flare. The event occurred in the middle of a large geomagnetic storm which intensified a day later. That was the first evidence of a Sun–Earth connection, which led to the new research field now known as solar-terrestrial physics. As luck would have it, Carrington observed the largest solar flare ever; no solar flare observed since has been brighter. It is now believed that a solar flare happens as a result of a plasma instability which reconfigures the arched magnetic field lines in the low solar corona.

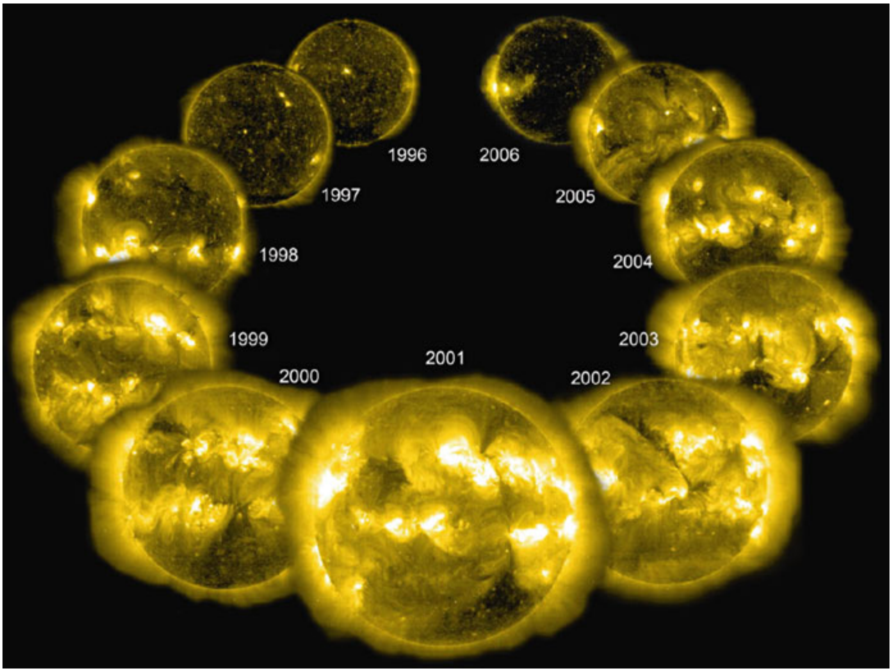


Fig. 4 Extreme ultraviolet images of the Sun (Courtesy of ESA and NASA, http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=47710)

This injects accelerated energetic electrons down to the chromosphere where they collide with atoms and emit X-rays. Nowadays, these X-rays are routinely detected by instruments aboard US geostationary satellites.

Coronal Mass Ejections (CMEs)

Associated with the magnetic field reconfiguration near the Sun is the ejection of hot, dense plasma into the corona and beyond, into interplanetary space. Such a huge (up to approximately 10^{12} kg) high speed solar wind event carrying strong, twisted magnetic fields away from the Sun is termed a coronal mass ejection (CME) event. Figure 5 is an image taken by the Large Angle and Spectrometric Coronagraph (LASCO) C2 instrument aboard the NASA/ESA Solar and Heliospheric Observatory (SOHO) spacecraft. In these images, direct sunlight is blocked by an occulting disk, so that only light emitted by the solar corona at heights greater than three solar radii is detected, in order to reveal structure in the high corona; the approximate size of the Sun is shown by the white circle in Fig. 5.

Dark areas in Fig. 5 represent coronal holes, where the low density solar wind is moving most rapidly away from the Sun along radial interplanetary magnetic field lines. The bright area at the bottom shows an arched magnetic field structure from the

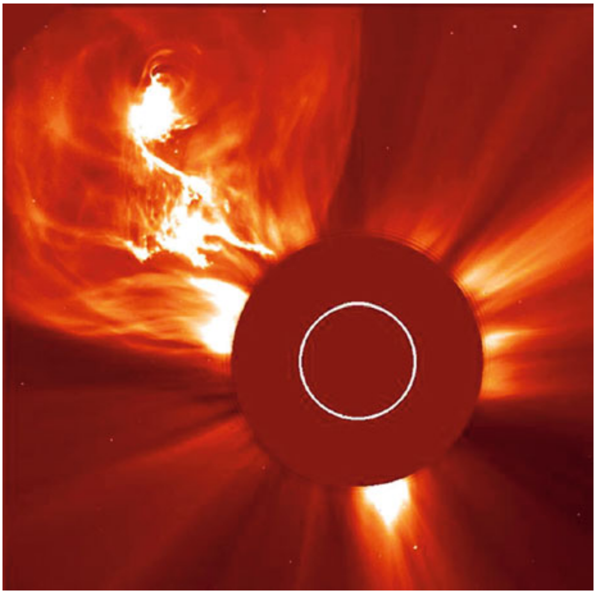


Fig. 5 A dramatic coronal mass ejection (CME) event was produced by the Sun on January 4, 2002 (Courtesy of ESA and NASA, http://sohowww.nascom.nasa.gov/gallery/images/c2fireball.html)

Sun's surface extending out to about four solar radii and containing very hot plasma of high density. A more complex arched magnetic field region at the top left has exploded, creating an enormous volume of very hot, dense plasma ballooning out into interplanetary space and evolving as it does so. This is a dramatic example of a coronal mass ejection (CME) event.

When such a CME is directed toward the Earth, it is called a halo CME; it compresses the Earth's magnetosphere, within which the Earth's dipolar magnetic field is contained. Thus the magnetopause, the boundary of the magnetosphere (which is usually at about 10 Earth radii (R_E) upstream of the Earth on the sunward side) is compressed and moves inward to about 6 R_E. Then satellites in GEO – at a geocentric distance of 6.6 R_E – under noon time conditions will be directly exposed to the solar wind plasma. They have then to operate in a totally different plasma environment from the one for which they were originally designed to function.

This situation creates a geomagnetic storm (or, simply, a magnetic storm). Today, it is said that solar activity creating a magnetic storm at the Earth is particularly "geoeffective." A CME takes from one to several days to travel from the Sun to the Earth; a CME having a faster solar wind speed generally produces a stronger magnetic storm. Magnetic storms are generally most prevalent in the declining phase of the solar cycle.

Sometimes, solar flares and CMEs occur together. At other times there can be a solar flare without a CME being generated, and sometimes there can be a CME without a solar flare. Large X-ray flares and large CMEs sometimes, but not always, occur simultaneously. This complicated situation makes the forecasting of space weather events (Bothmer and Daglis 2007) in the Earth's environment very difficult. However, space weather forecasting is an important subject because of the potential damage not only to space assets but also to long electrical conductors on the ground, such as oil and natural gas pipelines, and also to ground-based electrical power grid systems. Pipelines can suffer increased corrosion and electrical grid systems can be disrupted by induced currents caused by space weather disturbances, which also create bright auroral displays at high latitudes. On the occasion of especially intensive events these displays are seen at much lower latitudes, as occurred during the so-called Carrington event in 1859.

The Earth's Atmosphere and Near-Space Environment

Where solar energy is absorbed, the temperature of that material thereby increases. Most of the Sun's visible radiation is absorbed by the Earth's surface, and the air at the surface is thus warmed. Ultraviolet radiation is absorbed by oxygen and ozone in the stratosphere, at heights from about 15–50 km, and the stratosphere is heated. As illustrated in Fig. 6, solar X-rays are absorbed in the thermosphere, at altitudes above 100 km; in the process, the gaseous molecules or atoms are split into positive ions and electrons, forming an electrically charged gas termed the "ionosphere." Radiation from the Sun (at the top) in different wave bands, shown in the center of Fig. 6, is absorbed at different levels of the atmosphere. On the right, typical height profiles

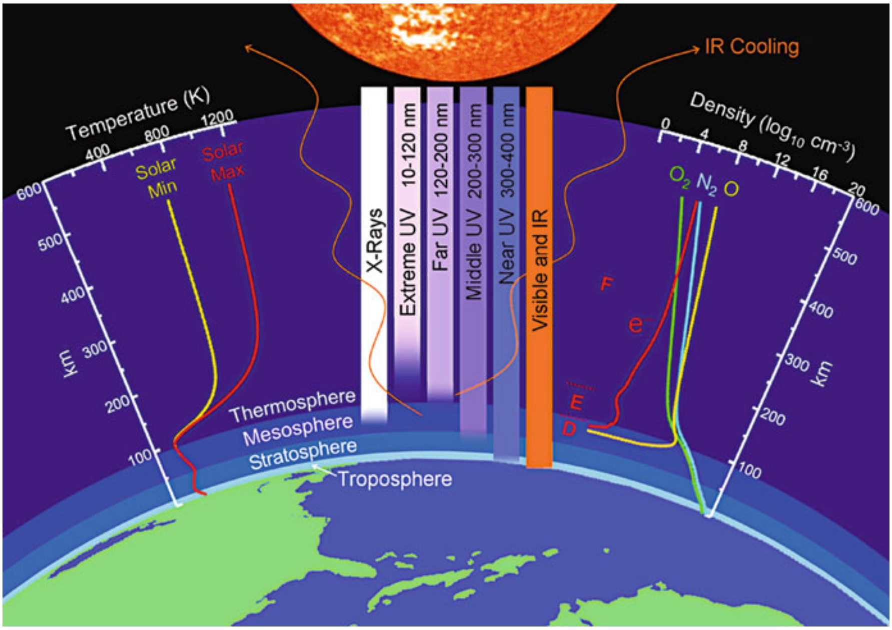


Fig. 6 Solar radiation of different wavelengths is absorbed at different levels in the atmosphere (Courtesy of NASA, http://science.nasa.gov/science-news/science-at-nasa/2010/15jul_thermosphere/)

of the density of oxygen molecules (green curve) and of oxygen atoms (yellow) are plotted; the ionospheric electron density profile (e^-) is shown in red. On the left is shown the neutral gas temperature for near solar minimum (yellow) and maximum (red) conditions.

In the upper atmosphere at 100 km altitude, there is only one ion-electron pair for every 10^8 electrically neutral molecules, whereas at ~300 km altitude for every ion-electron pair there are about a thousand neutrals. Thus, everywhere, the Earth's ionosphere is a weakly ionized plasma, as is evident from the right hand side of Fig. 6.

It is still uncertain whether the 0.1 % solar irradiance variation over the solar cycle has any effect on varying the Earth's weather and climate on a decadal timescale. However, it is clear that the temperature of the stratosphere varies a little over the 11-year cycle, and other stratospheric features have long-term variations. The temperature of the thermosphere varies markedly from solar minimum to solar maximum, and so do many ionospheric parameters, the most important of which is the maximum electron density in the ionosphere. This feature occurs near 300 km altitude at what is called the peak of the F-region.

Five Key Aspects of the Space Environment

Tribble (2003, 2010) has pointed out the five important features of the space environment as they affect most of the 800 or so active space missions today. We consider these features in some detail here.

The Residual Atmosphere in Low Earth Orbit (LEO): Near-Vacuum Conditions

The Earth's atmosphere (having a total mass of 5×10^{18} kg) is kept close to the surface by the gravitational force of attraction on the molecules of gas in the atmosphere (mainly nitrogen, with oxygen and some minor species including argon, water vapor, carbon dioxide, and methane). This gravitational force arising from the Earth's mass (6×10^{24} kg) is often called the force of gravity. The gravitational force acting on an object of mass m kg is 9.8 mN, with $g = 9.8$ m/s^2 being the acceleration due to gravity at the Earth's surface. The mean global temperature of the atmosphere at the Earth's surface is 15 °C (or centigrade), which translates to 288 K (Kelvin, or absolute).

The pressure (1,013 hPa, or 1.013 bar, at the Earth's surface) and the density (1.3 kg/m^3) both decrease exponentially with increasing altitude. At 15 km altitude, the pressure and density have fallen to about one tenth of its sea level value, and at 30 km to one hundredth. At about 100 km altitude, where, conventionally, space is said to begin, there still remains about one millionth of the Earth's entire atmosphere. This is the pressure attained in quite a good vacuum system, and so conditions above 100 km altitude (the base of the thermosphere) are near-vacuum conditions.

At altitudes above 120 km, the temperature can be as low as 500 K under solar minimum conditions and up to 2,000 K at solar maximum, as indicated in Fig. 6.

The density of the atmosphere at 500 km altitude is approximately 10^{-13} kg/m^3 at solar minimum and up to a hundred times bigger, i.e., 10^{-11} kg/m^3, at solar maximum. The thermosphere expands and contracts over the Sun's 11-year cycle of activity, causing the density at a particular altitude to change markedly. There is also a long-term change, possibly associated with increasing amounts of carbon dioxide in the troposphere (global warming – climate change – primarily due to the increasing rate of burning fossil fuels). The density at 400 km altitude at the solar minimum in 1997 was about 10 % smaller than it was around 1986, and that was 10 % less than 11 years earlier. During the pronounced solar minimum of 2008, the density at 400 km altitude was unexpectedly low, about 28 % lower than in 1997.

Thermal Radiation: From the Sun and the Earth

In space there is no oxygen and there is no ozone to absorb the strong dose of ultraviolet radiation coming from the Sun. Therefore, a satellite's surface is bombarded by an intense flux of ultraviolet radiation. This degrades the surface materials to a greater or lesser extent; it can cause photoelectrons to be emitted (see chapter "▶ Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing").

Direct thermal radiation from the Sun (typically 1,365 W/m^2), plus some solar radiation reflected back into space by clouds and by the Earth's surface, heat up a satellite's surface. There is an additional contribution to radiation that heats a satellite from the Earth-atmosphere system; this is terrestrial thermal infrared (IR) radiation with a spectral peak at wavelengths near 10 μm (10×10^{-6} m); its contribution is about 240 W/m^2. When designing a satellite's thermal control system, all these sources have to be accounted for. Within the satellite, heat is conducted from the sides that are illuminated to the dark side, from where it is radiated away into the blackness of space (at 2.7 K). There is no conduction of heat away from the satellite in the near-vacuum conditions of space. Surface materials with the desired absorptivity and emissivity characteristics are chosen so that the temperature inside the satellite lies within the correct operating temperature range.

A polar orbiting LEO satellite is eclipsed by the Earth for about 30 min during every 90 min orbit. This phenomenon exerts a strong thermal cycle on the satellite. A satellite in geosynchronous orbit (GEO) is eclipsed by the Earth for only about 45 min during the maximum eclipse period in the equinoxes; for most of the time there are no eclipses at all. The differential expansion of illuminated and dark satellite metallic surfaces is another effect that has to be considered.

Plasma: The Electrically Charged Gas, the Ionosphere

During the daytime, different wavelengths of solar extreme ultraviolet (EUV) and X-ray radiation cause the formation of different ionospheric layers at different heights (see Fig. 6). At heights above about 170 km, the ionospheric plasma moves upward to create a long lasting F-region peak. Above that the plasma density falls off with increasing height. During the night, the electrons collide with ions and recombine to form neutral molecules or atoms, the more so at lower heights. Because the flux of solar EUV and X-radiation is much larger at solar maximum than at solar minimum, there is a marked variation of ionospheric parameters over the solar cycle.

Energetic Charged Particles: Van Allen Radiation Belts, Solar Protons, and Cosmic Rays

In space we speak of radiation – that means not only electromagnetic radiation but also fluxes of energetic charged particles which constitute a radiation hazard to astronauts (especially if they are carrying out extra vehicular activities, EVAs) and to electronic equipment on satellites. The most energetic charged particles are galactic cosmic rays (GCRs) coming through interstellar space from beyond our Milky Way galaxy, ions with energies exceeding 1 GeV (10^9 eV). (Here, 1 eV is the energy gained by an electron when it is accelerated through a potential difference of 1 V; one electron Volt [1 eV] is equivalent to a thermal energy of $\sim$11,000 °K.) Solar energetic particle events, also termed solar proton events (SPEs), have individual ion energies of more than 1 MeV (10^6 eV); these energetic protons come from active regions on the Sun. During intense solar flares, the fluxes of these particles increase dramatically. Almost all (about 95 %) of the ions are protons, with some doubly charged helium ions and some highly charged ions of heavier elements.

At lower energies, 1 keV (10^3 eV) charged particles from the ionosphere and from the solar wind are accelerated up to about 1 MeV by complex plasma wave–particle interactions inside the Earth's magnetosphere to create the Van Allen radiation belt electrons and ions (being mainly – 95 % – protons). The Van Allen belts occur in two doughnut shaped regions around the Earth, as shown in Fig. 7; the strongest fluxes shown in red occur in the inner and outer zones. The actual fluxes of Van Allen charged particles can vary from typical values by up to two orders of magnitude, both up and down, from day to day.

The fluxes of energetic ions and electrons lost from the Van Allen radiation belts of the Earth's magnetosphere are largest in the vicinity of the South Atlantic geomagnetic anomaly. This is a region of reduced geomagnetic field strength which lies to the east of Brazil; it is shown in red in Fig. 8.

Micrometeoroids and Orbital Debris

Strictly speaking, this subject lies outside the topic of space weather; however, it is valuable to mention it here. The chapter on space debris provides additional useful information on this subject.

Very many tiny dust particles of the cometary material (from which the solar system was originally formed) continuously enter the upper atmosphere from space.

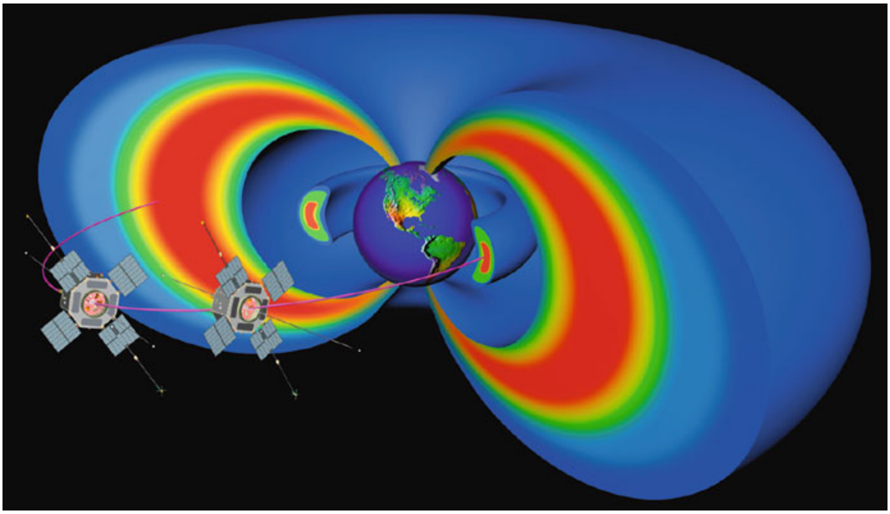


Fig. 7 The Radiation Belt Storm Probes (RBSP) were launched in 2012 by NASA to study the Van Allen Belts (Courtesy of NASA, http://rbsp.jhuapl.edu/gallery/artRender/pages/artRender_01.php)

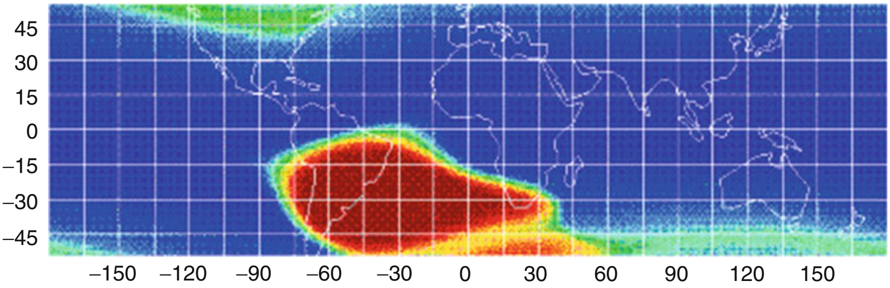


Fig. 8 The South Atlantic geomagnetic anomaly, shown in *red* (Courtesy of NASA, http://imagine.gsfc.nasa.gov/docs/ask_astro/answers/961004.html)

Known as micrometeoroids, the total mass of these primordial particles arriving per day at speeds of about 10 km/s or more is about a hundred metric tons (10^5 kg). The larger particles (each with a mass of a fraction of 1 g (10^{-3} kg) up to 1 g) typically burn up at approximately 100 km altitude, emitting a flash of light – this is the explanation for how a meteor (a "shooting star" or a "falling star") is created. Even larger particles (having masses more than 1 g and up to a significant fraction of 1 kg or even more) do not burn up but come down to the Earth's surface as meteorites.

