ASTROPHYSICS AND SPACE SCIENCE PROCEEDINGS

Natchimuthukonar Gopalswamy
S. Sirajul Hasan
Ashok Ambastha
Editors

Heliophysical Processes

Springer

Astrophysics and Space Science Proceedings

For further volumes:
http://www.springer.com/series/7395

Heliophysical Processes

Natchimuthukonar Gopalswamy
Editor

NASA Goddard Space Flight Center

S. Sirajul Hasan
Editor

Indian Institute of Astrophysics

Ashok Ambastha
Editor

Udaipur Solar Observatory
Physical Research Laboratory

Springer

Editors
Dr. Natchimuthukonar Gopalswamy
Heliophysics
NASA Goddard Space Flight Center
8800 Greenbelt Road
Greenbelt, MD 20771
USA
nat.gopalswamy@nasa.gov

Prof. S. Sirajul Hasan
Indian Institute of Astrophysics
IInd Block, Koramangala
Sarjapur Road
Bangalore-560034
India
hasan@iiap.res.in

Prof. Ashok Ambastha
Udaipur Solar Observatory
Physical Research Laboratory
Udaipur-313001
India
ambastha@prl.res.in

ISSN 1570-6591 e-ISSN 1570-6605
ISBN 978-3-642-11340-6 e-ISBN 978-3-642-11341-3
DOI 10.1007/978-3-642-11341-3
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010921920

Cover design: SPi Publisher Services

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The first Asia-Pacific Regional School of the International Heliophysical Year (IHY) 2007 program was held at the historic Kodaikanal Observatory of the Indian Institute of Astrophysics during December 10–22, 2007. The School was cosponsored by the IHY program, the Indian Institute of Astrophysics (Bangalore, India), and the Asian Office of Aerospace Research (Air Force Office of Scientific Research, United States). Selected lectures, presented at the IHY School, form the basis of the chapters included in this volume.

The IHY 2007 program marked the fiftieth anniversary of the International Geophysical Year (IGY, 1957–1958). The IHY activities specifically recognized the vast extent of space explored by humankind today, compared to only the geospace probed during the IGY. The IHY program ran from 2007 to 2009, and engaged in scientific investigations involving global cooperation, deployment of small instruments in developing countries in collaboration with the United Nations, public outreach to communicate the beauty, relevance and significance of space science to the general public and students, and the preservation of history related to the IGY program. This IHY School held at Kodaikanal is one of the activities planned as part of public outreach.

The School offered an intensive two week course in topics related to heliophysics, aimed primarily at graduate and post-doctoral research students, and covered a broad range of physical processes in heliospace, extending from the center of the Sun to the edge of the solar system. Students learned how the Sun influences the heliosphere through its electromagnetic and mass emissions. These Schools provide an introduction to heliophysics for students who normally do not have an opportunity to take such a course at their home institution. The IHY School in Kodaikanal was attended by about 40 students, mostly from India but there were also a few students from the Republic of Korea, Sri Lanka, and Nigeria. The School organized 40 one-h lectures by 26 lecturers. There were also laboratory sessions in using data bases and observing techniques at different wavelengths. The School was directed by N. Gopalswamy, A. Ambastha, and R. Ramesh. The editors thank R. Ramesh (the coordinator of the IHY School), K. E. Rangarajan (the convener of Kodaikanal schools) and K. Sundararaman (Scientist-in-charge of the Kodaikanal Observatory) for their tireless efforts in running the school smoothly.

The editors take this opportunity to thank Dr. Pertti Mäkelä for his valuable assistance in preparing this volume. We also thank the following reviewers who assisted in improving the chapters: T. Zurbuchen, A. Vinas, B. Vrsnak, V. Jordanova, M. Dikpati, P. Mäkelä, K. R. Sivaraman, G. S. Lakhina, Shyam Lal, P. K. Manoharan, Nandita Srivastava and Harish Chandra. Finally, we appreciate the support provided by R. Ponnappan and J. Moses (AFOSR) and M. Guhathakurta (NASA) in running the IHY School and bringing out this volume.