This material adds to the significant amount of space debris (2,000 metric tons) already present in the thermosphere. Such debris largely consists of several thousand

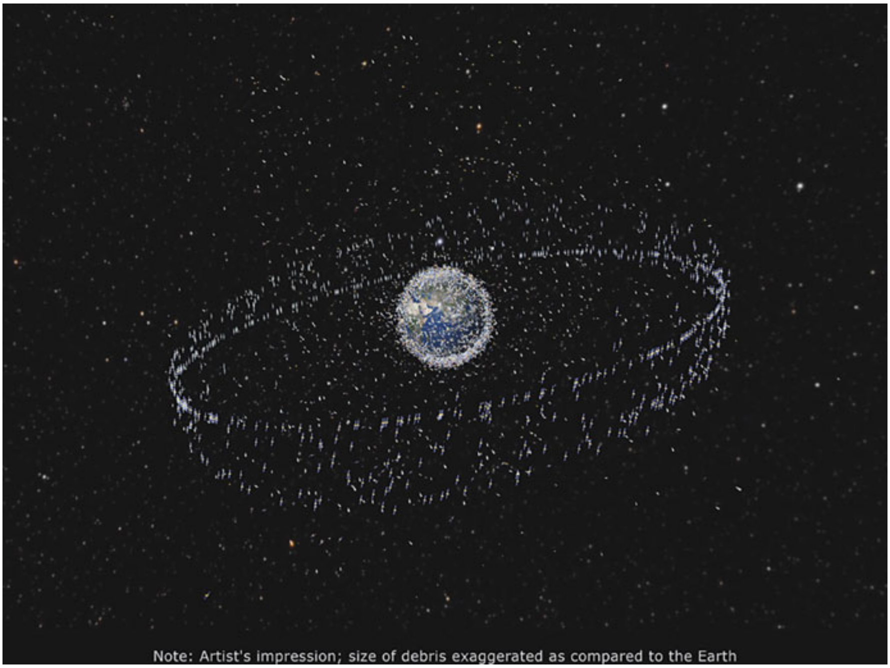


Fig. 9 Objects in orbit around the Earth which are tracked from the ground (Courtesy of ESA, http://www.esa.int/SPECIALS/ESOC/SEMN2VM5NDF_mg_1_s_b.html)

dead satellites and numerous spent upper stage rockets, plus various other parts and materials. Space debris has been created in situ by exploding rocket fuel tanks (due to the small amounts of fuel left in them), exploding electrical batteries or radioisotope thermoelectric generators, or by spent rockets or satellites colliding with other pieces of space debris; where the larger pieces of space debris are to be found is indicated in Fig. 9. Most of these objects are in LEO, a good number are in GEO, and some are in other orbits; the objects are exaggerated in size in order to make them visible in the figure.

How the Space Environment Affects Satellite Operations

Here we consider the many effects which space mission planners, space systems engineers, and space instrument designers must consider – and the several possible problems that must be resolved – before a space mission is attempted. Otherwise, premature failure may result, with considerable human efforts expended in vain and a sizable sum of money wasted.

Near-Vacuum Conditions

Even though the atmosphere is very thin in LEO, it still exerts a drag force on a satellite moving through it at high speed and on other particles in orbit. The magnitude of this force is proportional to the density of the air, the cross-sectional area which the satellite or particle presents, its drag coefficient (note that satellites are not normally as streamlined as sports cars are!), and the square of its orbital velocity, which is large, approximately 8 km/s. Under solar maximum conditions the drag force acting on a satellite at 500 km altitude is up to a hundred times larger than at solar minimum. For a satellite orbiting near 300 km altitude, the drag force is an order of magnitude larger than at solar minimum. The consequence of this drag is that the perigee of all LEO satellites is reduced more quickly near solar maximum, and so the lifetime of the mission is shorter than it would be under solar minimum conditions. For the International Space Station (ISS) orbiting at, say, 400 km altitude, the consequence is that at solar maximum much more propellant has to be transported to the ISS to boost the altitude of its orbit than at solar minimum.

A satellite surface glows when neutral atmospheric atoms and molecules impinge on its surface. This is because of its high velocity, about 8 km/s with respect to the gaseous atoms and molecules; on impact this gives the gas particles sufficient energy (about 5 eV) to react chemically with its surface and cause a glow in the visible part of the spectrum. This glow may interfere with the performance of an optical experiment, such as carried out by a telescope, if the beam of the instrument goes through the glowing region.

The major constituent of the air at a few hundred kilometers altitude is atomic oxygen, formed by the breakdown of molecular oxygen in the thermosphere by energetic solar ultraviolet and X-ray photons. The number density of these atoms can vary by a factor of 10 (i.e., the density lies between $10^{13}/\mathrm{m}^3$ and $10^{14}/\mathrm{m}^3$) near 400 km altitude. These oxygen atoms are highly reactive, chemically; for example, they will oxidize a front-silvered mirror, turning it black. Further, this action changes the thermal characteristics of the surface and even weakens it physically.

In space, because of the near-vacuum conditions, conventional lubricants do not work well at all. A moving joint, such as in the elbow action required to deploy an arm, can seize up. Therefore, special lubricants have been invented. Satellite and spacecraft surface outgas when first in space, that is to say particles and molecules of gas stuck to the surface are released into space. Thus, the local pressure around the satellite builds up during the first few days in orbit. Such detached particles (even though the satellites were put together in a clean room on Earth) can stick onto the surface of an optical instrument, say a mirror, or a lens, and degrade its performance below optimum. Even a partial mono-molecular layer can ruin the performance of a front-silvered mirror. The thrusters which are used to control the orientation of a satellite or spacecraft will introduce gas in the vicinity and may, similarly, reduce the performance of an optical instrument.

There are no known effects of the residual Earth's atmosphere on the performance of positioning satellites in inclined elliptical medium Earth orbits (MEO) or on communications satellites in MEO or GEO. Neither are there any such effects acting on spacecraft operating outside the Earth's magnetosphere.

Thermal Radiation

The surface materials covering the satellite of spacecraft may have their absorptivity properties and their emissivity characteristics changed by bombardment by energetic charged particles. Both these properties may also be altered by exposure to the strong flux of solar ultraviolet radiation.

Plasma

High frequency (HF, 3–30 MHz) radio communications under the ionosphere are strongly affected by SPE events, which lead to communications blackouts, especially in the polar regions. For radio communications between the ground and a satellite, and vice versa, the radio frequency used must exceed the maximum plasma frequency of the ionosphere. This is the F-region peak plasma frequency; its value is usually between 10 and 30 MHz, but it varies considerably with geographic location, time of day/night, phase of the solar cycle, geomagnetic activity, and several other quantities.

Small-scale spatial and temporal variations of the ionospheric plasma density cause scintillations, that is to say rapid amplitude and phase changes of radio signals received on the ground from spacecraft/satellites. These are most marked in equatorial and auroral regions, especially under conditions of enhanced solar and geomagnetic activity. Such effects require corrections to be made to GPS navigation satellite signal (1–2 GHz, L band) delays, and to spaceborne radar altimeter observations. The fact that the ionosphere is a birefringent medium has to be considered; this is to say that there are two values for the refractive index at each radio frequency, due to the fact that two wave modes (ordinary (O) and extraordinary (X)) propagate.

At GEO, where the ambient plasmaspheric density is generally small, below 10^7 m^{-3} (i.e., below 10 cm^{-3}) or in the interplanetary medium (when the solar wind speed is unusually large so that the magnetopause is compressed from its usual upstream distance of 10 Earth radii (R_E) from the Earth's center to about 6 R_E)), spacecraft charging can cause problems. This happens when sunlight incident on a spacecraft surface creates photoelectrons (by the photoelectric effect) so that the surface is left charged positively; the flux of thermal electrons from the surrounding plasma is insufficient to neutralize this charge when their density is low. Between the sunlit and the dark sides of the spacecraft, a strong electric field develops. If this field becomes large enough, arcing occurs – that is to say a spark, a mini lightning flash, an electrical discharge takes place – through the interior of the spacecraft where the sensitive electronic equipment is placed. If the discharge passes through a circuit it "zaps" it, so that it no longer operates.

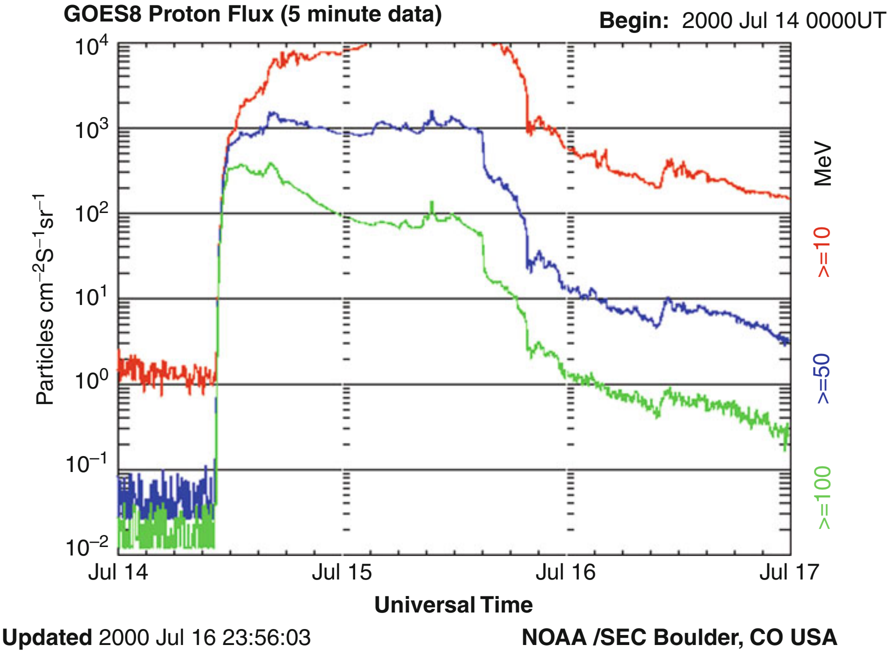


Fig. 10 Solar proton fluxes at GEO during a huge geomagnetic storm (Courtesy of NASA, http://pwg.gsfc.nasa.gov/istp/events/2000july14/)

Energetic Charged Particles

Energetic charged particles of energies exceeding a fraction of 1 MeV constitute a serious radiation hazard not only to astronauts but also to all forms of integrated circuits contained in electronic equipment aboard a satellite in LEO or GEO or on more distant roving spacecraft. This is because energetic ions cause extra currents to flow in these circuits, which can change the binary state of a computer memory; such a flip is called a single event upset (SEU). It may result in a phantom command to the satellite, which is hard for ground controllers to understand. The memory either may recover from the SEU, or the damage may be permanent, in which case it is called a single event latch-up (SEL).

Figure 10 shows observations made in geostationary orbit of the huge solar proton event which occurred on July 14, 2000, the so-called Bastille Day event. The proton data are shown for three energy thresholds; there is a three, or even four, orders of magnitude flux increase in a short time, and the enhanced fluxes persist for several days.

Deep dielectric charging and discharging occurs when 1 MeV “killer” electrons from the Van Allen belts penetrate the satellite walls and deposit charge on the insulating (dielectric) material of electronic circuit boards. The dielectric material can break down, causing electrical shorts unless the boards are carefully grounded.

Associated with a halo CME coming directly toward the Earth, the appearance of a coronagraph image (as shown in Fig. 5) suddenly changes. It becomes covered in white specks, sometimes called "snow." Each speck is produced when an energetic charged particle hits the charge-coupled device (CCD) detector in the instrument.

Another effect is that the performance of a satellite's solar array degrades with time due to solar proton bombardment. For example, the efficacy of the solar array aboard SOHO fell from 100 % to 82 % over 10 years. A further energetic ion effect is that of sputtering; when an ion of about 1 MeV or more hits a spacecraft surface, it kicks a few atoms from the surface. Over time a thin coating could be removed by this process or the surface material could be weakened.

Micrometeoroids and Space Debris

Particles with diameters exceeding 10^{-2} m (i.e., 1 cm), and moving at very high speed (10 km/s or even more) with respect to a satellite in LEO, pose an immediate danger to its integrity. This is because they can pass right through its outer case, be it made of metal or some composite material. Fortunately, the 21,000 such particles shown in Fig. 9 can be tracked from the Earth by radar and by optical telescopes, and their orbits calculated. If a collision appears to be imminent, the orbit of, for example, the ISS or a satellite having thrusters aboard can be changed a little to avoid a collision happening.

Of greater danger is the larger flux of particles with diameters between 0.1 and 1 cm whose orbital parameters cannot be known from observations made on the ground. There are believed to be about 6×10^5 such particles in LEO which can, for example, crack the glass covers of solar cell arrays or penetrate metallic foils. However, they can be stopped using honeycomb structures which act like the fenders (bumpers) on cars. Smaller particles whose diameters are a fraction of 1 mm can damage glass (optical) surfaces. The orbits of all debris particles tend to become more circular due to atmospheric drag acting at perigee. Because this drag is largest near solar maximum, this is the time in the solar cycle when most space debris is removed.

Protection of Application Satellites Against Space Weather Effects

Application satellites are complex and expensive spacecraft that cost many millions of dollars to manufacture and launch. Clearly, the protection of these assets against space debris, coronal mass ejections, energetic charged particles, increased thermospheric gas densities, and other hazardous conditions arising from space weather disturbances makes a good deal of sense. Maneuvering spacecraft to avoid collision with space debris is one possibility. Improved tracking and sharing of data on possible conjunctions of operational spacecraft is currently being pursued. There is a global network of solar observatories as well as of solar observing satellites that monitor the Sun 24 h a day, 7 days a week. When halo CMEs or other potentially destructive events are observed, that information is communicated to satellite operators. They thus have time to "power down" satellites and take other preventive measures that give spacecraft a much greater chance of surviving these events. To date, only a small number of satellites have been lost due to solar events or to collisions with debris, but something like the Carrington event, should it occur again, could have very severe effects on space assets.

Conclusion

In this chapter we have outlined the effects of:

1. The residual terrestrial atmosphere which causes a drag force to act on a satellite and of the highly reactive oxygen atoms present
2. The heating of a satellite due both to visible radiation from the Sun and to infrared radiation for the Earth
3. The ionospheric plasma which affects the strength of radio signals propagating from the ground to the satellite and vice versa, together with scintillations
4. Energetic charged particles – both ions and electrons – from the cosmos, from the Sun, and from the terrestrial Van Allen belts, which can penetrate into the heart of a satellite and damage its solid-state electronic circuits
5. Micrometeoroids and space debris which, if large enough (greater than 1 mm in size), can physically damage the structure of a satellite

The efficacy of all of these effects varies markedly through the 11-year cycle of solar activity. Fortunately, the forecasting of all such effects is improving as our understanding of the very many different processes which are involved improves with the continuing research being undertaken.

Cross-References

- ▶ Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing
- ▶ Lifetime Testing, Redundancy, Reliability, and Mean Time to Failure
- ▶ Orbital Debris and Sustainability of Space Operations

References

V. Bothmer, I.A. Daglis, *Space Weather: Physics and Effects* (Springer Praxis, Berlin/New York, 2007), 438 pp

M.J. Carlowicz, R.E. Lopez, *Storms from the Sun: The Emerging Science of Space Weather* (Joseph Henry Press, Washington, DC, 2002), 220 pp

I.A. Daglis (ed.), *Space Storms and Space Weather Hazards* (Kluwer, The Netherlands, 2001), 482 pp

I.A. Daglis (ed.), *Effects of Space Weather on Technology Infrastructure*. NATO Science Series (Springer, Dordrecht, 2004), 176 pp

J.W. Freeman, *Storms in Space* (Cambridge University Press, Cambridge, 2001), 139 pp

D.H. Hathaway, R.M. Wilson, What the sunspot record tells us about space climate. Sol. Phys. **224**, 5–19 (2004)

J. Lilenstein (ed.), *Space Weather; Research Towards Applications in Europe*. Astrophysics and Space Science Library, vol. 344 (Springer, Dordrecht, 2007), 330 pp

M. Moldwin, *An Introduction to Space Weather* (Cambridge University Press, Cambridge, 2008), 134 pp

M.J. Rycroft, The plasma and radiation environment in Earth orbit, in *Encyclopedia of Aerospace Engineering* (Wiley, Chichester, 2010), Chapter eae323

C.J. Schrijver, G.L. Siscoe (eds.), *Heliophysics: Plasma Physics of the Local Cosmos* (Cambridge University Press, Cambridge/New York, 2009), 435 pp

A.K. Singh, D. Siingh, R.P. Singh, Space weather: physics, effects and predictability. Surv. Geophys. **31**, 581–638 (2010)

P. Song, H.J. Singer, G.L. Siscoe (eds.), *Space Weather*. Geophysical Monograph, vol. 125 (American Geophysical Union, Washington, DC, 2001), 440 pp

A.C. Tribble, *The Space Environment: Implications for Spacecraft Design*, 2nd edn. (Princeton University Press, Princeton, 2003), 248 pp

A.C. Tribble, The effect of the space environment on spacecraft technologies, in *Encyclopedia of Aerospace Engineering* (Wiley, Chichester, 2010), Chapter eae568

Further Reading

http://goespoes.gsfc.nasa.gov/
http://royalsociety.org/summer-science/2011/aurora-explorer
http://sdo.gsfc.nasa.gov/data/
http://sohowww.nascom.nasa.gov
http://stereo.gsfc.nasa.gov
http://swri.org/9what/releases2011/solarwind.htm
http://www.esa-spaceweather.net
http://www.exploratorium.edu/spaceweather/
http://www.n3kl.org/sun/noaa.html
http://www.solarstorms.org
www.spaceweather.eu
www.spaceweathercenter.org
www.spenvis.oma.be
www.swpc.noaa.gov

Part IX

Appendices

Glossary of Terms

Joseph N. Pelton and Scott Madry

Ablation	In the context of a launch vehicle re-entry, this is the burning off of material, usually thermal protection shielding, as a result of contact with the atmosphere.
ABM	Apogee Boost Motor. Also see Apogee Kick Motor. This is a rocket motor, typically a solid fuel system that can be used to move a satellite or spacecraft from a highly elliptical transfer orbit into a circular geosynchronous orbit.
ABM	Anti-Ballistic Missile. This is a type of missile used as a defensive weapon to destroy or alter the path of a ballistic missile.
ABI	Advanced Baseline Imager that is used in meteorological satellites
Aborted Mission	This is a shutdown of a launch prior to liftoff due to a detected problem. Alternatively it can be the active destruction of a launch vehicle by the Range Safety Officer when there is a malfunction of a rocket and safety considerations, which indicates destruction of a rocket. In the case of launch vehicles with a crew, there is often an escape capability that can be utilized before the rocket is destroyed.

J.N. Pelton (✉)
International Space University Arlington, VA, USA
e-mail: joepelton@verizon.net

S. Madry
Global Space Institute, Chapel Hill, NC, USA
e-mail: Scottmadry@mindspring.com

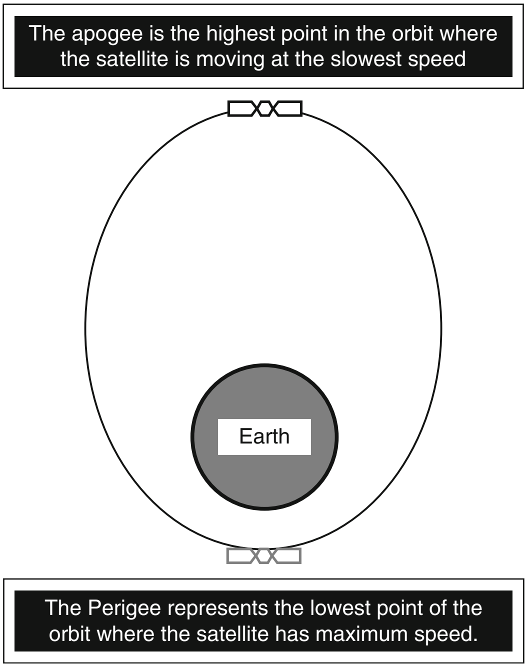
J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_108

Absolute Temperature	This is the temperature measured from the lowest possible theoretical temperature of absolute zero on the Kelvin scale where it has been calculated that all activity stops. The thermal temperature of “empty” outer space is universally measured at above 3° Kelvin due to the residual noise of the “Big Bang.” See Kelvin and Celsius Temperature Scales.
Access Services or Direct Access Services	This is a term used in satellite communications to refer to direct service to end users or consumers. Access service refers to the ability of satellites to provide direct access data, voice (Voice over IP), or digital video services to businesses, small offices, or home offices (SoHo). Most satellite data traffic around the world is heavy route service to provide broadband TCP/IP Internet services between major system nodes to provide heavy route or “trunked” data traffic between switching centers. Satellites, however, are increasingly able to provide “access service” to end users via VSAT or microterminals located at homes, businesses, or even desktops.
ACTS	Advanced Communications Technology Satellite. This was a NASA-funded satellite communications research project that carried out experiments involving transmissions in the Ka-band (30/20 GHz) as well as testing the concepts related to onboard processing and hopping beam antenna systems that could “hop” from various geographic locations for varying dwell times based upon traffic demand or requirements related to overcoming high levels of rain attenuation. These experiments on the ACTS experimental satellite were in a number of ways similar to more recent experiments carried out by the Japanese Space Exploration Agency (JAXA) with the WINDS research satellite.
ADCS	Attitude Determination and Control System. This is the subsystem on a spacecraft or launch vehicle that determines the current attitude of the vehicle in real time and is linked to a control system able to correct the spacecraft attitude orientation to the desired direction, typically through the firing of vernier jets.
ADPCM	Adaptive Differential Pulse Code Modulation. This is a digital compressed form of Pulse Code Modulation used in digital satellite networks.
ADM	Advanced Delta Modulation. This is a modulation system developed by Dolby Laboratories. This type of modulation is used in Australia for distributing digital audio via satellite.

Aerospace	This is a broad term that refers to activities related to vehicles or instruments that fly either in the Earth's atmosphere or in outer space. It can refer to the aerospace industry or aerospace vehicles and equipment. An aerospace vehicle typically refers to one that is designed to fly in airspace and into outer space and usually in a controlled manner. An aerospace vehicle can be a powered winged vehicle designed for atmospheric flight, a lifting body that has limited flight maneuverability or a rocket that might have systems designed for both vertical takeoff and landing capability. The definition of an aerospace vehicle as well as a space plane is a matter of some regulatory importance since difference agencies around the world may or may not have regulatory control over takeoff, liftoff, and/or landing responsibilities depending on whether the vehicle is defined as an aircraft, a space plane, or a rocket.
Aerospike rocket engine	This refers to the particular design of the exhaust system for a rocket engine. There are several basic designs for rocket nozzels. One is a "conventional" bell nozzle where the combusting fuel is forced through a narrow nozzle passageway and then expands. In the case of the case of the aerospike rocket engine, there are a number of smaller nozzle outlets arranged in a pattern. This approach allows more control of the rocket exhaust at different altitudes and changing atmospheric conditions. Aerospike rocket engine exhaust outlets discharge the combusted fuel against a truncated wedge and create a distinctive plume and series of shockwaves that create the "aerospike" pattern. One variation on the aerospike engine is a linear configuration where the smaller exhaust outlets discharge the combusted chemical propellant against a wedge and this is known as a linear aerospike engine. One can also deploy the smaller exhaust outlets in a circular fashion, and this becomes known as a plug nozzle.
A-GPS	Augmented Global Positioning Satellite (GPS) System
AIAA	The American Institute of Aeronautics and Astronautics, headquartered in Reston, Virginia
AK	Authentication key that is use in precision navigation and timing satellite systems.
AKM	Apogee Kick Motor. This is another name for an apogee boost motor. Also see apogee boost motor. Such a motor is not required if the launch vehicle can directly inject the satellite or spacecraft into the desired orbit.

ALIP	Alternating Linear Induction Pump. Such a pump is used in chemically fuelled launch vehicles.
Alliant-ATK Aerospace	Alliant Technologies Company, which merged with Thiokol, the major US supplier of solid rocket expendable vehicles, has now merged with Orbital Sciences Corporation to become Orbital ATK. Also see Orbital ATK.
Alternating Current	This is the mode of electricity where current alternates its path of flow. See Direct Current which is the most often the mode of operation on most spacecraft.
AM	Amplitude modulation or the varying of a signal to create an analog of sound, image or some form of information.
AMC	Applicable Means of Compliance. This is terminology used in space system standards in terms of meeting and complying with set standards.
Amos Satellite System	Amos is based in Tel Aviv. This is a series of Israeli communications satellites, Amos 1–6. All Amos satellites are operated by Spacecom and are developed by its partner, Israel Aerospace Industries. This is the case for all satellites except for Amos 5 that was developed in partnership with the Russian Company JSC Information Satellite Systems.
AMSU	Advanced microwave sounding unit that is a key instrument in meteorological satellites
AMTEC	Alkali Metal Thermal and Electric Energy Conversion. This is a term used with regard to launch systems and the specified standards for such energy conversion processes.
Anechoic Test Chamber	This is a test chamber that is designed to measure the Radio Frequency (RF) performance of a satellite in space prior to its launch.
Angara	A new Russian launch vehicle under development. See Appendix “▶ Major Launch Systems Available Globally.”
Anik	A name for Telesat of Canada’s satellite network for BSS and high speed digital services. Anik means “brother” in Inuit. Anik F2 is a high throughput satellite
Antares Launch Vehicle	This is the largest launcher developed by Orbital ATK (formerly Orbital Sciences Corporation). This with the Cyngus capsule is equipped to provide commercial launches to the International Space Station (ISS).
AOA	Abort Once Around. This is a term developed by NASA for the case of the Space Shuttle where the launcher has enough velocity to travel once around the Earth but then lands without achieving orbit.
Aperture	This is the size of a satellite transmitting or receiving antenna. The larger the antenna aperture size the greater

	the gain and thus the effective performance of a radio antenna.
APD	Avalanche Photo Detector
APL	The Applied Physics Lab that is a part of the US-based Johns Hopkins University located near Baltimore and conducts research in a wide range of advanced technologies including aerospace and electronics that is carried out for NASA, the US Department of Defense and other entities..
Apogee	The highest point or apex in an elliptical orbit. In Earth orbit, for instance, this would be when a satellite is farthest away from the Earth's surface but traveling at the slowest speed. Also see perigee which represents the reverse condition



(Graphic courtesy of J. Pelton)

AP Star	Hong Kong-based provider of BSS, FSS, and direct to home entertainment services.
Arabsat	The regional satellite system, based in Riyadh, Saudi Arabia, that provides satellite services to the Middle East and North Africa.

Argos	This is a telemetry communications system that is used primarily to support the relay of environmental information from ground- or sea-based sensors.
Ariane Launch Vehicle	This is the European Launch vehicle that operates from the launch facility in French Guiana. This center can now support launches by the Russian developed human-rated Soyuz launch vehicle. The Ariane 5 is able to launch multiton application satellites into GEO orbit.
ARSAT-1	A direct broadcast satellite for television service to Argentina.
AS	Antispoofing.
ASAT	This is a reference to anti-satellite technology and systems that could be used to destroy or disable satellites in Earth orbit.
ASI	Agenzia Speziale Italiana. This is the Italian Space Agency.
Asia Broadcasting Satellite System	This was formerly known as Agila. Its headquarters are officially in Bermuda. It provides BSS and directs to the home television services.
Asia Cellular Satellite	This was formerly known as Garuda. This is headquartered in London, UK, and is managed by Inmarsat. It provides satellite and terrestrial cellular services in the Asian region.
Asia Sat	This organization is based in Hong Kong and provides fixed, direct to the home, and broadcast services in the Asia region.
ASIC	Application-Specific Integrated Circuit. These highly specialized and increasingly complex and highly miniaturized integrated circuits are key to the architecture of spacecraft of all types as well as to the design of satellite handsets for telecommunications services. See also MMIC. This stands for the microwave monolithic integrated circuits that are also key components used in various spacecraft including applications and scientific satellites.
AST	The Associate Administrator for Commercial Space Transportation of the Federal Aviation Administration of the United States. This is usually denoted as the FAA-AST.
Astra	This is the name that is used to market direct to the home television services offered by SES of Luxembourg. Astra satellites now provide over a wide range of broadband digital services in Europe.
ATC	Air Traffic Control. This is the responsibility of ICAO at the international level and national or regional air traffic

	regulatory entities such as the FAA in the USA and EASA in Europe. With the advent of space planes and other near Earth activity involving high altitude platform systems, and robotically controlled vehicles at high altitude, there has been increasing concern, interest, and consideration of the issue of space traffic management and the interface between missions involve air traffic, robotic aircraft operations, near Earth orbit space activities, and space planes and other vehicles that may travel in both air and outer space. Also see ATM.
Atlas	This is a US expendable launch vehicle that is manufactured by the United Launch Alliance. The Atlas 5 is the largest launch vehicle in this launch series.
ATM	Air Traffic Management. This is the responsibility of ICAO at the international level and national or regional air traffic regulatory entities such as the FAA in the USA and EASA in Europe.
ATO	This is a term developed by NASA where the launcher has enough velocity to achieve orbit but without achieving the desired orbit required for the mission. A de-orbit would then be achieved from this irregular orbit and the mission thereby aborted.
Astra	This is the name used by SES for its direct to home television satellite service. Astra 2E is a high throughput satellite that provides a range of digital services.
AVHRR	Advanced very high-resolution radiometer, a key instrument used in meteorological satellites
Biedou	One of two Chinese precision navigation and timing satellite systems. This is the first-generation system, and its translation from Chinese is the "Big Dipper." Compass is the second-generation Chinese system, and it will eventually replace the Beidou system.
BIH	Bureau International l'Heure
Blue Origin	This is the commercial launch company founded by Jeff Bezos that is developing the New Shepard launcher. Blue Origin is also developing new launch motors for the United Launch Alliance.
BNSC	British National Space Center.
Boeing Corporation	This Aerospace corporation is a partner with Lockheed Martin in the United Launch Alliance (ULA) to build the Delta and Atlas expendable launch vehicles and is also a major supplier of space craft for space applications, space scientific research, and military and strategic applications.

BOL	Beginning of Life. This relates to the launch of a satellite and its initial check out after launch. See also End of Life (EOL).
BPS	Bits per second. It can sometimes stand for bytes per second.
Brazilsat	This is a satellite that provides television and voice and data services to Brazil. Television services is provided to all of Brazil, while voice and data services is largely provided to rural and remote areas of Brazil and especially Amazonia.
B-Sat	This is a Japanese broadcasting satellite.
BSI	British Standards Institute. This is the organization of the United Kingdom that establishes and maintains technical standards.
BSS	Broadcast Satellite Service: This is an official designation used by the ITU for its radio frequency allocations for this service.
B2B	Business to business data (B2B) relay satellite. Also see data relay satellites and store and forward satellites.
Bytes	This is the same as 8 bits of data.
C/A Code	This is the civilian GPS Course Acquisition Code.
CATEX	Categorical Exclusion. This is a term used with regard to space safety standards and their implementation.
C	The speed of light or 300,000 km/s
C Band	These are the 6/4 radio frequency bands that are used for FSS satellite communications
CCD	Charge Coupled Device
CCSDS	Consultative Committee for Space Data Systems. This is a process for the systematic sharing of space system performance and safety.
CD	Conference for Disarmament. This is a process that is conducted from Geneva by the United Nations. Also see United Nations Office of Disarmament Affairs (UNODA).
CD	In terms of precision navigation and timing, this refers to clock drift. This often is used to refer to compact disc in popular lexicon. Thus, the meaning must be read in context.
CDGPS	Canada-wide Differential Positioning System
CDMA	Code Division Multiple Access. This is a digital multiple access system that is commonly used in digital satellite communications. This is also sometimes called spread spectrum. Also see TDMA.
CDI	Center for Defense Information.

CDR	Critical design review. This is a key step in finalizing the design and specifications for a satellite, spacecraft, or launch vehicle.
CEN	Comite’ Europe’ en de Normalisation. Technical standards setting body for Europe. See also CENELEC.
CENELEC	Comite’ Europe’ en de Normalisation Electrotechnique.
CEOS	Committee on Earth Observation Satellites. This is an international coordinating organization that works to coordinate standards and practices in this area.
CEP	Circular Error Probability.
Chinasat	This is the domestic telecommunications and television satellite system for China. This system together with Sinosat provides a good deal of the coverage for services in rural and remote areas. Over 100,000 satellite terminals to provide television, telecommunications and rural and remote education and health care services are deployed in remote parts of China.
CNSA	Chinese National Satellite Agency
CISPR	This is the International Space Committee on Radio Interference. (French Acronym)
CME	Coronal Mass Ejection. This is a solar storm that ejects ionic particles from the sun’s corona at very accelerated velocity and can endanger satellites. Solar flares that represent solar radiation events can also endanger satellites.
CNES	This the French National Center for Space Research which is located in Toulouse, France.
CNSA	Chinese National Space Agency
CODEC	This refers to a coder/decoder that is used in digital satellite communications. A digital encoder today can use various forms of coding to send information much more efficiently than in the past. A decade or so ago one bit of information per 1 Hz of bandwidth was fairly standard in terms of transmission efficiency. Today using very efficiency turbocoding up to 4 to 5 bits of information per Hz is possible. These breakthroughs in encoding have included multiphase coding, Reed-Solomon, Viturbi encoding, and Turbo-coding.
COMMStellation	A network of 78 microsatellites in polar orbit planned for deployment in 2018.
Compass	The Chinese second-generation precision navigation and timing satellite system that replaces the Beidou Satellite System.
Constellation	This refers to a network of satellites typically designed to provide global or near global coverage of the earth. This

	can range from a two-satellite network that provides store-and-forward data relay to a large constellation such as that represented by the Iridium mobile satellite network in low earth orbit with 66 satellites to proposed so-called mega leo satellite networks such as One Web that would have thousands of small satellites in a swarm constellation.
COPUOS	The Committee on the Peaceful Uses of Outer Space. The UN Committee of some 70 countries that address policies and issues related to outer space.
CORS	Continuously Operating Reference Station. This is part of the augmented GPS system.
COSPAS	Cosmitscheskaja Sistema Poiska Awarinitsch Sudow (COSPAS) is the Russian search and rescue satellite network. This system together with the SARSAT satellite network is used to rescue downed pilots and distressed people at sea. (Also see SARSAT).
Cross link	This is one of the terms used to refer to intersatellite links that can refer to links between satellites at close range between a LEO or Meo Constellation or a much longer link between satellites in GEO Orbit. Also see intersatellite link.
Cube Satellite	This is a quite small satellite that is typical 10 cm × 10 cm × 10 cm in size that is in the 1 kg to 10 kg range. There can be various unit sizes from a 1 unit cube sat up to a 6-units version. (See also Small Satellites, Nano Satellites, and Micro Satellites.)
DABS	Digital Audio Broadcast Service. This refers to satellites designed to provide direct broadcast services primarily to vehicles. These services can include news, music, entertainment radio, and security and safety services. XM-Sirius Radio and Worldspace are the two prime providers of this service.
DAGGER	This is the name given by the US Department of Defense to the second-generation GPS receiving unit that is utilized by US forces.
DASS	The US Distress Alerting Satellite System.
Data Resolution	This is a key term for remote sensing. The sensing that is done by remote sensing satellites can have a resolution (or fineness of detail) in four different areas. These four different categories of resolution reveal different types of information. These four areas are called: (i) **Spatial resolution** is how much detail is captured in terms of pixels per image. This means what detail can be clearly seen. (ii) **Temporal resolution** is when the image by the sensing

	satellite was actually collected and recorded; (iii) **Spectral resolution** is what bandwidth within the electro-magnetic spectrum was the data collected. This could be in a part of the visible spectrum, above it in the ultra-violet range, or below it in infra-red range; and finally (iv) **Radiometric resolution** that provides what might be called the relative "brightness" of the elements in the image. In the postprocessing of the data, these four key elements of resolution can reveal a great deal information.
DBSD	This is the hybrid MSS satellite system that was obtained from ICO. Offices are in Bellvue, Washington and Reston, Virginia
Delta Launch Vehicle	This is a long time and well-proven US Launch Vehicle with a 95 % launch success rate. The Delta II and Delta IV are still in use. It, like the Atlas, is currently designed and manufactured by the United Launch Alliance.
DGNSS	Differential Global Navigation Satellite System.
DGPS	Differential Global Positioning Satellite System.
Digital Globe Corporation	This is the current name for what was once known as GeoEye. Its current assets include the following imaging satellites known as IKONOS, QuickBird, WorldView-1, GeoEye-1, WorldView-2, and WorldView-3. These satellites are capable of collecting over one billion square kilometers of imaging data in the course of 1 year cycle.
Direct Current	Direct Current or DC electricity. AC stands for Alternating Current.
DirecTV	This is a direct broadcast satellite system, headquartered in Colorado, that provides news and entertainment services to North America and also provides digital download services as well.
Dish	This is a US-based direct broadcast satellite system, the main competitor to DirecTV. This system has also been known as Echostar. This company also now owns Hughes Network Systems (HNS), which is the largest provider of very small aperture antennas, and now provides satellite services (known as Hughes Net) in rural areas.
DLR	This is the German space agency.
DOCSIS	Data over Cable Systems Interface Standard. This standard is the foundation for high-speed access to content on the Internet. Originally this was developed for North American cable operators, but it is used by many satellite operators such as Via Sat.
DoD	The United States Department of Defense.

Doppler Shift	The change in frequency when a device transmitting moves away or toward the signal source.
Dual Use	The use of civilian satellite to provide capacity to support military communications capabilities. This usually involves nontactical communications to support television broadcasts to overseas troops and other such services. In some cases, commercial systems deploy networks in military spectrum bands such as X-band as is the case with XTAR. In such cases, commercial systems may be used for tactical communications and war-fighting services.
DVB-RCS	This is the Digital Video Broadcast-Return Channel Service that provides direct television services to the home, but also provides high speed data services to corporate satellite antennas.
Earth Station	A facility that can transmit to and receive from a satellite. This typically refers to a larger aperture facility that might be design to carry out tracking, telemetry, commands, and/or monitoring of in-orbit satellite facilities or to carry out these activities during a launch of a satellite.
EASA	The European Aviation Safety Agency is headquartered in Cologne, Germany, and is responsible throughout Europe for rulemaking related to the safe operation of aircraft, initial certification of aircraft as air worthy and certification of maintenance standards to ensure airworthiness continues. More recently EASA has begun a process to establish safety standards and certification procedures for space planes with wings that would operate in Europe.
Echostar Corporation	This is the owner of EchoStar Satellite Services L.L.C. that provides advanced satellite communications solutions including video distribution, data communications, and backhaul services for media and broadcast, enterprise, government, and military customers. In addition, the company provides spacecraft operations and command and control services for EchoStar's fleet of 24 owned, leased, and managed in-orbit satellite. Prior to 2008, EchoStar operated the Dish Network service brand; the Dish Network brand was spun off as Dish Network Corporation on January 1, 2008. It is also owner of Hughes Communications Inc. and Hughes Network Systems. Also see.
EDM	Electronic Distance Measuring
EEPROM	Electrically Erasable Programmable Read Only Memory.
EDO	Extended Duration Orbiter. This is a term no longer used that referred to a Space Shuttle configured for an extended duration flight.