Greenbelt, Maryland, USA, *Natchimuthukonar Gopalswamy*
Bangalore, India, *Siraj Hasan*
Udaipur, India, *Ashok Ambastha*
October 2009

Contents

List of Contributors

Ashok Ambastha Udaipur Solar Observatory, Physical Research Laboratory, Udaipur 313001, India, ambastha@prl.res.in

Nanan Balan Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, UK, B.NANAN@sheffield.ac.uk

B. Bavassano Istituto di Fisica dello Spazio Interplanetario (Istituto Nazionale di Astrofisica), Via del Fosso del Cavaliere 100, 00133 Roma, Italy, Bruno.Bavassano@ifsi-roma.inaf.it

A. Bhardwaj Space Physics Laboratory, Vikram Sarabhai Space Centre, Trivandrum 695022, India, Anil_Bhardwaj@vssc.gov.in

R. Bruno Istituto di Fisica dello Spazio Interplanetario (Istituto Nazionale di Astrofisica), Via del Fosso del Cavaliere 100, 00133 Roma, Italy, Roberto.Bruno@ifsi-roma.inaf.it

H. Chandra Physical Research Laboratory, Ahmedabad 380 009, India, hchandra@prl.res.in

A.C. Das Physical Research Laboratory, Ahmedabad 380 009, India, anathcdas@gmail.com, acd@prl.res.in

Natchimuthukonar Gopalswamy NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA, nat.gopalswamy@nasa.gov

Arnold Hanslmeier Department of Geophysics, Astrophysics and Meteorology, Institute of Physics Univ.-Platz 5, 8010 Graz, Austria, arnold.hanslmeier@uni-graz.at

Eberhard Möbius Space Science Center and Department of Physics, University of New Hampshire, Durham, NH, USA, eberhard.moebius@unh.edu

Thomas Earle Moore NASA's Goddard Space Flight Center, Greenbelt, MD, USA, thomas.e.moore@nasa.gov

Dibyendu Nandy Indian Institute of Science Education and Research, Kolkata, India, dnandi@iiserkol.ac.in

R. Ramesh Physical Research Laboratory, Navrangpura, Ahmedabad 380 009, India, rramesh@prl.res.in

John D. Richardson Massachusetts Institute of Technology, Cambridge, MA, USA, jdr@space.mit.edu

Takashi Sako Solar-Terrestrial Environment Laboratory, Nagoya University, Nagoya 464-8601, Japan, sako@stelab.nagoya-u.ac.jp

C. Sivaram Indian Institute of Astrophysics, Bangalore, India, arun_kenath@yahoo.com

K.R. Subramanian Indian Institute of Astrophysics, Koramangala, Bangalore 34, India, subra@iiap.res.in

Manish Tiwari National Centre for Antarctic and Ocean Research, Vasco da Gama, Goa 403 804, India, manish@ncaor.org

The Sun in the Universe

C. Sivaram

Abstract This chapter provides an over view of: (1) Processes in the solar core and corona; (2) The Sun's place in the galaxy and universe and uniqueness of the Sun for life on earth; (3) The Sun's role in Olber's paradox (darkness of night sky) which gave the first clue about the finiteness of the universe, crucial for cosmology; (4) Discovery of Helium in the Sun and its importance for cosmology; (5) Heavy-element abundances in the Sun compared with stellar abundances; (6) Importance of the solar neutrino problem and its final resolution in the Sudbury detector; (7) The Sun's role in the detection of Dark matter and the current status of axions, wimps etc.; (8) The future evolution of the Sun and its effect on Earth.

1 Introduction

From an astronomer's point of view, the Sun is just a fairly ordinary main-sequence middle-aged low mass star with a G2V spectral type and an X-ray emitting corona. But in several respects, Sun is not typical compared to nearby Sun-like stars. Typical Sun-like star does not have constant luminosity. High-precision photometric observations reveal most F and K dwarfs vary at millimag level or greater. Both long term and short term (day to day) variability correlate with Ca II H and K emission index (Radick 2001). In both cases, Sun's variability appears much smaller than stars with similar level of chromospheric activity. Gray: STARS and SUN ... ed., Donahue and Bookbinder gives brief review of magnetic activity cycles in stars of solar type and compares with the Sun, and determines low photometric variability of the Sun.

The Sun's location in the Milky Way is also notable in two ways. At present, it is only at a distance of 10–12 pc from the mid-plane. It is very near the co-rotation circle (Mishurov 2000). Encounter with spiral arms is infrequent and there is less threat from supernova explosions (Fig. 1).

C. Sivaram
Indian Institute of Astrophysics, Bangalore, India
e-mail: sivaram@iiap.res.in

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_1,

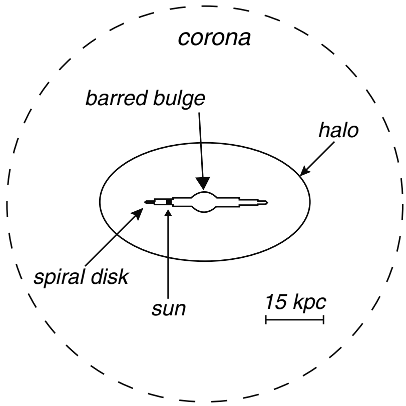


Fig. 1 Milky Way as seen edge on

Table 1 Average coronal energy losses (in 10^4 erg cm^{-2} s^{-1})

Loss by	Quiet Sun	Active region	Coronal hole
Conduction	30	30–103	6
Radiation	10	1,000	2
Solar wind	5	×10	90
Total flux	30	103	100

A HIPPARCOS – based study indicates that out of 37 stars with age estimates of 3–6 Gyr, Sun has v_{ISR} (space velocity relative to LSR) of 13.4 km s^{-1}, whereas for other stars, $v_{LSR} \sim 42 \pm 17$ km s^{-1}. The Sun has the lowest v_{ISR}. There are several such anomalous properties of the Sun.

MS and giant stars of all spectral types (i.e., with exception of degenerate stars) show UV and X-ray emission and show evidence of chromospheric and coronal activities as measured by the OSO series, the IEU and Einstein satellite. F, G, K and M-stars have coronae similar to the Sun where radiation is generally attributed to surface convection of these stars (Table 1).

Late supergiants and giants do not seem to have coronae while A-type stars have neither coronae nor chromosphere. Chromospheric emission from T Tau stars originates from mass-infall from accretion discs while for other stars this energy is not received beyond the stellar atmosphere.

2 Solar Flares

Flares can release energy equivalent of billions of atomic bombs in a few minutes. Explosions release bursts of X-rays, charged particles, etc. which may later hit earth, endangering satellites and causing power blackouts. Sudden release of energy in flare results from magnetic reconnection, wherein oppositely directed magnetic field lines come together and partially annihilate.

Only recently have space probes provided observational evidence for these phenomena. For example the pointed magnetic loops below spots where magnetic reconnection takes place.[1]

Non thermal energy excess to sustain solar corona is only 10^{-4} of Sun's total energy output. It is easy to propose several mechanisms to divert 0.01% of total solar output to heat corona. Improving spatial and temporal resolution reveals more and more fine structure like magnetic pores, dark mottles, spicules, supergranular cells, filaments, EUV bright points, etc.

Summary of popular heating mechanism (Erdelyi et al. 2007)

Hydrodynamic heating	
Carrier	Mechanism
Acoustic waves	Shock dissipation
Pulsation waves	
Magnetic AC waves mechanism	
Slow waves	Shock damping
	Resonant absorption
Fast MHD waves	Landau damping
Alfven waves	Mode coupling, resonant
	Heating, phase mixing
	Viscous heating, turbulent
	Heating, Landau damping
Direct current	
Current sheets	Reconnection

3 Effects of Solar Radiation

1. *Drell drag*: Echo satellite brought down too soon. SKYLAB came down (1979) after 34,981 orbits. Increased solar activity not considered.
2. *Poynting-Robertson effect*: small particles spiral into Sun.
3. *Yarkovsky effect*: important for asteroids.
4. *YORP effect*: asteroid spin rotation affected.