EGNOS	European Geo-Stationary Navigational Overlay System
Ekran	The Ekran system (which was also known as Stationar-T) was the Soviet Union's first operational geosynchronous satellite and the world's first direct-to-home TV service. It provided color television broadcast to Siberia and the Far North. This has now been replaced by the Yamal satellite system. Also see Yamal satellites.
Electromagnetic Spectrum	The key to all application satellites is the ability to communicate with the satellite via radio wave frequencies. Multispectral sensing, infrared sensing, and radar sensing are key to remote sensing and meteorological satellites. Understanding of high-intensity radiation from the sun is key to being able to protect application satellites from solar storms. Thus, virtually all elements of the electromagnetic spectrum from the longest wavelengths to the shortest and highest intensity wavelengths are important in some way to the field of application satellites. The most important frequencies for communications, remote sensing, meteorological satellites, and precision navigation and timing are in the UHF band (300 MHz to 3,000 MHz), the SHF band (3,000 MHz to 30 GHz), and the EHF band above 30 GHz through the visible light band. The International Telecommunication Union (ITU) through an elaborate global consultation process allocates frequencies for these various uses.
Electro-1	Russian Satellite.
EMP	Electro Magnetic Pulse.
ELV	Expendable Launch Vehicle. Launcher that are used only once and their various stages typically designed to fall into the sea.
EOL	End of Life. This relates to the end of life for a satellite. Also see Mean Time to Failure (MTTF).
EPIRB	Emergency Position Indicating Radio Beacon (EPIRB) that sends emergency signals to SARSAT-COPAS.
ESA	The European Space Agency whose headquarters are in Paris France, but this organization has various facilities all over Europe. Its ESTEC facility is in Noorwyck, The Netherlands.
ETS	This is a series of experimental satellites launched JAXA. ETS stands for experimental test satellite. The series extended from ETS-1 to ETS-8.
EU	The European Union
EUMETSAT	This is the organization that designs and arranges for the launch of meteorological satellites for Europe. Many of the

	Eumetsat satellites operate in tandem and close cooperation with US meteorological satellites.
Eutelsat	This is a satellite system that began as a public international organization that provided regional satellite services to the European region, but is now a privatized organization that provide global fixed and broadcast satellite services globally.
Express	This is a Russian telecommunications satellite, although most of these were constructed by Alcatel of France. These satellites are operated by the Russian Satellite Communications Company (RSCC). Currently the Company provides space segment capacity to users in 35 countries and, with its orbital and frequency capacity, is one of the world's ten largest satellite operators. In 2012, the RSCC constellation includes 11 satellites that are positioned along the geostationary arc extending from 14 °W to 140 °E. The Company's ground infrastructure includes five satellite communications centers in European Russia, Siberia, and the Far East, as well as the Shabolovka Technical Center in Moscow.
FAA	The United States Federal Aviation Administration that is responsible for regulating commercial space launches
FAA-AST	The FAA Office of Commercial Space Transportation
Falcon Launch Vehicle	This is the new commercial launch system developed by Space X. This includes the Falcon 1 and the Falcon 9 and Falcon 9 Heavy that are much larger launch vehicles.
FCC	The Federal Communications Commission of the United States that approves satellite communications systems and frequency assignments and orbital locations for new applications for such systems.
Feng-Yun	This is the name of the Chinese meteorological satellites. These satellites consist of both polar orbiting satellites and geosynchronous satellites and have evolved in design and capability over time. The Feng-Yun meteorological satellite that was defunct and shot down by the Chinese military is of particular note in that this generated nearly 3,000 sizable debris elements that has been of major concern because of the growing amount of orbital debris in low earth orbit and because of this debris is in proximity to the International Space Station.
FM	Frequency modulation or the varying of a signal's frequency to create a model of sound, image, or some other form of information.

Frequency	This is the measure of a radio waves' pattern of variation with time. The basic formula in terms of determining a radio wave's frequency is given by the following formula: **C** (speed of light) = the wavelength (cm) × its frequency (1/lambda) or (per second) Frequency is today expressed as a Hertz or (HZ) in honor of the scientist who developed so much of our knowledge about radio wave transmissions.
FSS	Fixed Satellite Services. This is an official designation used by the ITU for its radio frequency allocations for this service.
FTP	File Transfer Protocol.
GAGAN	GPS Aided GEO Augmented Navigation (this is the Indian Wide Area Augmentation System (WAAS) Network).
Gain	This is a measure of a radio antenna's performance. The Gain is determined by considering its aperture size, its operating frequency, the accuracy of its shape, and its effective efficiency. This is calculated for a parabolic-shaped satellite antenna dish as follows: $\text{Gain} = \text{E} \times \pi \text{A} = \text{E}\, \pi^2 \text{r}^2/\text{lambda}^2$ E is a dimensionless parameter between 0 and 1 called the aperture efficiency. (Note: The aperture efficiency of typical parabolic antennas is 0.55 to 0.70.) $\text{A} = \pi \text{r}^2$ is the area of the antenna aperture, that is, the mouth of the parabolic reflector. Lambda is the wavelength of the radio waves.
Galaxy Satellites	There are satellites serving North America that are now owned by Intelsat, but were originally launched by Hughes Communications.
Galileo Precision Navigation and Timing Satellite System	This is the planned system that Europe is planning to deploy.
GBBF	Ground-Based Beam Former. This is a technology used for MSS services with auxiliary terrestrial component (MSS-ATC) or MSS with Complementary Ground Component (MSS-CGC)
GCCS	Geostationary Satellite Communications Control Segment (GCCS), a system that has been established by the US Federal Aviation Administration.

GDOP	Geometric Dilution of Precision. This is a term used with regard to Precision Navigation and Timing Satellites.
GeoEye	GeoEye Inc. (formerly Orbital Imaging Corporation or ORBIMAGE) was an American commercial satellite imagery company based in Herndon, Virginia, in conjunction with it major investor Orbital Science Corporation. GeoEye was merged into the DigitalGlobe corporation on January 29th, 2013. The company was originally founded in 1992 as a division of Orbital Sciences Corporation. This company was set in light of the provisions of the 1992 Land Remote Sensing Policy Act which permitted private companies to enter the satellite imaging business. The division was spun off in 1997. It changed its name to GeoEye in 2006 after acquiring Denver, Colorado-based Space Imaging for $58 million. Space Imaging was initiallye founded and controlled by Raytheon and Lockheed Martin. Its principal asset was the IKONOS satellite. See DigitalGlobe.
GEO Orbit	Geosynchronous orbit. This is a special circular orbit in the Earth's equatorial plane that is sometimes known as the Clarke Orbit. This orbit is 35,870 km above the Earth's surface (or 22,230 miles). This unique orbit allows a satellite to revolve around the world exactly once a day, and thus, earth station antennas do not have to track the satellite since their remain constantly pointed at a satellite in that orbit. This orbit is thus highly desired for satellite communications and meteorological applications as well as other uses.
GEOSAR	This is a Search and Rescue satellite in GEO orbit.
GE Satellites	These are satellites that provide direct broadcast, direct to the home television and fixed satellites services. This company is owned by the GE Corporation and is based in Bethesda, Maryland in the USA.
GIOVE	Galileo In-Orbit Validation Element. This is a Galileo test satellite.
GLM	Geostationary lightning mapper (GLM). This is a new feature in the latest meteorological satellites.
GlobalStar	A global mobile voice and data communications system that is US based. It began as a low earth orbit constellation but has transitioned to a Geo orbit-based system.
GLONASS	This is the Russian precision navigation and timing satellite service. It has now been restored to full global capability.

GLS	GNSS Landing System (GLS) that is a system defined by the FAA for use of GNSS systems for aircraft takeoff and landing.
GMS (Himawara Series)	This is the Japanese Geostationary Meteorological Satellite. It is also known as the Himawara meteorological satellite series.
GNSS	Global Navigation Satellite Systems. These are also called Positioning, Navigation, and Timing (PNT) satellites. Also see GPS, Glonass, Beidou, Compass, Quasi-Zenith, Galileo, and Indian Regional Navigational Satellite System.
GOES	Geostationary Operational Environmental Satellites. This is a US-operated GEO orbit-based meteorological satellite system. The third generation is now in service, and the fourth generation, namely, GOES R,S, T and U, will be deployed so as to maintain GEO orbit coverage through 2036.
GOMS-Elektro	This is the Russian Geostationary Operational Meteorological Satellite System. This is also known as the Elektro system.
GPS	The Global Positioning Satellite network that provides precision navigation and timing services. This is a US-based service that provides a global constellation of some 30 satellites that is deployed by the US military but is used for free for civil applications around the world. The element of so-called selected availability that provides less precise targeting and positioning capability is no longer used.
G Star	A satellite system for the USA and North America jointly owned by GTE & SES Americom.
GUS	Ground Uplink Station. These are used in many systems such as in the case of the GCCS-WAAS for aviation.
HAPS	High Altitude Platform Station. This is the term approved by the International Telecommunication Union to refer to high altitude platforms maintained a constant location (such as 21 km) to provide communications or other services.
HEO	Highly Elliptical Orbit. This is an orbit with a low perigee but a very high apogee. This is sometimes called a Molniya Orbit after the original Russian domestic satellite that used three Molniya satellites in 12 h orbits with 8 h above the horizon to provide coverage for the Russian (USSR) country-scape. Sometimes this is also called an Extremely Elliptical Orbit (EEO).

HIRS	High-resolution infrared radiation sounders (HIRS). This is an instrument used on meteorological satellites.
Hisdesat	This is a group of investors that includes Hispasat that together with Loral Communications owns the XTAR satellite that provides communications in bands reserved for military satellite communications.
Hispasat	This is a company based in Madrid Spain that provides satellite services to Europe and South America. They are also investors in XTAR that involves a partnership with Loral Space and Communications Inc. (Also See XTAR)
HTS	High Throughput Satellites. These are satellites of high efficiency that can transmit in the range of 10 to 150 Gigabits/s.
Hughes Communications Inc.	This is the parent company that operates **HughesNet** which is a satellite-based broadband Internet provider and **Hughes Networks Systems** that operates high throughput satellites to support Hughes Net. It is also the largest provider of Very Small Aperture Antennas for digital satellite business networks. This company, which is located in Germantown, Maryland, is entirely owned by Echostar Corporation L.L.C. Also see Echostar Corporation.
Hughes Jupiter High Throughput Satellite	This high throughput satellite to offer high speed Internet and other digital services has been recently launched by Hughes Communications.
Hybrid MSS	This is a mobile satellite system that also combines with terrestrial cellular mobile service in urban areas. This is sometimes characterized as MSS-ATC or MSS-CGC.
Hylas	This is a system that is operated by the Avanti Corporation that is based in London, UK. It provides FSS data and broadband data services.
Hyper-spectral imaging	This is the latest form of remote sensing that collect data in much smaller increments of spectra. Thus, data are obtained over dozens of different spectra ranges (perhaps a hundred different samples covering the entire light waves as well as infrared and ultraviolent) so that much more precise interpretation can be made of the data as to crop disease, use of urban land, etc. This is in contrast to previous spectral sensing that broke the entire spectra down into just a few parts (like just five or six parts for the entire spectra). (Also see Multi-Spectral Sensing)
IAASS	The International Association for the Advancement of Space Safety. This is a professional organization, headquartered in the Netherlands, and is devoted to the topic of space safety. It sponsors the Space Safety Magazine and the Journal of Space Safety Engineering.

IADC	the Inter Agency space Debris Coordinating Committee
IBRD	International Bank for Reconstruction and Development that is headquartered in the USA. This is more commonly known as the World Bank
ICG	International Committee on Global Navigational Satellite Systems that is coordinated under the good offices of the UN Committee on the Peaceful Uses of Outer Space (COPUOS) and the secretariat for which is provided by the UN Office of Outer Space Affairs.
ICAO	The International Civil Aviation Organization based in Montreal, Canada, and brought into being by the 1944 Chicago Convention.
ICO Global Communications	This company that was originally based in London and went through bankruptcy is reorganized and is now based in Bellvue, Washington, in the USA. See website www.ico.com for latest information as to its MSS service offerings.
IJPS	Initial joint polar-orbiting satellite (IJPS) system that is a US meteorological satellite program
IMF	The International Monetary Fund. See also the IBRD.
IMO	The International Maritime Organization. This is the United Nations Specialized Agency to coordinate international policies and regulations with regard to all things related to maritime safety and operations.
India Geosynchronous Satellite Launch Vehicle (GSLV)	This is a now proven expendable launch vehicle that can lift application and scientific satellites into geosynchronous orbit or beyond that has been developed by ISRO.
India Polar Satellite Launch Vehicle (PSLV)	This is a now proven expendable launch vehicle that can lift application and scientific satellites into low earth polar orbit that has been developed by ISRO. It can lift about 600 kg into sun synchronous polar orbit.
IRNS	Indian Regional Navigational Satellite System.
Inmarsat	This is a large international satellite network that provides mobile communications satellite services and is based in London, United Kingdom. It was once a public international organization but has been privatized and operates as a private corporation.
Inmarsat Xpress	This is
INRSS	Indian Navigation Regional Satellite System
INS	Inertial Navigation System
InSat	The name of the Indian domestic satellite communications network that is deployed by the Indian Space Research Organization.

Intelsat	A large global satellite network headquartered in Luxembourg, but its major operations are in the Washington, DC, area. This organization was once a public international organization but has been privatized and acquired by equity investors. It has acquired the PanAm Sat organization that operates from Atlanta, Georgia.
Intelsat Epic	This is the latest Intelsat satellite that is a high throughput satellite and operates in the Ka band.
Intersputnik	This is the Russian-led international communications satellite network. This was created to compete with Intelsat during the years of the Cold war between the U.S.S.R. and the USA.
IP Star	The company that launched the Thaicom satellites also started and launched for the Asian regional service area the high-efficiency IPStar satellites that were as their name suggested optimized for IP-based broad band services. These satellites operated with high-efficiency coder/decoder (CODEC) technology. IP Star has a licensed-operating arrangement with 14 countries in the Asia-Pacific region and claims to be the prime broadband IP provider in the region. It provides mobile satellite services to ships at sea as well as broadband IP-based FSS services. By means of its high-efficiency CODEC, it was one of the first company to deploy high throughput satellites.
Iridium	A global voice- and data-based mobile satellite network. It provides complete global coverage using a network of low earth orbit satellites in near polar orbits. Its generation Next satellites will provide expanded capacity and include hosted payload for aeronautical mobile satellite service.
IRNSS	Indian Regional Navigational Satellite System.
ISDCC	The Interagency Space Debris Coordinating Committee of the United Nations
ISL	Intersatellite Link. This is also called a crosslink. Some of the low earth orbit satellite networks such as Iridium have ISLs to connect satellites in their LEO constellation.
ISRO	The Indian Space Research Organization which is based in Bangalore India and various other locations that are largely in southern India. This space organization for India is responsible for all of its space launch vehicles, its space research activities, and all of its space applications programs related to telecommunications, remote sensing, navigation, and meteorological satellites.
ISS	International Space Station. This international project is managed as a series of agreements among the various

participating space agencies, known as franchise agreements. This process is used to coordinate the various elements or modules that make up this largest of spacecraft in Low Earth Orbit.

The International Space Station (Image Courtesy of NASA)

ITAR	International Traffic in Arms Regulation. The US regulatory review process of technology considered to be of strategic or military importance.
ITOS	Improved TIROS operating system (ITOS)
ITRF2008 and ITRF2014	International Terrestrial Reference Frame2008 that will soon be replaced by ITRF2014.
ITU	International Telecommunication Union. This is the United Nations specialized agency headquartered in Geneva, Switzerland, that coordinated standards for telecommunications on a global basis and also represents the forum for the allocation of radio frequencies.
Japan H2 and H2A Launch Vehicles	The H2 and H2A represent the current largest expendable launch capabilities of Japan. The H2A is possible of launching spacecraft to the Moon or to launch heavy communication satellite payloads to GEO orbit.
JAXA	This is the Japanese Aerospace eXploration Agency (JAXA) that develops the H2A and H2A Transfer Vehicle (HTV) to provide lift capability to the ISS as well as new technology for space application satellites as well as new spaceplane technology.

JPL	The Jet Propulsion Lab that carries out a wide range of research in aerospace, propulsion, and electronics for NASA and the US Government.
JPSS	Joint polar satellites system (JPSS). This is meteorological satellite program that is a joint program between the US Department of Defense and the National Oceanic and Atmospheric Administration (NOAA)
JSAT	Japan Satellite Corporation is headquartered in Tokyo, Japan. JSAT has deployed Horizon-1 over North America as a joint venture with Intelsat to provide direct broadcast services. Horizon-2 operated by JSAT also provides direct broadcast services for the Asia region.
Ku Band	These are the 14/12 GHz radio frequency bands that are used for satellite communications.
Ka Band	These are the 30/20 and 28/18 GHz frequency bands that are used for FSS and BSS satellite communications.
KA Star Corporation	This is a Ka-band satellite corporation that was formed in 1995. It obtained financing from venture capital including from Kleiner Perkins. At one point, it was renamed iSky Corporation. The two Ka-Band satellites it eventually launched were built by SSL and renamed Wild Blue. These satellites were developed for service to the US. Also an agreement was signed with Telesat for capacity from their Anik F2 satellite. The Wild Blue Satellites were eventually acquired by Via-Satellite. (Also see Via Sat.)
KHz	Kilo Hertz or 1,000 cycles per second
Koreasat	These are satellites launched to provide telecommunications services including direct broadcast satellite television to Korea. The latest version of these satellites has been designed to provide regional service to Asia.
Kosmos	This was the early experimental meteorological satellites of the Soviet Union in the period 1965–1969.
LAAS	Local Area Augmentation Services
LBS	Location-Based Services.
LEO Orbit	This stands for low earth orbit and because the satellites are much closer to the Earth's surface requires many more satellites flying in a defined constellation to provide complete Earth coverage. Thus, a typical LEO constellation is composed of about 50–70 satellites and fly in orbits that are about 500 km to 1,200 km in altitude (or 300 miles to 750 miles). Satellites in this orbit have much less transmission delay and experience much less "path loss" in

	terms of the spread out of a signal as it travels from a satellite to the Earth.
LHP	Loop Heat Pipe
Light Square Corporation	This company that deployed the SkyTerra-1 satellite but has now gone bankrupt due to the problem with FCC reversal of its regulatory approval of frequencies for use with the ancillary terrestrial component of the overall mobile cellular service in the USA.
Link Budget	This is a calculation of the power needed to complete a link between a communications satellite and transmitting and/or receiving antenna.
Link Margin	This is the additional power that is added to the minimum power needed to complete a link to provide confidence that service will be maintained in light of factors that might impinge on the quality of service such as rain attenuation, atmospheric scintillation.
Lockheed-Martin	This Aerospace corporation is a partner with Boeing in the United Launch Alliance (ULA) to build the Delta and Atlas expendable launch vehicles and is also a major supplier of space craft for space applications, space scientific research, and military and strategic applications.
Long March Launch Vehicles of China	These launch vehicles vary in capacities from the Long March 1 to the Long March 5.
Loral Space and Communications Corporation	Loral Space and Communications is a satellite communications company that has undergone Chapter 11 bankruptcy. It owns 62.8 % of Telesat, which provides reliable and secure satellite-delivered communications solutions to broadcast, telecom, corporate, and government customers on a global basis although Telesat also provides Canadian domestic satellite services as well. Loral also owns 56 % of XTAR, a joint venture between Loral and HISDESAT, a consortium comprised of leading Spanish telecommunications companies, including Hispasat, S.A., and agencies of the Spanish government. Loral Space and Communications Corporation was also the former parent company of Space Systems/Loral the satellite manufacturer that is now known simply as SSL. In November 2012, McDonald Dettwiler Associates (MDA) completed the acquisition of this company that split off from Loral Space and Communications. (See also MDA)
LORAN	The LOng RAnge Navigation ground-based radio navigation system. This capability has been largely superseded by the GPS network.