3.1 Solar Radiation Propulsion: Solar Sail

The solar radiation pressure is of the order of 1 dyne m^{-2}. A large sails ($\sim$1 km^2), using this solar radiation pressure, can reach Mars in weeks!

[1] Referred http://hesperia.gsfc.nasa.gov/sftheory.

3.2 Solar Twins

A new solar twin, HD 98618, was found among a sample of 16 Sun-like stars whose spectra was observed in fine detail with 10-meter Keck telescope. This is the second close solar analogue after 18 Scorpii. Both orbit about 8 kpc from galaxy centre. And both have roughly the same heavy element concentration as the Sun.

The stars are less than 1% hotter, spin a little faster, making them less than half a billion years (less than 8%) younger. They do not have hot Jupiters. Astronomers urge that top priority be given to these two stars when searching for extra-terrestrial intelligence or future searches for life. The time scale for the development of these star systems is the same as that for the Sun.[2]

4 Distance to Sun

The Earth–Sun distance is a key scale in fixing distances in the universe. For instance, stellar parallax using the Earth–Sun distance as the baseline, gives distances to several nearby stars. Thus determination of astronomical unit (AU) accurately taxed several astronomers who used the transit of Venus, passage of asteroids like Eros, etc. to determine this basic astronomy ruler. Nowadays, bouncing radar off the atmospheres of planets has enabled this to be determined to the nearest kilometer.

5 Newton and the Density of the Sun

Although Newton could not accurately estimate the AU, his ingenuity could nevertheless estimate the average solar density as just above that of water. He did this just by knowing that the Sun subtends an angle of half a degree in the sky.

Applying Kepler's III law: $\mathrm{GMT}^2 = 4\pi^2\mathrm{R}^3$, where the mass is given by: $\mathrm{M} = {}^4\!/_3\left(\mathrm{r}^3\pi\rho\right)$, where, ρ and r are the Sun's average density and radius. $(r/R)$is about half a degree $\sim$0.01 rad. And the result follows.

6 The Sun and Olber's Paradox

On a moonless night, the total intensity of light falling from the night sky off all the other stars put together is less than a hundred millionth of what we get from the Sun during daytime. The utter 'darkness' of the night sky gave the first clue as

[2] Ref: www.arxiv.org/astro-ph/0603219.

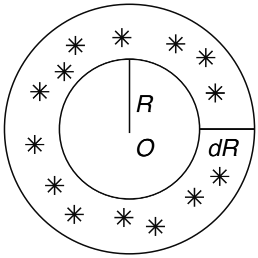


Fig. 2 Universe divided into shells of thickness *dR*

to what type of universe we may be in. This is the famous Olber's paradox. Olbers (and others before him, especially Kepler) had pondered as to what effect an infinite static universe (favoured by many philosophers through the ages!) would have on the appearance of the sky. A simple calculation shows that the night sky in this case would be at least as bright as the daytime sky.

Consider a spherical shell of stars around any observer O as in Fig. 2.

For a shell of thickness dR and radius R, the volume of the shell would be $4\pi R^2 dR$. If L be the average luminosity of a star in the shell and n be the number density in the shell, then for an infinitely large universe, the total light intensity at any arbitrary O would diverge as $I \sim \int_0^\infty n dR$

$$\text{Number of stars in the shell} = 4pR^2 dR.n$$

So the darkness of the night sky shows that the concept of an infinite static universe is untenable. Indeed Hubble discovered that the universe is expanding and astronomers realized that all stars including the Sun have only a finite lifetime.

With these conditions the brightness of the night sky is much limited. Many treatises have been written on this subject: e.g. E Harrison 'Darkness at night'.

7 Discovery of Helium

Solar eclipses have provided important clues to understand the universe. The famous eclipse of 1919 played an important role in confirming light deflection by the Sun, proving *Einstein's general relativity*. The most famous total solar-eclipse discovery was made in 1868 in India. Spectrographic observations revealed brightest emission lines of the chromosphere for the first time. At first bright yellow line was thought to be D line of Sodium, but was soon realized that it was displaced from D1–D2 doublet and was called D3. The element that caused bright D3 line was called Helium (after Helios – the Sun). This new element was not identified on Earth till 1895, by Ramsey. Helium is the second most abundant element in the universe after hydrogen. It is one-fourth as abundant as hydrogen. Even the oldest stars in the universe (with 10^{-5} of solar iron) show 20% He abundance.

Although stars are powered by conversion of hydrogen to helium, it can be easily calculated that all the stars during a Hubble time can convert a bare 2% of the hydrogen to helium. But even the oldest objects, more than ten billion years old, show much higher He abundance. The hot big bang model, in which the universe began with a very hot dense phase, naturally accounted for the one-fourth abundance of helium. So He played important role in cosmology. Thermonuclear reactions in the *first three minutes* of the hot expanding universe would convert about a fourth of the hydrogen to helium as observed. In addition, the big bang produced minute traces of D, He-3 and lithium, which are not made in stars. Their observed abundances also agree with observation. All the other heavier elements were made in the stars.

8 Gold in the Sun

From observation of Au *3122.798* line, it is estimated that there is 10^{16} tons of gold in the Sun. Gold, uranium and other heaviest elements are produced in supernovae type II, which is formed due to the collapse of massive stars. The nuclear r-process is responsible for the creation of these heavy elements. Each supernova type II explosion perhaps produces about 10^9 tons gold, however blast energy (total power released) in SN II is about 10^{40} units, worth trillion times the cost of gold produced.

9 Cosmic Power Houses

The following is a comparative study that shows the magnitude of power generated by the Sun compared to other sources. Sun's luminosity (power) is of the order of 4×10^{26} W (4×10^{20} MW), whereas that for a Boeing 747 $\sim$ 300 MW. Total power consumed by man (total power installed capacity) $\sim 10^{13}$ Watts (India $\sim$120,000 MW $\sim$1% of world power). In one second, the Sun radiates as much energy as the mankind will consume in several million years!

The mass of Sun is of the order of 2×10^{30} kg. If the Sun were made entirely of coal or petroleum it would burn up completely in less than 6,000 years! A Boeing 747 consumes 160 tons of jet fuel on a 12 h flight! The SATURN V rocket which took the Apollo 11 manned flight to the moon has a jet power of 200×10^6 HP and the fuel consumed $\sim$2,000 tons. Comparatively the Sun is equivalent to 3×10^{23} HP $\sim 10^{18}$ jumbo jets.

The source of Sun's power is nuclear energy: mass is converted to energy according to $E = mc^2$ in thermonuclear reactions that convert 4 hydrogen (H) nuclei into a helium (He) nuclei. In one second the Sun converts, in its core, 564 million tons H to 560 million tons He and 4 million tons converted to energy. That is: 4×10^6 ton $\times\, c^2 = 4 \times 10^{26}$ W

There are stars in the universe emitting several million times the Sun's energy (power). The Sun has been radiating energy at this rate for about 5 billion years now and would be radiating for another at least 5 billion years. Our galaxy has 200 billion stars and the power emitted by galaxy $\sim 10^{37}$ Watts.

10 Most Energetic Objects in the Universe

1. Supernovae: star explosion
2. QUASARS
3. Gamma Ray Bursts

Supernovae: Release a total of $\sim 10^{44}$ Joules of energy in a few weeks, as much as what the Sun radiates in its lifetime of 10 billion years!

QUASARS: they are very far away objects (billions of light years away). They are powered by supermassive black holes (billion solar masses). In one second, a quasar may emit in X-rays and gamma rays what the Sun emits in 10 billion years.

Gamma Ray Bursts: they are caused due to the collapse of very massive stars or merger of dense neutron stars. In a few seconds, they emit in gamma rays what the Sun emits in radiation in its entire life time of 10 billion years.