LSA	Launch Service Alliance. This alliance includes Arianespace, Boeing on behalf of SeaLaunch, and Mitsubishi of Japan on behalf of the H-II and HIIA launch vehicles.
MDA	This stands for McDonald Dettwiler Associates. This corporation that is headquartered in Richmond British Columbia, Canada, is a manufacturer of many types of space products that range from the Canada arm on the ISS, to Dextre, space robotics, to remote sensing satellites such as Radarsat, to many other application satellites; in November 2012, it acquired the US satellite manufacturer Space Systems/Loral that is now known as SS/L.
Measat	This is the Malaysia domestic satellite system. Some of the satellites offer the ability to provide services to Southeast Asia.
MegaLEO Constellation	This is a term that is sometimes applied to LEO constellations with a very large number of satellites in their network such as OneWeb and the proposed network of SpaceX.
MEO Orbit	This stands for Medium Earth Orbit. This is an orbit that allows total Earth coverage with a network of some 12 to 18 satellites depending on the orbit. Satellite networks deployed in MEO are deployed in what are called constellations and typically fly in a defined pattern that is above the lower Van Allen Belt and in orbits that range from about 8,000 km to 16,000 km (or 5,000 miles to 10,000 miles)
MEOSAR	A search and rescue satellite in medium earth orbit.
Mesbah	This is a store and forward data relay satellite that has been deployed by Iran.
Meteor	Polar orbiting satellites for meteorological monitoring that were launched by the Soviet Union. Meteor-1 s in the late 1960s through 1978. Meteor-2 s were launched from 1973 through 1993. Meteor-3 were launched starting in 1985. The latest version of these satellites is known as Meteor M. At least, 25 satellites of this type have been launched by the Soviet Union.
MHz	Mega Hertz or a million cycles per second
Minotaur Launch Vehicle	This is one of the launchers developed by the Orbital Sciences Corporation (now Orbital ATK). It is between the Pegasus and the Taurus in terms of lift capability.
MMIC	Microwave Monolithic Integrated Circuits. These are key solid-state electronic components that used in various spacecraft including applications and scientific satellites.

Modem	This stands for Modulator/Demodulator. This is a basic function to modulate a signal to transmit through a communications satellite. The various modulation schemes involve AM or amplitude modulation, FM or Frequency modulation, digital pulse code modulation (PCM), or Delta modulation. In modern digital satellite communications, analog modulation such as AM or FM is no longer used because of the advantages provided by digital encoding and that allow highly efficient digital compression techniques to be applied.
MPEG	This stands for Motion Picture Expert Group. This is the body that develops and agrees technical standards for compressed digital transmission of video. MPEG-2 is often used for television transmission via digital satellite communications systems at 4 megabits/s.
MSS	Mobile Satellite Service. This is an official designation used by the ITU for its radio frequency allocations for this service.
MSS-ATC	This refers to mobile satellite service-ancillary terrestrial component or hybrid satellite systems that combines terrestrial cellular wireless service with mobile satellite services. This type of hybrid MSS services is called MSS-CGC or Mobile Satellite Service-Complementary Ground Component.
MTTF	This stands for Mean Time To Failure. It is used to calculate the expected lifetime of applications satellites.
Multispectral Remote Sensing	This was the type of sensing done by remote sensing satellites that took images of the Earth divided into on a few spectral ranges but splitting the entire visible spectrum into perhaps five to eight spectral ranges. Today the latest technology slices the spectrum into much narrower spectral images with what is called hyper-spectral imaging. (Also see hyper-spectral sensing)
MUOS	The Mobile User Objective System. A mobile communications satellite in low earth orbit that is deployed by the US Department of Defense.
Nadejda	This was the Soviet polar orbiting environmental satellite network.
Nano Satellite	A Nano Satellite is a quite small satellite. It will typically be in 1 kg to 10 kg range (also this can be a cubesat). Even smaller are what is called a Pico Satellite which is typically in 100 g to 1 kg range and a Femto Satellite which is in the 10–100 g range

NASA	The National Aeronautical and Space Administration, the space and aeronautical agency of the United States Government that is headquartered in Washington, DC, but carries out most of its research and development and operations at its various centers. The largest of these Centers, where the bulk of NASA's 14,000 or so employees reside, are Ames Research Center in Mountain View, California; Glenn Research Center in Cleveland, Ohio; Goddard Space Research Center in Greenbelt; Johnson Spaceflight Center at Houston, Texas; Kennedy Spaceflight Center at Cape Canaveral, Florida; Goddard Research Center in Greenbelt, Maryland; Marshall Spaceflight Center in Huntsville, Alabama; and Stennis Research Center, Mississippi.
NASA COTS	This is the commercial orbital transportation system (COTS) program that was first designed to provide lift capability to the International Space Station (Space X and Falcon 9 with the Dragon Capsule) and (Orbital Sciences Corporation and the Antares with Cygnus capsule). This has now transitioned to the program to provide for the transportation of astronauts to and from the ISS. The finalist contractors for this are the Boeing Corporation (uprated Atlas plus CST-100 capsule) and Space X (uprated Falcon 9 plus the Dragon Capsule).
NASDA	The National Aeronautical and Space Development Agency (NASDA). This was the previous name for the Japanese Space Agency before it was combined with the National Aerospace Labs (NAL) and the Institute of Space and Astronautical Science (ISAS) at the University of Tokyo to form JAXA.
Navstar	This is the actual name of the satellites in the GPS System. NAVSTAR stands for NAVigation Satellite Timing and Ranging satellites.
NDGPS	The Nationwide Differential GPS System.
NESDIS	National Environmental Satellite Data Information Service
New Shepherd Launch Vehicle	This is the launch system being developed by Blue Origin.
NGS	The US National Geodetic Survey.
Nigcomsat	This is the name of the Nigeria satellite system that is designed to provide television and telecommunications services. The original satellite built and launched by China was a failure, but this satellite has now been replaced and is operating normally.

Nilesat	This is the name of the domestic satellite communications network that provides domestic television and telecommunications services for Egypt.
NIMBUS	This was one of the early national meteorological satellites that was an experimental satellite by NASA carried in cooperation with NOAA.
NOAA	The National Oceanic and Atmospheric Administration.
Northrop Grumman Corporation	This is one of the largest US aerospace corporations. There prime area of emphasis is US military projects that include space systems.
Nova Greece	This is the direct broadcast satellite system for Greece that provides BSS services and digital video broadcast data services.
NPOESS	National polar-orbiting operational environmental satellite system.
N-Star	This is a Japanese mobile satellite system.
NUDET	The nuclear denotation sensors on the GPS satellites. The detection sensors on the GPS have allowed scientists to determine that asteroid strikes on Earth are at least four times more frequent than had previously been thought to be the case.
NWS	This is the United States' National Weather Service.
OBSS	Orbiter Boom Sensor System of the Space Shuttle.
ODS	Orbiter Docking System for the Space Shuttle.
OICET	Optical Interorbit Communications Engineering Test (OICET). This was a Japanese test satellite to conduct experiments with optical intersatellite links (ISLs)
Okean	This is a joint Russian-Ukrainian Earth observation satellite that is designed primarily for ocean monitoring.
OneWeb	This is a proposed "swarm constellation" or mega LEO network that would provide global Internet access using about 800 small satellites (150 kg class) in low earth orbit including spares. These satellites turn to the side on their axis as they approach the equatorial orbital arc to avoid interference with satellites in GEO orbit and then return to pointing to Earth as they pass the equatorial zone.
OOSA	Office of Outer Space Affairs. This is the United Nations Office that supports the activities of the United Nations Committee on the Peaceful Uses of Outer Space.
Optus	The Australia satellite communications service provider that is based in Sydney, Australia. This organization provides a state of the art network for all of Australia. It also provides services in the Asia-Pacific region.

Orbital Sciences Corporation	This is the company that developed the Pegasus, Taurus, and Antares launch vehicles. They are also a major manufacturer of medium sized spaececraft. Orbital Sciences has now merged with ATK to become Orbital ATK.
Orbital Express	This is the name of a project by the US Defense Advanced Research Projects Agency. This involved the capture of a client satellite to carry out a simulation of on-orbit servicing in low earth orbit.
Orbital ATK	This is the new name of the merged US aerospace corporations Orbital Sciences and Alliant Technologies (ATK).
OSI	Open Standards Integration model for telecommunications. This is the internationally accepted standard for modern digital communications that is the basis for Number 7 signaling and Integrated Digital Standards for Networking (ISDN). The seven layer OSI Model and the functionality of each layer is described in the chart provided below.

The seven layers of the OSI model used in ATM switching
Applications Level: Actual content such as e-mails, video images, voice, and data. This represents the highest level in the OSI Model
Presentation Level: Provides for such functions as encryption or data conversion
Session Level: Starts and stops sessions and creates the correct order
Transport Level: This ensures that the entire and complete message is delivered
Network Level: This routes information to a particular location based on network address
Transport Level: This routes data packets from node to node based on station addresses and the actual transmission mode
Physical Level: This actually provides the physical conduit to connect nodes in a network

(Chart courtesy of J. Pelton)

O3b	This is a global satellite organization that provides services optimized for Internet-based telecommunications using a constellation of Medium Earth Orbit satellites from the equatorial plane. Its name stands for the "Other Three Billion" or the population of planet earth that lives in the equatorial region which are largely represented by developing economies. This system is operated by the SES Global organization of Luxembourg. This system has many major investors that in include SES, Liberty Media, and Google. (See also One Web)

Paksat	This is the Pakistan domestic satellite communications network.
PDOP	Position Dilution of Precision
Pegasus Launch Vehicle	This is the aircraft launch small launch vehicle developed by Orbital Sciences Corporation (now known as Orbital ATK)
Perigee	This is the nadir or lowest point in an orbiting body. For a satellite in an elliptical Earth orbit, the perigee would represent the point closest to Earth.
Phoenix Project of DARPA	This is a project of the US DARPA to be able to simulate capture, servicing, and other functions in GEO Orbit. Note this term has been applied in other projects such as a French spaceplane.
Pixel	This is the smallest element of an image that can be individually processed in a video display image as captured by a multispectral remote sensing satellite. This is what defines the spectral resolution of a remote sensing satellite image.
PLB	Personal Locator Beacon. This is used by stranded pilots or marooned people at sea to send an alert signal to SARSAT-COPAS.
Plug Nozzle	This is a circularly configured series of rocket engine nozzle exhaust that are expelled against a wedge to create a specialized rocket exhaust pattern.
PMT	Platform Messaging Transceiver (PMT) that operates via the Argos II satellite system.
PNT	Precision Navigation and Timing (PNT) satellite systems. The best known satellites of this type are the GPS satellite network of the US and the GLONASS system of Russia. There are now a number of these satellite systems deployed or being deployed by Japan (Quasi-Zenith), China (Biedou and COMPASS), Europe (Galileo), and India (Indian Regional Navigational Satellite System).
POES	Polar Orbiting Environmental Satellite system.
PPS	Precise Positioning Service.
PRN	Pseudo Random Noise.
Proton Launch Vehicle	This is the Russian Launch system. It is offered through the International Launch Services to the US launch market.
PRN#	Pseudo Random Noise number.
PPT	Platform Transmitter Terminals that operate with the Argos network.
QAM	Quadrature Amplitude Modulation. A sophisticated modulation technique that allows encoding of information

	based on a four-level range of variable amplitudes. This can be used in low interference satellite communications transmission links to effectively encode information with great density in terms of bits per Hz transmitted
Quazi Zenith Orbit/	This orbit that is also referred to as Figure 8 orbit is essentially a Geo orbit that this rotated 45° from the equatorial orbital plane.
Quasi Zenith Satellite System of Japan	This has been used by Japan for its combined navigational and mobile satellite communications satellite network.
Q/V Band	This is the 48 GHz/38 GHz band that is allocated to satellite communications. This is a difficult band to utilize because of the difficulty of manufacturing equipment to operate at these extremely high frequencies and small bandwidths, but also because precipitation attenuation is very difficult to overcome at these extremely high frequencies well up into the millimeter wave band.
Radar Sat	This is the name of the Canadian radar sensing satellite. Radarsat 1 was the first in the series that was followed by Radarsat 2.
RF	Radio frequency. This is the part of the electro-magnetic spectrum which extends from low frequency radio emissions up through the terahertz frequencies and ends with infrared and light frequencies
Radiometric resolution	This is the recorded information that provides what might be called the relative “brightness” of the images collected.
Rain attenuation	The distortion of satellite transmission that occurs during heavy rainfall that occurs in the higher frequencies used for satellite communications such as Ku and Ka bands.
Range Safety Officer	This is the official in charge of the safety of a launch range with the responsibility to decide on the abort of a launch or the active destruction of a launch vehicle if safety considerations so dictate.
Roscosmos	This is the name of the Russian space agency. It is responsible for Russia’s participation in the International Space Station and a wide range of space technologies and space application satellite systems.
RTK	Real-time kinematic for the Global Positioning Satellite (GPS) system.
RTLS	Return to Launch Site. This is a term that covers both a commercial launch vehicle returning to its original launch site and the case of a reusable launch vehicle shedding auxiliary rocket launchers and return to the launch site in the case of an aborted mission.
RLV	Reusable Launch Vehicle

Rockot Launch Vehicle	This is one of the Russian launch vehicles that derives from an intercontinental ballistic mission (ICBM) and is also sometimes spelled ROKOT. It has limited capacity of launching about 200 kg into low earth orbit.
SA	Selective Availability. This is the ability to intentionally degrade the accuracy of determination of the civilian GPS signal. Under the Executive Order of President Clinton, it was directed that selective availability would not be implemented.
SARSAT	The Search and Rescue Satellite System.
SARSAT-COSPAS	This is the combined international satellite system that is used for search and rescue. See also Cospas.
Satellite Commands	These are transmitted instructions to a satellite to alter its orbit, flip a switch, or otherwise change some aspect of its operation or to correct a problem that has been detected
Satellite Constellations	A configured network of satellites usually in low earth orbit (typically 50–70 satellites) or in medium earth orbit (typically 12–18 satellite) to provide communications, precise navigation and timing services, remote sensing or meteorological satellite services.
SatMex	This is the domestic satellite system for Mexico.
SBAS	Satellite-Based Augmentation System.
SES	This is very large satellite communications company headquartered in Luxembourg, and it currently operated a fleet of 54 satellites for direct broadcast video services as well as fixed satellite services and Internet data services. It is a part owner and operator for the O3b network that it owns in partnership with Google, Liberty Media, and other large organizations. Its web page indicates that it connects to over one billion people worldwide. Originally founded in 1985 as Société Européenne des Satellites, the company was renamed SES Global in 2001 and in 2006 renamed simply SES. (See O3b)
Shavit	This is an expendable launch vehicle developed by Israel that with the Shavit 2 can lift up to 800 kg into polar orbit.
Sierra Nevada Corporation	This is one of the major players in new commercial space transportation systems. They have acquired SpaceDev that developed the Dreamchaser space plane and were until 2014 a finalist to provide astronaut transportation to the International space station for NASA under the COTS competition. Sierra Nevada also designs and builds small spacecraft.
Sinosat	This is one of the domestic satellite systems that provides telecommunications services in China. Also see Chinasat.

Skybox	A remote sensing constellation of low earth orbit satellites that was built using off the shelf components and was deployed to give close to real-time updated information. This system has now been acquired by Google.
Sky Television	This is the operator of a direct broadcast satellite that provides service to all of the United Kingdom as part of the News Corporation holdings.
Skyterra	This is the very large MSS satellite with the 22 m multibeam deployable antenna that was part of the planned Light Squared mobile satellite network for the USA with ancillary terrestrial component. Light Squared went bankrupt when it lost its terrestrial mobile frequencies due to a change in FCC authorization.
SNAS	Satellite Navigation Augmentation System (This is the term used in the Chinese PNT Systems.)
Solidaridad	This the name of the direct broadcast satellite system for Mexico.
Soyuz Launch Vehicle	This is the human-rated Russian launch system that together with the Progress system transports astronauts to the International Space Station. The Soyuz can be used for launch to Leo, Meo or Geo orbits for application satellites.
Spaceway	This was the name that was given to three Ka-band high throughput satellites that were designed by Hughes Communications to provide broadband services using advanced broadband digital services and on-board processing as developed on the ACTS experimental satellite. Over time it was determined that two of these satellites would be converted to support direct broadcast satellite services to support the DirecTV BSS service. Spaceway 3 was retained by Hughes Network Services to support HughesNet services to provide broadband IP services to remote areas or areas not adequately served by terrestrial broadband networks. Hughes Communications and HNS have now been acquired by Echostar.
SpaceX	The Space Exploration Technologies Corporation that launches the Falcon 1 and Falcon 9 launch vehicles. This company is also planning the launch of a large-scale constellation of small satellites to support Internet optimized services.
Spatial Resolution	In remote sensing, this refers to how much detail is captured in terms of pixels per image. This means what detail can be clearly seen – a forest, a tree, a limb, or a leaf.
Spectral resolution	In remote sensing, this refers to what bandwidth is the data collected within the electro-magnetic spectrum. Was the

	image “seen” in the visible spectrum, or above it in the ultra-violet, below it in infra-red. The more precise the bandwidth range, the more information can be revealed.
Spectrum	This is the measure of bandwidth that is used for various satellite applications. The radio wave spectrum allocated for commercial satellite communications, for instance, is typically either 500 MHz or 1,000 MHz across. One practical problem is that the ITU that is responsible for allocation of radio spectrum for practical or scientific use has divided the world into three regions and the allocations can be and indeed are different for different regions of the world.
Spot Image	This French company is one of the oldest and most well-established providers of commercial remote sensing imaging and thus is the main competitor to DigitalGlobal Corporation. This organization was initially established with support from the French Space Agency (CNES)
SPS	Standard Positioning System. This is the GPS civilian course acquisition (C/A) code signal.
Sputnik	The name of the first Soviet satellite that was launched in October 1957 that began human uses and exploration of outer space.
SRB	Solid Rocket Booster. This acronym particularly applies to the Solid Fuel Rockets for the Space Shuttle.
SSAS	This is the Argos-based Ship Security Alert Systems (SASS)
SSL	Space Systems/Loral has been one of major manufacturers of applications satellites dating back to its formative days as Philco Ford. It has particularly designed and built many communications satellites over the years. In November 2012, Loral Space and Communications sold Space Systems/Loral to McDonald Dettwiler Associations for approximately $1 billion, and this organization is now known simply as SSL.
SSO	Sun Synchronous Orbit. This orbit typically has a 90-min period and maintains constant visibility with the sun when it is not behind the earth in its orbits. This is a – particularly valuable orbit for multispectral imaging or hyper-spectral imaging for remote sensing applications.
SSPTS	Station-Shuttle Power Transfer System.
Sthil Launch Vehicle	This is a three stage launch vehicle that is a converted Submarine Launched Ballistic Missile that is used for launching artificial satellites into orbit. It is based on the R-29RM designed by State Rocket Center Makeyev and

	related to the Volna Launch Vehicle. It is notable as the only launch vehicle that has put a satellite into orbit from a submarine launch.
Store and Forward Satellites	Satellite systems that relay data up to satellite that then stores it and then downloads it to the designated location at a later time. The relay times from one location to another depends on the number of satellites in the network and their orbital altitudes. See also business to business data (B2B) relay.
STS	Space Transportation System. This is the formal name that NASA used to designate the Space Shuttle program.
SV	Space Vehicle
Synthetic Aperture Radar Sensing	This is a radar remote sensing system that collects data simulating the equivalent of a much larger aperture radar system.
TAI	Temps Atomique International or International Atomic Time. This is the basis for Coordinated Universal Time (UTC).
Taurus Launch Vehicle	This is the larger launch vehicle developed by the Orbital Sciences Corporation to provide a greater lift capacity than the Pegasus. The Taurus was the basis for the upgraded capacity launch vehicle developed by OSC for the Nasa Commercial Orbital Transportation System (COTS) development for cargo deliveries to the International Space Station, the Antares launcher together with the Cygnus capsule. (See Antares)
TDMA	This is one of the most commonly used digital multiple access systems used in satellite communications and all forms of digital communications. This acronym stands for time division multiple access.
Teledesic	This was a proposed mega-Leo constellation of nearly a thousand small satellites to provide internet services. This project that was backed by Edward Tuck, Bill Gates, and McCaw Communications went bankrupt and was never deployed. Current initiatives such as One Web are currently seeking to deploy a similar type system.
Telemetry	This is to obtain data from a satellite as to its in-orbit bus and power operations and to detect errors or component failures.
Telenor	Over the past two decades, Telenor Satellite Broadcasting has established itself as a major satellite operator in key target markets throughout Europe and the Middle East with its network of Thor Satellites that includes its more recently launched Thor 7 Satellite. Telenor, which is based

	in Oslo, Norwa, also has a longer term indefeasible right of use on spot beam transponders on the Intelsat 10 satellite. It has leased capacity on one of its Thor satellites to SES for service in Sweden.
Telesat	The name of the Canadian satellite system that has now expanded to provide global coverage. It has now deployed a number of high throughput satellites in other regions of the world. Loral Space and Communications Inc. of the USA owns 62 % of Telesat. Satellite services are provided via the Anik satellites, the Nimiq direct broadcast satellite to rural and remote areas, and a number of the Telstar satellites provide services through the partnership with Loral Space and Communications Inc. of the USA.
Temporal resolution	In remote sensing, this refers to exactly when the image by the sensing satellite was actually taken and optically or electronically recorded.
Terrestar	This is the satellite with the large 18 m deployable satellite antenna. This satellite was deployed as part of a hybrid mobile satellite system with ancillary terrestrial component for mobile cellular services in the USA.
Thaicom	This is a domestic satellite designed to operate in Thailand and adjacent countries. The company that launched the Thaicom satellites also started and launched for the Asian regional service area the high efficiency IPStar satellites that were as their name suggested optimized for IP-based broad band services. These satellites operated with high-efficiency coder/decoder (CODEC) technology. (See IP Star.)
Thuraya	This is the Geo orbit satellite network that provides land mobile satellite services primarily in the Middle East, but also in other parts of Africa, Asia, and Europe. The large multibeam antenna on Thuraya satellites allows high power spot beams of high intensity and thus can operate to hand held user terminals.
THz	TeraHertz or one trillion cycles per second
Timation	This is the name of the US Navy system that was deployed prior to the GPS or Navstar system.
TIROS	This was an early experimental meteorological satellite by NASA.
TOE	Time of Ephemeris.
Tracking	This is to determine precisely where a satellite is in orbit.
Transit	The name of the first US satellite positioning system.
Transponder	This is the key technical component of a communications satellite that involves the electronics for translation of an

	uplink frequency to a downlink frequency for transmission back to an earth station. Transponders are how the capacity of a communications is frequently stated. Typically transponder frequency ranges are 36 MHz, 72 MHz, but can be broadband up to 125 MHz or even larger. In the early days of satellite communications, a television channel could require an entire 36 MHz transponder, but in today's world of digital transmission and digital compression techniques some 18 digital television channels, each at 4 megabits/s, can be transmitted through a single 72 MHz transponder.
Tsikada	The name of the first positioning satellite system deployed by the Soviet Union.
TTC&M	This refers to Tracking, Telemetry, Command, and Monitoring of in-orbit satellites. Tracking is to determine precisely where a satellite is in orbit. Telemetry is to obtain data from a satellite as to its in-orbit operations and to detect errors or component failures. Commands are transmitted instructions to a satellite to alter its orbit, flip a switch, or otherwise change some aspect of its operation or to correct a problem that has been detected. Monitoring is to follow the on-going operation of a satellite to detect if its level of service is meeting specifications or if it is somehow substandard or is being subjected to interference.
TTFF	Time to First Fix.
Turksat	This is the Turkish domestic communications satellite system.
ULA	The United Launch Alliance that is a joint project of Boeing and Lockheed Martin.
United Nations	This is the public international organization formed after World War II to address all matters related to international cooperation and peacekeeping. All so-called specialized international organizations related to various functions come under the United Nations structure. Entities in the UN structure that have a particular relationship to space application areas include **the International Telecommunication Organization (ITU), the International Maritime Organization (IMO), the International Civil Aviation Organization (ICAO), the United Nations Office on Disarmament Affairs (UNODA), and the United Nations Committee on the Peaceful Uses of Outer Space (COPUOS).** Other entities such as the **International Bank for Reconstruction and Development**

	(IBRD) and the International Monetary Fund (IMF) become involved in terms of financing space projects, and the **United Nations Educational, Scientific and Cultural** Organization (**UNESCO**) become involved due to educational and scientific matters. Other organizations that have some sort of functional relationship in terms of space applications include the **World Health Organization** (i.e., using space for health and medical care), **the UN Environmental Programme and the World Meteorological Organization** due to their reliance on environmental and weather satellites.
United Nations COPUOS	This is the United Nations Committee on the Peaceful Uses of Outer Space that is headquarters in Vienna, Austria. It is also composed of a Technical SubCommittee and a Legal SubCommittee, each of which meets once a year in Vienna, Austria. There is also a Working Group on the Long Term Sustainability of Outer Space Activities.
UNEP	The United Nations Environmental Programme.
UNESCO	United Nations Educational, Scientific and Cultural Organization.
UNODA	United Nations Office of Disarmament Affairs that is headquartered in Geneva, Switzerland.
USAT	Ultra Small Aperture Terminal. This is typically under 0.5 m in diameter.
USGS	United States Geological Survey.
UTC	Coordinated Universal Time.
VDP	Vertical Dilution Precision.
Vega Launch Vehicle	This is the new European developed light weight launcher that is launched from the Guyana launch facility. See Appendix "▶ Major Launch Systems Available Globally."
Via Sat	This corporation that is headquartered in San Diego, California, USA, started out as VSAT manufacturer now operates the largest high throughput satellites Via Sat-1 and Via Sat-2 and also owns the Wild Blue Satellite – all of which operate in the Ka-band. While Via Sat-1 and Wild Blue are essentially for US-based satellite services, Via-Sat-2 satellites will over services on an international basis.
Vivasat	This is a company that has developed technology associated with on-orbit repair, servicing, and refueling of satellites.
Volna Launch Vehicle	This is a Russian launch vehicle that is closely akin to the Shtil Launcher in that is a three-stage launcher based on a

	submarine launched ballistic missile. (See Shtil launch vehicle)
Vulcan	A new core launch vehicle under development by the United Launch Alliance (ULA) for the 2019 time period.
VSAA	Very Small Aperture Antenna.
VSAT	Very Small Aperture Terminal.
VTS	Vehicle Tracking System.
WAAS	Wide Area Augmentation System (WAAS) that is deployed in the USA though the use of ground systems to augment the GPS network.
WADGPS	Wide Area Differential Global Positioning System.
W Band	This band with allocations for commercial satellite communications at 60 GHz in the millimeter wave band is even more difficult to exploit than the Q/V bands. See also.
WCRP	World Climate Research Program.
WGS	World Geodetic System.
WGS84	World Geodetic System Version 1984
Wild Blue	These are Ka-band satellites designed for service in the USA that were started by the Ka Star Corporation. These satellites are now owned by Via Satellite.
WINDS	This was an experimental satellite of JAXA and the National Institute of Information and Communications Technology (NICT) to experiment with on-board processing and switching and testing of 1 gigabit/s broadband digital channels on narrow Ka-band beams using a phased-array satellite antenna. This tested similar technologies to that tested by NASA ACTS satellite, plus the phased array antenna.
WMO	World Meteorological Organization.
Worldspace	This is a direct audio broadcasting satellite for the African region. It has gone through a bankruptcy process and is now owned by Yazmi Holdings, Maryland, USA. Only Afristar was deployed although a global system of three satellites including also Caribstar and Asiastar was at one time anticipated.
WRC	World Radio Conference. This is the plenipotentiary meeting of the members of the International Telecommunication Union that decides on the allocation of radio frequencies and other important issues involving radio communications, including satellite communications.
XM Radio	This system that is now merged with Sirius Radio provides digital audio broadcast services, including news, music, and security services in North America.

Yahsat	Al Yah Satellite Communications Company (Yahsat) is a private joint stock company that is fully owned by Mubadala, the investment arm of the Government of Abu Dhabi, the capital of the United Arab Emirates. Yahsat provides service largely to middle eastern countries but does service as far south as the Republic of South Africa and over as far as Afghanistan and Pakistan.
Yamal	Yamal is a communication satellite developed by Gazprom Space Systems for Russian Direct-To-Home television. This system carries Lyngsat television.
Zenit	This is the Ukrainian launch vehicle that includes the Zenit 1 and Zenit 2.

The World's Launch Sites

Arthur Norman Guest and Joseph N. Pelton

Contents

Introduction

There are over three dozen launch sites around the world. With the growth of the newly emerging spaceplane transportation and space adventures business, there are also a growing number of "spaceports" designed for commercial liftoffs and landings. While there are an increasing number of commercial spaceports in the United States,

A.N. Guest
International Space University, San Francisco, CA, USA
e-mail: guest.arthur@gmail.com

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_88

there are also numerous other sites under consideration around the world in locations such as Singapore, Barcelona, the United Arab Emirates, Malaysia, and various sites in Europe such as at least one in the United Kingdom and one in Sweden.

Today, perhaps the busiest launch sites for various types of spacecraft are at Cape Canaveral in California (Eastern Range) and Vandenberg (Western Range), the Baikonur and Plesetsk Cosmodromes in Russia, the Kourou launch facility of the French Space Agency (CNES) and the European Space Agency in French Guiana that now also supports Soyuz and Vega, the Kagoshima and Tanegashima launch facilities in Japan, the Jiuquan and Xichang launch ranges in China, and the Satish Dhawan launch facility in India.

The current concentration on official governmental launch sites will undoubtedly change as more and more commercial spaceports evolve in the coming years. Many of the commercial spaceports being licensed in the United States by the FAA-AST are for horizontal takeoff and landing and are intended only for the operation of suborbital flights and thus do not support commercial launch of application satellites to Earth orbit.

For years, the space programs of the United States and Russia were comparable in size and by far the world’s largest space operations. But this has changed as European (both national and regional), Chinese, Japanese, Indian, and even Israel space programs have matured. Further, there are many emerging space programs in Australia, Argentina, Brazil, Canada, Iran, Israel, North and South Korea, Pakistan, Taiwan, and the Ukraine. The following provides background information on launch sites around the world that provide launch operations or have done so in the past. Not all licensed commercial spaceports are included, particularly those that only provide suborbital flights.

Algeria

The French government for the period 1947– 1967 operated a special weapons test center from Hammaguira, Algeria, near the Moroccan border. This facility is no longer operational. From 1961 to 1965, France also operated a rocket and weapons test facility near Reggane, Algeria, which is in the Saharan Desert. Essentially all French space launcher operations now operate from Kourou.

Australia

Site Name: Spaceport Australia
Location: Woomera, Australia (Latitude 31.1°S Longitude 136.6°E)
Launch Vehicles Supported: Kistler K-1 (canceled)

Originally known as the Woomera Rocket Range, this site was originally created to support British and Australian missile operations. In 1967, WRESAT, an Australian science satellite, was launched on a US Redstone rocket from Woomera. The original launchpads have been dismantled, but the site is being redeveloped as Spaceport Australia to support future commercial launch opportunities. Spaceport Australia was to be used to support Kistler Aerospace's K-1 launch vehicle. The development of the K-1 rocket was canceled in 2007. The Australian Space Council is currently supporting future launch operations from this site. Spaceport Australia is located in the desert and well positioned for polar orbit launches.

Site Name: Asia Pacific Space Center (proposed)
Location: Christmas Island, Australia (Latitude 10.4°S Longitude 105.7°E)
Launch Vehicles Supported: Aurora (proposed)

The Asia Pacific Space Center (APSC) was proposed to support the Aurora rocket, a new Russian/Australian launch vehicle derived from the Soyuz launch vehicles. The site has been proposed to be located in the Indian Ocean on Christmas Island.

Brazil

Site Name: Alcantara Launch Center
Location: Alcantara, Brazil (Latitude 2.3°S Longitude 44.4°W)
Launch Vehicles Supported: VLS-1 (proposed)

Located near Sao Luis on the Atlantic Ocean coastline, this site supports launches by the proposed Brazilian VLS-1 launch vehicle as well as the Sonda 3 and Sonda 4 sounding rockets. The development of the VLS-1 launch vehicle has been delayed since an accident on the launchpad in 2003. The potential of using the Alcantara Launch Center to support other commercial launch vehicles has been proposed, although no immediate plans exist. The launch center's location (2° south of the Equator) offers it substantial advantage for geosynchronous satellite launches.

China

Site Name: Jiuquan Satellite Launch Center (JSLC)
Location: Gobi desert, Inner Mongolia (Latitude 40.6°N Longitude 99.9°E)
Launch Vehicles Supported: Long March 2C/2D/2 F & Long March 4B/4C

On April 24, 1970, the People's Republic of China became the fifth nation to launch an artificial satellite into Earth orbit. This satellite was named Mao-1 in honor of Chairman Mao Tse Tung and this spacecraft was launched into orbit by a Long March 1 vehicle from the Jiuquan Satellite Launch Center in Inner Mongolia. Jiuquan Satellite Launch Center (also known as Shuang Cheng Tzu) was the first launch complex to be built by China, and it is located north of Jiuquan City in the Gobi desert 1,600 km west of Beijing, China. Jiuquan has been limited to southeastern launches into 57°–70° orbits in order to avoid overflying Russia and Mongolia. Jiuquan is used for recoverable Earth observation and microgravity missions, but due to the site's isolated geographical location, most Chinese commercial flights takeoff from other spaceports.

On October 15, 2003, a Long March 2 F rocket with the spacecraft Shenzhou 5 was launched from the Jiuquan Satellite Launch Center. Inside the spacecraft was Yang Liwei, China's first astronaut. The flight made China the third nation to send a person into space.

Site Name: Xichang Satellite Launch Center (XSLC)
Location: Xichang City, China (Latitude 28.3°N Longitude 102.0°E)
Launch Vehicles Supported: Long March 2C & Long March 3A/3B/3BE/3C

This site offers better access to geostationary orbits than Jiuquan since it is much closer to the equator. It was built 65 km north of Xichang City in 1978 but its first orbital launch was not until 1984. Xichang launches Long March 2C and Long March 3-series launch vehicles. A number of commercial communications satellites on behalf of a number of operators have been launched from this site.

Site Name: Taiyuan Satellite Launch Center (TSLC)
Location: Shanxi Province, China (Latitude 37.5°N Longitude 112.6°E)
Launch Vehicles Supported: Long March 2C/2D and Long March 4B/4C

This center was started in 1968 as a test base for missiles and rockets that were too large to fly from Jiuquan. The Taiyuan Satellite Launch Center, also known as Wuzhai, opened its single space launchpad in 1988 for launching Long March 4 launch vehicles. These launches have been primarily for placing remote sensing, meteorological, and reconnaissance satellites into polar orbit. In 1996 and 1997, Long March 2C rockets carried Iridium satellites into orbit.

Site Name: Wenchang Satellite Launch Center (WSLC)
Location: Hainan Island, China (Latitude 19.7°N Longitude 111.0°E)
Launch Vehicles Supported: Long March 5 (proposed)

The Wenchang Satellite Launch Center is being constructed to support the Long March 5 launch vehicles. The site was specifically chosen for its proximity to the equator to support launching satellites into geostationary orbit. The site is expected to be completed in 2013 in order to support launch of Long March 5 in 2014.

Europe

Site Name: Guiana Space Center (Centre Spatial Guyanais)
Location: Kourou, French Guiana (Latitude 5.2°N Longitude 52.8°W)

Launch Vehicles Supported: Ariane 5, Soyuz, and Vega. The Center is also being prepared for Ariane 6 which is still under development.

On December 24, 1979, the European Space Agency became the seventh entity to launch an artificial satellite into Earth orbit. The satellite was named CAT and it was launched by an Ariane 1 rocket from the Guiana Space Center (or Centre Spatial Guyanais, CSG). CSG is owned by the French national space agency, CNES. It is used by the European Space Agency (ESA) and its commercial space launch arm Arianespace to launch ESA's Ariane 5 launch vehicles. CSG is one of the most favorable sites for launching satellites into geostationary orbit since it is only 5° north of the equator. French Guiana's coastline permits launches into both equatorial and polar Sun-synchronous orbits with inclinations up to 100.5°. Hundreds of sounding rockets and balloons and space satellites have been launched from Centre Spatial Guyanais. Most recently CSG has been expanded to support the Russian Soyuz and Vega launch vehicles. The first Soyuz launch from Kourou was on October 21, 2011. The first Vega launch from Kourou was on May 6, 2013. Construction is currently underway to enable the support of launches of the Ariane 6 launch vehicle.

India

Site Name: Satish Dhawan Space Center (SHAR)
Location: Sriharikota Island, India (Latitude 13.9°N Longitude 80.4°E)
Launch Vehicles Supported: PSLV and GSLV

On July 18, 1980, India became the eighth nation to launch an artificial satellite into orbit. The satellite was named Rohini 1. It was launched on the Satellite Launch Vehicle (SLV) from Satish Dhawan. The Satish Dhawan Space Center is located on Sriharikota Island on India's east coast state of Andhra Pradesh. In India, all space activities, launcher development and satellite design and manufacture, are the responsibility of the Indian Space Research Organisation (ISRO) and its commercial partners and subsidiary organizations. ISRO has launched many space research and application satellites on its PSLV (Polar SLV) and GSLV (Geostationary SLV) vehicles. This launch site, that is only 14° north of the Equator, is well suited for geosynchronous launches.

India's first mission to Mars, the Mars Orbiter Mission, known as the Mangalyaan orbiter, was launched on a Polar Satellite Launch Vehicle on November 5, 2013, from the Indian Space Research Organisation's Satish Dhawan Space Center on Sriharikota Island.

Israel

Site Name: Palmachim Air Force Base
Location: Negev Desert, Israel (Latitude 31.5°N Longitude 34.5°E)
Launch Vehicles Supported: Shavit

On September 19, 1988, Israel became the ninth nation to launch an artificial satellite into orbit. The satellite was named Horizon 1 (Ofeq 1 in Israeli). It was launched on Shavit rocket from Israel's Palmachim Air Force Base. This location is still Israel's main spaceport. This site is south of Tel Aviv and is near the town of Yavne in the Negev Desert. Launches from Palmachim are done into a retrograde orbit in order to allow all launches to head west toward the Mediterranean Sea as opposed to heading east toward neighboring states.

Italy

Site Name: Broglio Space Center (San Marco Platform)
Location: Offshore near Malindi, Kenya (Latitude 2.9°S Longitude 40.3°E)
Launch Vehicles Supported: None (site not operational)

Italy's San Marco Range (renamed the Broglio Space Center in 2004) consists of a pair of platforms that float in Formosa Bay 3 miles off the coast of Kenya in the Indian Ocean. The San Marco platform served as the launchpad and the Santa Rita platform held the firing control blockhouse. The range started firing sounding rockets in 1966 and eight satellites were boosted to space from there by 1976. Italy used the offshore platform for another launch in 1988. This platform is not currently operational but could be employed for future launches. The position close to the equator makes this well suited for geosynchronous launches.