11 Nuclear Reactions in the Solar Core

The following reactions that occur at a temperature of 10^7 K at the core of the Sun, are responsible for the power generated by the Sun.

$$\mathrm{p} + \mathrm{p} \rightarrow \mathrm{D} + \mathrm{e}^+ + {}^1\nu_e \text{ (electron neutrino escapes right through!)}$$
$$\mathrm{D} + \mathrm{p} \rightarrow \mathrm{He}^3 + \gamma \text{ (gamma photon)}$$
$$\mathrm{He}^3 + \mathrm{He}^3 \rightarrow \mathrm{He}^4 + \mathrm{p} + \mathrm{p}$$

Direct evidence for nuclear reaction in Sun is obtained by the detection of solar neutrinos. Four protons fusing to form He releases 26.7 MeV of energy $= 4.5 \times 10^{-5}$ ergs released by $4 \times 1.66 \times 10^{-24}$ g of hydrogen.

To generate 4×10^{33} ergs sec^{-1}, 564 million tons hydrogen per second becomes 560 million tons helium per second. Mass difference of 4 million tons sec^{-1} converted to energy.

$$\mathrm{E} = \mathrm{mc}^2 = 4 \times 10^6 \times 10^3\,\mathrm{kg} \times (3 \times 10^8)^2 = 3.6 \times 10^{33}\,\mathrm{ergs\ sec}^{-1}$$

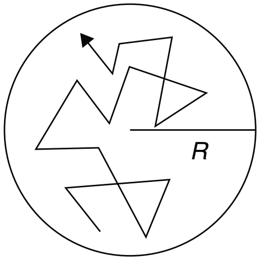


Fig. 3 Diffusion of particles in solar interiors

At this rate the life time of the Sun is given by:

$$\sim \frac{2 \times 10^{32} g}{5.64 \times 10^{14}} = 3 \times 10^{17} \mathrm{s}.$$

This corresponds to about 10 billion years.

Thermonuclear reactions in the Sun produce enormous flux of neutrinos.

$$\mathrm{p} + \mathrm{p} \to \mathrm{D} + \mathrm{e}^{+} + {}^{1}\nu_{\mathrm{e}};\ \mathrm{D} + \mathrm{p} \to \mathrm{He}^{3} + \gamma;\ \text{etc}$$

The flux of neutrinos on Earth from Sun is $\sim 10^{12}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ (10^{14} neutrinos pass through your body per second!)

These neutrinos have now been detected in Gallium and Chlorine experiments. Also the Kamiokande in Japan has detected these solar neutrinos.

12 Supernova Neutrinos

When a massive star explodes as much as 10^{58} neutrinos are released. In February 24 1987, about 20 neutrinos from SN 1987A from Large Magellanic Cloud was detected in water detectors. Water detectors use a few kilotons of heavy water for the neutrino detection.

Within the solar interiors, the diffusion of photon takes nearly a million years because the matter density within the solar interiors is very large (50 times that of steel) (Fig. 3). The mean free path between scattering is of the order of λ. The total number of scatterings is given by: $N = \left(\frac{R}{\lambda}\right)^2$

Time taken to diffuse through a distance R is $t \sim \frac{R^2}{\lambda c}$ (not $\frac{R}{c}$).

For the Sun, $\mathrm{N} \approx 10^{24}$; $\mathrm{R} = 7 \times 10^{5}\,\mathrm{km} \Rightarrow \mathrm{t} \sim 10^{6}$ years.

That is the Sun light takes a million years to diffuse out from the core to the surface. Whereas the scattering cross section for the neutrinos is very small. Hence they pass through the Sun (and Earth) without much resistance.

13 Sudbury Heavy Water Neutrino Detector (SNO)

Earlier neutrino detectors had detected only one-third of the neutrino flux, from the Sun, of what was expected from the solar models. This inconsistency was sorted out by the Sudbury Heavy Water Neutrino Detector in Canada. Heavy water enables detection of neutral current weak interaction. All 3 neutrino flavors (electron, muon and tau) can participate in neutral currents (NC).

$$\nu_x + d \rightarrow \nu_x + p + n, \text{where } x \rightarrow (e, \mu, \tau).$$

Also elastic scattering (ES) can be detected by these detectors.

$$\nu_x + e \rightarrow \nu_x + e$$

However earlier detectors were sensitive only to charge current (CC) reactions, that is

$$\nu + d \rightarrow e^+ + n + n, \text{etc.}$$

Mu and tau neutrinos cannot have CC interaction as threshold energy to produce mu and tau leptons is several hundred MeV (GeV for tau). So if in Sudbury all 3 reactions, i.e., CC + NC + ES are added up, no neutrino deficit is seen. We get full flux predicted by the solar model. The electron neutrinos are altered to mu and tau neutrinos by oscillations as they traverse from Sun to earth. Solar model is correct!