Japan

Site Name: Uchinoura Space Center (Kagoshima Space Center)
Location: Kagoshima, Kyushu Island, Japan (Latitude 31.2°N Longitude 131.1°E)
Launch Vehicles Supported: M-V (retired)

In February 1970, Japan became the fourth nation to launch an artificial satellite into Earth orbit. The satellite was named Ohsumi. This launch was on a Lambda 4S-5 rocket, and the Kagoshima site, which was constructed in 1962, was used to support this initial Japanese orbital flight. The site is built on the southern tip of Kyushu Island to support an eastward launch over the Pacific Ocean. This site was initially employed for the launch of sounding and meteorological rockets and then later used for scientific and application satellite launches. Japan's first six satellites were launched from the Kagoshima Space Center which has been called the Uchinoura Space Center since 2003 and the establishment of JAXA. The large M-V orbital rocket was first launched there in 1997. Over 20 orbital launches and much larger number of suborbital flights have been initiated from this site. The site is currently used to launch sounding rockets.

Site Name: Tanegashima Space Center (TNSC)
Location: Tanegashima, Japan (Latitude 30.4°N Longitude 131.0°E)
Launch Vehicles Supported: H-IIA and H-IIB

Japan's space agency (JAXA) operates the Tanegashima Space Center orbital launch site on the southeastern tip of Tanegashima Island some 1,050 km southwest of Tokyo. The complex's northern Osaki Launch Site can support H-IIA and H-IIB rockets. It also has static test facilities for liquid-fuel rocket engines. This site also houses the H-II Range Control Center.

Russian and CIS Launch Sites

Site Name: Baikonur Cosmodrome
Location: Tyuratam, Kazakhstan (Latitude 45.6°N Longitude 63.4°E)
Launch Vehicles Supported: Proton, Strela, Dnepr, Zenit, Rockot, and Cyclone 2

Sputnik 1, the world's first artificial satellite, was launched on October 4, 1957, from this site. The first development of this launch site was in the early 1950s in the Baikonur/Tyuratam area of Kazakhstan in central Asia. The launchpad from which Sputnik 1 launched, as well as from which Yuri Gagarin, the first man in space, was launched, was constructed in 1955. It is in an isolated area about 370 km southwest of Baikonur. It was renamed by the Kazakhstan Government after the close town of Tyuratam in the 1990s. However, the global space community still refers to it as Baikonur Cosmodrome. Baikonur is a large cosmodrome with nine launch complexes encompassing 15 launchpads. All of Russia's manned space flights and interplanetary probes are launched from the Baikonur Cosmodrome. Baikonur is the only cosmodrome capable of launching Proton, Zenit, Energia, and Tsyklon SL-11 space rockets. Launches headed due east would be the most efficient launch path, but this launch orientation is not used from Baikonur because of concerns that lower stages of the rockets might fall into China.

Site Name: Plesetsk Cosmodrome
Location: Arkhangelsk Oblast, Russia (Latitude 62.8°N Longitude 40.1°E)
Launch Vehicles Supported: Kosmos 3 M, Rockot, Soyuz, Start-1, Angara

The first of several pads were constructed at the Plesetsk Cosmodrome starting in 1957 to support both R7 and A-class intercontinental missile tests. The pads and various classes of ICBM rockets were moved to active duty in 1960. Initially, the Plesetsk Cosmodrome was the world's most utilized launch site. In time, the launches moved to the Baikonur/Tyuratam Cosmodrome that was build to support newer launch systems. Today, Plesetsk can support Kosmos 3 M, Rockot, Angara, and other launch vehicles. Plesetsk Cosmodrome is located in Russia at 62.8°N and 40.1°E, which allows the launch of communications satellites and surveillance satellites to polar and highly elliptical Molniya orbits. The Plesetsk Cosmodrome is currently being renovated and expanded for possible future operations.

Site Name: Svobodny Cosmodrome (not operational)
Location: Amur Oblast, Russia (Latitude 51.4°N Longitude 128.3°E)
Launch Vehicles Supported: Start-1 and Rockot

This is the most recently constructed Russian cosmodrome, started in 1996. This cosmodrome was built on a former missile site called Svobodny-18 and lies about 100 km away from the Chinese border. This site can support the launch of Start and Rockot launch vehicles.

In addition, Soyuz and Vega launches are also possible from the Guiana Space Center as noted earlier.

South Korea

Site Name: Naro Space Center
Location: Goheung County, South Jeolla (Latitude 34.4°N Longitude 127.5°E)
Launch Vehicles Supported: Naro-1

South Korea's new spaceport is operated by the Korea Aerospace Research Institute (KARI) to support the launch of the Naro-1 rocket, formally known as the KSLV 1. The early attempts to launch the Naro-1 rocket failed, but as of January 10, 2013, the NARO-1 was able to place a 300 kg satellite into low earth orbit.

United States of America

Site Name: Kodiak Launch Complex
Location: Kodiak Island, Alaska (Latitude 67.5°N Longitude 146°W)
Launch Vehicles Supported: Athena, Minotaur I, Minotaur IV, Taurus, and Taurus II (proposed)

This site is a commercial launch facility and because of its Northern Latitude is well suited for launching satellites into polar orbit. The launch site is located on the Narrow Cape, of Kodiak Island, Alaska, south of the city of Kodiak and some 400 km south of Anchorage. Kodiak Island is a volcanic peak in the ocean located some 48 km off shore in the Gulf of Alaska. This commercial spaceport is operated by the Alaska Aerospace Corporation.

Site Name: Cape Canaveral Air Force Station (CCAFS)
Location: Cape Canaveral, Florida (Latitude 28.3°N Longitude 80.3°W)
Launch Vehicles Supported: Falcon 9, Atlas V, and Delta IV

This facility is operated by the US Air Force and has been the site for many launches to support US space mission involved with national security. The CCAFS is an installation of the US Air Force Space Command 45th Space Wing that is headquartered at the adjacent Patrick Air Force Base. The CCAFS is the primary launch head of America's Easter Range and currently the facility has four launchpads capable of supporting launches. The facility is south-southeast of the Kennedy Space Center and the two launch systems are linked by bridges and roadways. There is a 10,000-ft runway at the CCAFS that is close to the launch complexes. It is available for airlift aircraft that deliver heavy and particularly large payloads associated with launches from the Cape.

Several major American space exploration "firsts" were launched from CCAFS. These launches include the first US artificial satellite the Explorer I in 1958, the first US astronaut launches with Project Mercury, and the follow on Gemini launches in the 1961–1964 time period. It has also supported other crewed systems launches as well as flights to explore the solar system.

Site Name: Kennedy Space Center (KSC)
Location: Merritt Island, Florida (Latitude 28.5°N Longitude 81.5°W)
Launch Vehicles Supported: Space Shuttle (retired)

This site has served as the location for processing, launching, and landing space shuttles and their payloads, including components of the International Space Station. It has historically served as the launch site for the Apollo launches. This launch center is essentially designed to support crewed missions to places beyond Earth but also has supported other launch missions as well. KSC is located on Merritt Island. It is adjacent to the US Air Force launch facilities known as the Cape Canaveral Air Station and the newly formed commercial Space Florida launch facility. Kennedy was built first to support the Apollo lunar landings of the 1960s. After the last Apollo lunar launch in 1972, launch complex 39 supported Skylab space station in 1973–1974, then the Apollo-Soyuz Russian-American linkup in space in 1975, and now space shuttles since the late 1970s.

Site Name: The Mojave Spaceport
Location: California, USA (Latitude 35.0°N Longitude 118.2°W)
Launch Vehicles Supported: Various horizontal takeoff spaceplanes

The Mojave Spaceport was the first commercial spaceport licensed by the US Federal Aviation Administration Office of Commercial Space Transportation (FAA-AST). This facility is limited to the horizontal takeoff and landing of spaceplane systems and reusable spacecraft. This facility is located 160 km north of Los Angeles, California. Commercial operations are maintained at this site by Scaled Composites, Interorbital Systems (IOS), Orbital Sciences, and XCOR Aerospace.

Site Name: Spaceport America (formerly known as the Southwest Regional Spaceport)

Location: Las Cruces, New Mexico (Latitude 32°N Longitude 107°W)
Launch Vehicles Supported: Space Ship Two

This especially designed and purpose-build facility, located 72 km north of Las Cruces, New Mexico, was constructed in close proximity to the White Sands Missile Range. This facility is designed to support the spaceplane operations of Virgin Galactic and the Space Ship Two spaceplanes. The carrier vehicle will takeoff and carry Space Ship Two to altitude where the two vehicles will separate. Space Ship Two will land after a suborbital parabolic flight to a maximum height of 120 km. The alternative flight center is the Mojave Spaceport. The actual operation of this facility has been delayed due to the fatal accident with the Space Ship Two test flight on October 31, 2014.

Site Name: Vandenberg Air Force Base (VAFB)
Location: Lompoc, California (Latitude 34.4 °N Longitude 120.35 °W)
Launch Vehicles Supported: Delta II, Delta IV, Atlas V, Minotaur I, Minotaur IV, Taurus, Pegasus, and Falcon 1

This facility, which was first established in 1941 during World War II, is operated by the US Space Command's 30th Space Wing. It is located on the central Pacific coastline 20 km north of Lompoc, California, and 240 km northwest of Los Angeles. Vandenberg is the only military installation in the United States from which unmanned government and commercial satellites are launched into polar orbit. VAFB sends satellites to polar orbits by launching them due south. The base also test fires America's intercontinental ballistic missiles (ICBMs) westward toward the Kwajalein Atoll in the Marshall Islands. Vandenberg operates the Western Range tracking network, which extends all the way into the Indian Ocean to meet the Eastern Range tracking network. Western Range sites are on the California coast and downrange in the Hawaiian Islands. Vandenberg was to have provided a base for space shuttle launches on high inclination missions, but no shuttles ever have flown from there. California Spaceport is a colocated commercial launch facility at Vandenberg Air Force Base, largely in support of the U.S. Air Force, and primarily to support launches to polar orbit. There is also an adjacent commercial spaceport facility at this location as well.

Site Name: Wallops Flight Facility (WFF) and adjacent Mid-Atlantic Spaceport.
Location: Wallops Island, Virginia (Latitude 37.8°N Longitude 75.5°W)
Launch Vehicles Supported: Pegasus, Minotaur I, Minotaur IV, Taurus, Antares, and Cygnus launch system.WFF is officially a part of the NASA Goddard Space Flight Center. Wallops Flight Facility is the main launch site for the vehicles owned and operated by Orbital Sciences Corporation ATK, although they now use the adjacent Mid-Atlantic Spaceport. WFF also supports numerous sounding rockets and balloon launches.
Site Name: Ronald Reagan Ballistic Missile Defense Test Site

Location: Omelek Island, Kwajalein Atoll, Marshall Islands (Latitude 9.3°N Longitude 167.4°E)
Launch Vehicles Supported: Falcon 1, Falcon 9

The Reagan Test Site has been used by SpaceX Corporation, as a spaceport for launching its Falcon 1 launch vehicles in the 2006 time frame. The Falcon 9 and the Falcon 9 Heavy are now the prime launchers by the SpaceX Corporation. This site is beneficial for geostationary satellite launches when compared to other US launch sites due to its proximity to the Equator.

Multinational Sites

Site Name: Sea Launch, Odyssey Platform
Location: Mobile, typically from Latitude 0°N to Longitude 154°W
Launch Vehicles Supported: Zenit 3SL

Sea Launch is controlled by a consortium of aerospace companies and is headquartered at Long Beach, California. The consortium was headed by the Boeing Corporation until Sea Launch claimed bankruptcy in June 2009. Since emerging from bankruptcy in 2010, Sea Launch's majority owner is Energia Overseas Limited. Sea Launch, which has been in existence since 1999, operates its floating launch platform from a position near Kiribati (popularly known as the Christmas Islands). This position, almost exactly on the equator, is ideal for the commercial launch of satellites that are to be deployed into a geosynchronous orbit. Sea Launch operations begin at the ship's home port at Long Beach, California, where a satellite is prepared and loaded onto Sea Launch Commander, the assembly and command ship. Then, a Zenit rocket in a horizontal position is transferred to an environmentally controlled hangar on Odyssey, the partially submersible, self-propelled, launch platform. Odyssey once was a North Sea oil drilling platform. It is 436 ft. long and 220 ft. wide. After sailing to a launch point in the Pacific, a rocket is rolled out onto Odyssey's deck, erected, and fueled with kerosene and liquid oxygen. The rockets fired from Sea Launch typically carry telecommunications satellites to space on their way to geostationary orbit.

Other Commercial Spaceports

There are a number of other commercial spaceports in the U.S. at various levels of development (Three in Texas, plus commercial sites in Florida, California, Oklahoma (largely defunct), Mid-Atlantic, Wisconsin, etc.) as well as other sites around the world such as Kiruna, Sweden, Barcelona, Spain, the United Arab Emirates, Singapore, the United Kingdom, etc. It is recommended that one consult various websites, the FAA Office of Commercial Spaceflight as well as Joseph N. Pelton and Peter Marshall, *Launching into Commercial Space* (2016) AIAA, Reston, Maryland, with regard to the status of these various spaceport initiatives that are more fluid and change over time as to their current active status and regulatory licensing approvals.

Major Launch Systems Available Globally

Arthur Norman Guest and Joseph N. Pelton

Contents

A.N. Guest
International Space University, San Francisco, CA, USA
e-mail: guest.arthur@gmail.com

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_89

Introductory Note

This listing of launch vehicles and their performance represents current information at time of publication. There are constant upgrades and changes to launcher performance. It is thus recommended that one consult current website listings for the latest information on launch vehicles.

China

Long March 2C

Height: 43 m
Lift-off Mass: 245 t
Lift-off Thrust: 2,962 kN
Fairing Diameter: 3.35 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
LEO Capability: 3,850 kg
SSO Capability: 900 kg
Launch Site(s): JSLC/TSLC/XSLC

Long March 2C/CTS1

Height: 43 m
Lift-off Mass: 245 t
Lift-off Thrust: 2,962 kN
Fairing Diameter: 3.35 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: Solid Propellant
SSO Capability: 2,100 kg
Launch Site(s): JSLC/TSLC/XSLC

Long March 2C/CTS2

Height: 43 m
Lift-off Mass: 245 t
Lift-off Thrust: 2,962 kN
Fairing Diameter: 3.35 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: Solid Propellant
GTO Capability: 1,250 kg
Launch Site(s): JSLC/TSLC/XSLC

Long March 2D

Height: 41 m
Lift-off Mass: 250 t

Lift-off Thrust: 2,962 kN
Fairing Diameter: 3.35 m/3.80 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
LEO Capability: 4,000 kg
SSO Capability: 1,150 kg
Launch Site(s): JSLC/TSLC

Long March 2 F

Height: 58.3 m
Lift-off Mass: 497.9 t
Lift-off Thrust: 5,923 kN
Fairing Diameter: 3.80 m/4.20 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
LEO Capability: 8,600 kg
Launch Site(s): JSLC

Long March 3A

Height: 52.5 m
Lift-off Mass: 241 t
Lift-off Thrust: 2,962 kN
Fairing Diameter: 3.35 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: LOX/LH_2
LEO Capability: 2,600 kg
Launch Site(s): XSLC

Long March 3B

Height: 54.8 m
Lift-off Mass: 425.8 t
Lift-off Thrust: 5,923 kN
Fairing Diameter: 4.00 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: LOX/LH_2
LEO Capability: 5,100 kg
Launch Site(s): XSLC

Long March 3BE

Height: 56.3 m
Lift-off Mass: 456 t
Lift-off Thrust: 5,923 kN
Fairing Diameter: 4.00 m/4.20 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: LOX/LH_2
LEO Capability: 5,100 kg
Launch Site(s): XSLC

Long March 3C

Height: 54.8 m
Lift-off Mass: 345 t
Lift-off Thrust: 4,443 kN
Fairing Diameter: 4.00 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: LOX/LH_2
LEO Capability: 3,800 kg
Launch Site(s): XSLC

Long March 4B

Height: 48.0 m
Lift-off Mass: 250 t
Lift-off Thrust: 2,962 kN
Fairing Diameter: 2.90 m/3.35 m/3.80 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: N_2O_4/UDMH
SSO Capability: 2,230 kg
Launch Site(s): JSLC/TSLC

Long March 4C

Height: 48.0 m
Lift-off Mass: 250 t
Lift-off Thrust: 2,962 kN
Fairing Diameter: 2.90 m/3.35 m/3.80 m

Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: N_2O_4/UDMH
SSO Capability: 2,950 kg
Launch Site(s): JSLC/TSLC

Long March 5E

Height: 62.0 m
Lift-off Mass: 800 t
Lift-off Thrust: 10,640 kN
Fairing Diameter: 2.90 m/3.35 m/3.80 m
Booster Propellant: LOX/Kerosene
Stage-1 Propellant: LOX/LH_2
Stage-2 Propellant: LOX/LH_2
LEO Capability: 25,000 kg
GTO Capability: 14,000 kg
Launch Site(s): WSLC
Note: Long March 5E is the largest of six variations being developed (LM-5A to LM-5 F). The capability to launch 25,000 kg to low earth orbit and 14,000 kg to geosynchronous orbit makes the Long March 5E generally competitive with the largest launchers in the world.

Europe

Ariane 5ECA

Height: 59.0 m
Lift-off Mass: 780 t
Fairing Diameter: 5.4 m
Booster Propellant: Solid Propellant
Stage-1 Propellant: LOX/LH_2
Stage-2 Propellant: LOX/LH_2
GTO Capability: 10,500 kg
Launch Site(s): Kourou

Ariane 5ES

Height: 59.0 m
Lift-off Mass: 780 t
Fairing Diameter: 5.4 m
Booster Propellant: Solid Propellant

Stage-1 Propellant: LOX/LH_2
Stage-2 Propellant: LOX/LH_2
LEO Capability: 21,500 kg
Launch Site(s): Kourou

Ariane 6 (Under Development)

Launch Site(s): Kourou

Vega

700 km Polar Orbit Capability: 1,500 kg
Launch Site(s): Kourou
First Flight: January 2012
Note: This launcher can be configured for multiple simulatious launch of several small satellites of the 200 kg class.

India

Polar Satellite Launch Vehicle (PSLV)

Height: 44 m
Lift-off Mass: 294 t
SSO Capability: 620 kg
GTO Capability: 1,050 kg
Launch Site(s): Satish Dhawan Space Center (SHAR)

Geosynchronous Satellite Launch Vehicle (GSLV)

Height: 49.0 m
Lift-off Mass: 414 t
GTO Capability: 2,500 kg
Launch Site(s): Satish Dhawan Space Center (SHAR)

Geosynchronous Satellite Launch Vehicle (GSLV): Mark III (Proposed)

Height: 42.4 m
Lift-off Mass: 630 t
GTO Capability: 4,500 kg

Launch Site(s): Satish Dhawan Space Center (SHAR)
First flight: 2012 (proposed)

Israel

Shavit 2

Height: 18.0 m
Lift-off Mass: 30 t
Polar LEO Capability: 300–800 kg
Launch Site(s): Palmachim Air Force Base

Japan

H-IIA 202

Height: 53.0 m
Lift-off Mass: 289 t
Fairing Diameter: 4 m
Booster Propellant: Solid Propellant
Stage-1 Propellant: LOX/LH_2
Stage-2 Propellant: LOX/LH_2
LEO Capability: 10,000 kg
GTO Capability: 3,700 kg
Launch Site(s): Tanegashima Space Center

H-IIA 204

Height: 53.0 m
Lift-off Mass: 289 t
Fairing Diameter: 4 m
Booster Propellant: Solid Propellant
Stage-1 Propellant: LOX/LH_2
Stage-2 Propellant: LOX/LH_2
GTO Capability: 6,000 kg
Launch Site(s): Tanegashima Space Center

H-IIB 304

Height: 57.0 m
Lift-off Mass: 530 t

Fairing Diameter: 5.2 m
Booster Propellant: Solid Propellant
Stage-1 Propellant: LOX/LH_2
Stage-2 Propellant: LOX/LH_2
GTO Capability: 9,000 kg
LEO Capability: 16,500 kg (for HTV orbit)
Launch Site(s): Tanegashima Space Center

South Korea

Naro-1 (KSLV-I)

Height: 33.0 m
Lift-off Mass: 140 t
Fairing Diameter: 3 m
Stage-1 Propellant: LOX/Kerosene
Stage-2 Propellant: Solid
LEO Capability: 300 kg
Launch Site(s): Naro Space Center
Note: First two flights unsuccessful. No successful flights as of October 2011. As of Jan. 10, 2013, the Naro rocket was able for the first time to place 300 kg into low earth orbit.

Russia

Angara 1.1 (Proposed)

Height: 34.9 m
Lift-off Mass: 149 t
LEO Capability: 2,000 kg
Launch Site(s): Plesetsk Cosmodrome (proposed)
First flight: 2013 (proposed)

Angara 1.2 (Proposed)

Height: 41.5 m
Lift-off Mass: 171.5 t
LEO Capability: 3,700 kg
Launch Site(s): Plesetsk Cosmodrome (proposed)
First flight: Unknown. Post-2013 (proposed)
Note: There are larger versions of the Angara launcher proposed as well.

Proton M/Breeze M

Height: 53 m
Lift-off Mass: 690 t
LEO Capability: 22,000 kg
GTO Capability: 6,920 kg
GSO Capability: 3,250 kg
Launch Site(s): Baikonur Cosmodrome

Rockot

Height: 29.2 m
Lift-off Mass: 107 t
Fairing Diameter: 2.5 m
LEO Capability: 2,140 kg
Launch Site(s): Plesetsk Cosmodrome

Shtil

LEO Capability: 160 kg
Launch Site(s): Launched from a submarine (typically from Barents Sea)

Volna

LEO Capability: 100 kg
Launch Site(s): Launched from a submarine (typically from Barents Sea)

Start-1

Height: 22.7 m
Lift-off Mass: 47 t
Fairing Diameter: 1.8 m
SSO Capability: 350 kg
Launch Site(s): Plesetsk Cosmodrome

Soyuz FG/Fregat

Height: 49.5 m
Lift-off Mass: 305 t

LEO Capability: 7,800 kg
SSO Capability: 4,500 kg
Launch Site(s): Baikonur Cosmodrome

Soyuz-2.1a/Soyuz-ST

Height: 46.1 m
Lift-off Mass: 305 t
Fairing Diameter: 2.95 m
GTO Capability: 2,760 kg
SSO Capability: 4,500 kg
Launch Site(s): Baikonur Cosmodrome, Centre Spatial Guyanais

Vega

Height: 30 m
Diameter: 3 m
Liftoff mass: 137 tonnes
Payload mass: 1500 kg to LEO
Launch Site(s): Baikonur Cosmodrome, Centre Spatial Guyanais

Ukraine

Zenit2

Height: 57 m
Lift-off Mass: 445 t
LEO Capability: 13,740 kg
Launch Site(s): Baikonur Cosmodrome

Zenit 3SL

Height: 59.6 m
Lift-off Mass: 462.2 t
GTO Capability: 6,100 kg
Launch Site(s): Sea Launch Odyssey Platform

United States

Athena I

Height: 19.8 m
Lift-off Mass: 66.3 t
LEO Capability: 820 kg
SSO Capability: 360 kg
GTO Capability: 13,740 kg
Launch Site(s): Cape Canaveral, Vandenberg, Kodiak
Note: Retired 2001, proposed to return to service 2012.

Athena II

Height: 30.5 m
Lift-off Mass: 120.2 t
LEO Capability: 2,065 kg
SSO Capability: 1,165 kg
GTO Capability: 590 kg
Launch Site(s): Cape Canaveral, Vandenberg, Kodiak
Note: Retired 2001, proposed to return to service 2012.

Atlas V

Note: There are 10 variations of the Atlas V. The naming convention for the Atlas V is Atlas V XYZ, where X is the diameter of the payload fairing (4 m or 5 m), Y is the number of SRB boosters (0,1,2,3,4,5), and Z is the number of engines on the Centaury core stage.

Atlas V Variation	401	411	421	431	501	511	521	531	541	551
GTO Capability (kg)	3,460	4,450	5,210	5,860	2,690	3,900	4,880	5,690	6,280	6,860

Launch Site(s): Cape Canaveral, Vandenberg

Delta II

Note: There are numerous variations of the Delta II. The naming convention for the Delta II is Delta II VWXY-Z, where V is the type of first-stage engine (7 = RS-27A engine), W is the number of first-stage solid rocket motors (9,4, or 3), X is the type of second stage (2 = Aerojet AJ10–118 K engine), Y is the type of third stage

(0 = none, 0H = None, heavy configuration, 5 = Star-48B solid motor, 5H = Star-48B solid motor, heavy configuration, 6 = Star-37FM solid motor), and Z = the type of fairing (9.5 ft diameter, 10 ft diameter, or 10 ft diameter long).

Two-stage		Three-stage	
LEO capability (kg)		GTO capability (kg)	
Delta II 7320–9.5	2,809	Delta II 7326–9.5	934
Delta II 7320-10	2,703	Delta II 7326-10	898
Delta II 7420–9.5	3,185	Delta II 7425–9.5	1,110
Delta II 7420-10	3,099	Delta II 7425-10	1,073
Delta II 7920–9.5	5,030	Delta II 7426–9.5	1,058
Delta II 7920-10	4,844	Delta II 7426-10	1,029
Delta II 7920-10 L	4,805	Delta II 7925–9.5	1,819
Delta II 7920H-9.5	6,097	Delta II 7925-10	1,747
Delta II 7920H-10	5,959	Delta II 7925-10 L	1,739
Delta II 7920H-10 L	5,899	Delta II 7926–9.5	1,660
		Delta II 7926-10	1,581
		Delta II 7926-10 L	1,578
		Delta II 7925H-9.5	2,171
		Delta II 7925H-10	2,123
		Delta II 7925H-10L	2,102
		Delta II 7926H-9.5	1,981
		Delta II 7926H-10	1,934
		Delta II 7926H-10L	1,916

Launch Site(s): Cape Canaveral, Vandenberg

Delta IV

Note: There are five variations of the Delta IV.

	Number of common core boosters (first stage)	Number of solid boosters	Payload fairing diameter (m)
Delta IV medium	1	0	4
Delta IV medium + (4,2)	1	2	4
Delta IV medium + (5,2)	1	2	5
Delta IV medium + (5,4)	1	4	5
Delta IV heavy	3	0	5
	Payload capabilities (kg)		
	GEO	GTO	LEO

(*continued*)

	Number of common core boosters (first stage)	Number of solid boosters	Payload fairing diameter (m)
Delta IV medium	1,348	4,508	9,390
Delta IV medium + (4,2)	2,208	6,200	12,477
Delta IV medium + (5,2)	2,105	5,124	11,062
Delta IV medium + (5,4)	3,116	6,905	13,774
Delta IV heavy	6,573	13,248	22,977

Launch Site(s): Cape Canaveral, Vandenberg

SpaceX Falcon 1e

Height: 24.7 m
Lift-off Mass: 35.2 t
Fairing Diameter: 1.7 m
LEO Capability: 1,010 kg
Launch Site(s): Omelek Island

SpaceX Falcon 9

Height: 54.9 m
Lift-off Mass: 333 t
Fairing Diameter: 3.6 m
LEO Capability: 10,450 kg
SSO Capability: 8,560 kg
GTO Capability: 4,600 kg
Launch Site(s): Omelek Island, Cape Canaveral

Orbital ATK Pegasus

LEO Capability: 450 kg
SSO Capability: 300 kg
Launch Site(s): Launched from an aircraft. Sites include Wallops, Canary Island, Omelek Island, Cape Canaveral, Vandenberg

Orbital ATK Minotaur I

Height: 19.2 m
Fairing Diameter: 1.27 m

LEO Capability: 580 kg
Launch Site(s): Vandenberg, Kodiak, Cape Canaveral, WFF

Orbital ATK Minotaur IV

Fairing Diameter: 2.05 m
SSO Capability: 900 kg
LEO Capability: 1,900 kg
Launch Site(s): Vandenberg, Kodiak, Cape Canaveral, WFF

Orbital ATK Taurus

Height: 32 m
Lift-Off Mass: 77 t
LEO Capability: 700–1,050 kg (depending on configuration)
Launch Site(s): Vandenberg, Kodiak, Cape Canaveral, WFF

Orbital ATK Antares

Height
110/120: 2 stage 40.5 m (133 ft)
130: 3 stage 41.9 m (137 ft)
Diameter: 3.9 m (13 ft)
Mass: ~240,000 kg (530,000 lb)
Capacity: Payload to LEO: 6,120 kg (13,490 lb)
Launch Site: Mid-Atlantic Spaceport

Launcher One

This is a new small launcher under development for the 2018 time frame that is being developed by Virgin Galactic. Its specifications are not final. It is scheduled to be part of the launch operations for the One Web Constellation.

Stratolauncher

This is currently under development. It a new large-scale carrier vehicle that is intended to assist with cost-efficient launches in the future.

Vulcan Launch Vehicle

This is the new core launch vehicle being developed by the United Launch Alliance (ULA). This is a new core launcher that will replace the Atlas with greater lift capability. It will work with the Centaur vehicle as the second stage from 2019 to 2023. This lift capability can be supplemented with solid rocket boosters and can work with 4 ft. (1.3 m) or 5 ft. (1.6 m) fairings.

Global Communications Satellite Systems

Joseph N. Pelton

International Communications Satellite Systems[a]			
Name of satellite system	**Location of headquarters**	**Service provided**	**Web site**
COMMStellation (planned for possible launch in 2018)[b]	Microsat Systems Canada, Inc. Mississauga, Ontario	Broadband Internet	www.commstellation.com/constellation/
Eutelsat	Paris, France	FSS and BSS	www.eutelsat.com
Globalstar	Milpitas, California, USA	MSS	www.globalstar.com
Inmarsat	London, UK	MSS	www.inmarsat.com
Intelsat	Luxembourg	FSS	www.intelsat.com www.intelsatgeneral.com
Intersputnik	Moscow, Russia	FSS and BSS	http://www.intersputnik.com/satellites/00023/
Iridium and Iridium Next	Chandler, Arizona, USA	MSS	www.leosat.com
Leosat	Arlington, Va	Broadband Internet	www.leosat.com/
O3b	Paris, France	FSS – broadband internet	www.O3b.com
One Web (large scale LEO system to be deployed in 2018)[b]	Jacksonville, Florida	Broadband Internet	http://oneweb.world/

(*continued*)

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_90

Orbcomm	Fort Lee, New Jersey, USA	Store and forward data	www.orbcomm.com
SES Global – SES World Skies	Luxembourg	FSS, BSS, and DTH TV	http://www.ses-worldskies.com
SPACEX Large Scale LEO Constellation[b]	Plans still pending	Broadband Internet	

[b] Note: These three systems are not yet deployed

Regional Communications Satellite System[a]

Name of satellite system	**Location of system headquarters**	**Service provided**	**Web site**
Amos	Tel Aviv, Israel (Operated by Spacecom)	FSS and DTH TV	www.amos-spacecom.com/content.cfm
AP Star	Hong Kong, China (Operated by APT Satellite)	FSS and DTH TV	http://www.apstar.com/apt_apstar
Arabsat	Riyadh, Saudi Arabia	FSS and DTH TV	http://www.arabsat.com
Asia Broadcasting Satellite System (formerly Agila)	Bermuda, UK	BSS/DTH TV	http://www.absatellite.net/about/index.html
Asia Cellular Satellite (formerly Garuda Satellite)	London, UK (Managed by Inmarsat)	MSS plus terrestrial cellular	www.inmarsat.com www.acesinternational.com/corporate
Asia Sat	Causeway Bay, Hong Kong	FSS and DTH TV	http://www.asiasat.com/
GE Satellite	Bethesda, Maryland, USA	FSS and DTH TV	www.gesatellite.com
Hispasat	Madrid, Spain	FSS and DTH TV	www.hispasat.com
Hylas	London, UK (Operated by Avanti Corporation)	FSS/Broadband Data	www.avantiplc.com
ICO Global Communications	Bellevue, Washington, USA	MSS	www.ico.com
IP Star	Bangkok, Thailand	FSS	www.ipstar.com
JSAT and Horizon	Tokyo, Japan	FSS	http://www.jsati.com/satellite-services.asp
Optus	Sydney, Australia	FSS and DTH TV	http://www.optus.com
Telenor	Oslo, Norway	FSS and BSS	http://www.telenor.com/
Telstar 11, 12, 14, & 18	Palo Alto, California USA (Operated by Space Systems/Loral)	FSS and Direct to the Home Broadcast	http://www.telesat.ca/en/Telstar_Fleet

(*continued*)

Thuraya	Abu Dhabi, United Arab Emirates	MSS	http://www.thuraya.com
Worldspace	Yazmi Holdings, Maryland, USA	DABS	biz.yahoo.com/e/100629/wrspq.pk8-k.html
XTAR	Palo Alto, California, USA and Madrid, Spain	FSS in X-band to support military services	www.xtar.com
Yahsat	Abu Dhabi, United Arab Emirates	FSS Services	www.yahsat.com
Domestic Communications Satellite Systems[a]			
Country	**Name of satellite (s)**	**Type of service**	**Web site for latest system information**
Argentina	*ARSAT-1*	FSS and DTH TV (Pending service shortly)	http://www.arsat.com.ar/ingles/home_ing.html
Australia	*Optus*	FSS and DTH TV	http://www.optus.com
Brazil	*Brazilsat*	FSS and DTH TV	http://www.embratel.com.br/
Canada	*Anik, Nimiq DBS, Telstar*	*Anik* – FSS to Canada and North America *Nimiq* – BSS service to Canada *Telstar* – FSS to Canada and Atlantic region	http://www.telesat.ca
China	*ChinaSat-6*	Sinosat: FSS	http://www.fas.org/spp/guide/china/comm/chinasat.htm
	Sinosat	Chinasat: BSS	info@sinosat.com.cn
Egypt	*Nilesat 1 & 2*	FSS	http://www.nilesat.com.eg/services.htm
Greece	*Nova Greece*	BSS	www.satsig.net/ivsat-europe.htm
India	*Insat*	FSS and BSS	
Iran	*Mesbah*	Store and Forward data	www.universetoday.com/24552/iran-launches-satellite-into-orbit
Israel	*Amos*	FSS and DTH TV	http://www.amos-spacecom.com/
Japan	*B-Sat*	BSS	www.b-sat.co.jp/
	JSat	FSS and BSS	http://www.jsati.com/satellite-services.asp
	N-star	MSS	http://www.nttdocomo.com/services/miscellaneous/index.html
Korea	Koreasat	FSS and BSS	www.lyngsat.com/South-Korea.html

(*continued*)

Luxembourg	*ASTRA*	FSS and DTH TV	www.ses-astra.com
Malaysia	*Measat*	FSS and DTH TV	http://www.measat.com/satellite.html
Mexico	*SatMex*	FSS	http://www.satmex.com.mx/index1.php
	Solidaridad Satellites	BSS	http://space.skyrocket.de/doc_sdat/solidaridad-1.htm
Nigeria	*Nigcomsat*	BSS	http://www.nigerianbestforum.com/index.php
Norway	*Telenor*	FSS and BSS	www.highbeam.com/doc/1G1-110308909.html
Pakistan	*Paksat-1*	FSS and DTH TV	www.suparco.gov.pk/pages/paksat1.asp
Russia	*Express*	FSS and DTH TV	www.intersputnik.com
	Intersputnik		
	Stationar		
	Yamal		
Spain	*Hispasat*	FSS and DTH TV	www.hispasat.com/
Thailand	*Thaicom*	FSS and DTH TV	www.thaicom.net
Turkey	*Turksat*	FSS and DTH TV	www.ts2.pl/en/Satellite-Internet/Turkey
United Arab Emirates	*Yahsat*	FSS and DTH TV	www.yahsat.ae
United Kingdom	*Sky Television*	BSS	www.sky.com
United States	DBSD	Hybrid MSS	http://www.ico.com/_about/
United States	Direct TV	BSS	www.directv.com
United States	Dish (Echostar)	BSS	www.dish.com www.echostar.com
United States	Galaxy (Intelsat)	FSS and DTH TV	www.intelsat.com
United States	GE SAT	FSS and DTH TV	www.gesat.com
United States	G Star	FSS and DTH TV	www.geo-orbit.org/westhemipgs/fgstar4p.html
United States	Hughes Net	FSS and broadband data	www.hns.com
United States	JSAT Horizon	FSS and DTH TV	www.jsati.com
United States	Light Squared Network	Hybrid MSS	www.lightsquared.com

(*continued*)

United States	SES Americom	FSS and DTH TV	www.ses-americom.com
United States	Terrestar	Hybrid MSS	www.terrestar.com
United States	Wild Blue	FSS and DTH TV	www.wildblue.com
United States	XM-Sirius	DABS	www.xmradio.com

[a] This chart was prepared and copyrighted by Joseph N. Pelton. Copyright permission is granted to Springer Press for publication in the *Handbook for Satellite Applications*. Please check with the relevant web site to obtain the latest information

US Domestic Communications Satellite Systems

Joseph N. Pelton

Name of system (and HQ location)	Operator	Type of service	Web site for latest information
DBSD (Bellevue, Washington and Reston, VA, USA)	Former ICO Global Communications Subsidiary (one satellite plus terrestrial mobile) (now owned by DirecTV)	Hybrid terrestrial-satellite land mobile service	www.DirecTV.com/
DirecTV	DirecTV	Broadcast satellite service-DBS	www.directv.com
Dish	Dish (Echostar)	Broadcast satellite service-DBS	www.echostar.com www.dish.com
Galaxy and Intelsat Americas Sats	Intelsat	Fixed and DTH service	www.Intelsat.com
GE Sat	General Electric	Fixed satellite and DTH service	www.gesat.com
G Star	GTE & SES Americom	Fixed satellite and DTH service	www.geo-orbit.org/westhemipgs/fgstar4p.html
Hughes Net	Hughes Network Systems	IP-based broadband data services. Now provided by Jupiter Satellites	www.hns.com

(*continued*)

J.N. Pelton (✉)
International Space University, Arlington, VA, USA
e-mail: joepelton@verizon.net

J.N. Pelton et al. (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-3-319-23386-4_1

Name of system (and HQ location)	Operator	Type of service	Web site for latest information
JSAT (Horizon)	Japan Satellite	DTH and FSS services	www.jsati.com/news.asp?DocID=180
LightSquared (formerly Mobile Satellite Ventures)	LightSquared Network	Hybrid mobile satellite and terrestrial cellular service	www.lightsquared.com/
SES Americom	SES Global	DTH satellite and fixed satellite service	www.ses-americom.com
Sirius (Now part of XM Radio)	XM Radio Inc.	Direct broadcast radio	www.xmradio.com/bestofsirius/index.xmc
Terrestar	TerreStar Corporation (now owned by DirecTV)	Hybrid mobile satellite and terrestrial cellular service	www.DirecTV.com/
Viasat	Viasat-1 and Viasat-2 to be deployed shortly	Broadband service	
Wild Blue (Formerly Ka star)	Wild Blue Communications Inc.	Fixed satellite and DTH service	www.wildblue.com
XM Radio	XM Radio Inc.	Direct broadcast radio	www.xmradio.com

Table prepared and copyrighted by Joseph N. Pelton, printed by his permission