14 Some Subtleties of Solar Nuclear Reactions

$$p + p \rightarrow D + e^+ + \nu_e;\ (0.42\,\text{MeV})$$
$$p + e^- + p \rightarrow D + \nu_e; (1.4\,\text{MeV}, 1 \text{ in } 40)$$
$$e^- + Be^7 \rightarrow Li^7 + \nu_e;\ (0.8\,\text{MeV } 90\%, 0.38\,\text{MeV } 10\%)$$

Both pep and Be^7 neutrinos are monoenergetic. The pp flux matching with solar luminosity implies $\sim 2 \times 10^8$ neutrinos sec^{-1}, occurs only if bulk of solar energy is generated by the pp chain reaction.

Final step: $He^3 + He^3 \rightarrow He^4 + p + p$, two p's recovered and then $p + p \rightarrow D + e^+ + \nu_e$ starting the chain.

If 50% goes via: $He^3 + He^4 \rightarrow Be^7 + \gamma$, then this reaction does not produce neutrinos. So the flux would be only half! If CNO dominates in the Sun, both C^{13} and N^{15} decays produce MeV neutrinos and flux in Chlorine detectors would have been $\sim$600 SNU! And not 6 SNU as predicted (1 SNU $= 10^{-36}$ captures atom^{-1} sec^{-1}). This implies that less than 1% of solar energy is due to CNO cycle.

If only pp chain were there, then the reactions could go on for ever as pp is regenerated. But $\sim$15% branch is $He^3 + He^4 \rightarrow Be^7 + \gamma$ (apart from dominant

85% $He^3 + He^3 \rightarrow He^4 + p + p$). So the cycle is now broken as He^3 is being used up!

$He^3 + He^3$ reaction was overlooked by the pioneers Bethe and Critchfield in 1938 (also by Weizaker). Lauritsen pointed it out only in the 1950's. So the main reaction was overlooked for more than 15 years!

HEP neutrinos have the highest energy:

$$He^3 + p \rightarrow He^4 + e^+ + \nu\,(17.5\,\mathrm{MeV})$$

Flux on Earth is of the order of a millionth of the pp flux $10^4\,\nu'\mathrm{s\,cm^{-2}\,s^{-1}}$!

The boron-8 reaction, that is $B^8 \rightarrow Be^8 + e^+ + 14.7\,\mathrm{MeV}$ produced 14 MeV neutrinos. The flux associated with this is of the order of $6 \times 10^6\,\nu'\mathrm{s\,cm^{-2}\,s^{-1}}$, which is about 10^{-4} pp flux.

The reaction $He^3 + p \rightarrow Li^4 + \gamma$, would have been predominant if the mass of Li^4 were less than or about equal to the combined masses of He^3 and p. The chain would have terminated with high energy beta decay and Sun's lifetime reduced by 40%. It is Obvious that the Li^4 production is absent as high energy neutrino flux would be $5 \times 10^{10}\,\nu'\mathrm{s\,cm^{-2}\,s^{-1}}$ and not $6 \times 10^6\,\nu'\mathrm{s\,cm^{-2}\,s^{-1}}$ as observed.

The primary reaction $p + p \rightarrow D + e^+ + \nu_e$ has such a small cross section $\sim 10^{-47}\,\mathrm{cm}^2$ (10 sheds) that it has not been observed or produced in the laboratory! Also this small cross section gives the long life for the pp cycle enabling the Sun to shine for ten billion years. In stars a few times solar mass, CNO cycle ('half life' of only few million years) dominates, so that a ten solar mass star would last only ten million years.

15 Search for Solar Axions

Axions were proposed by particle physicists to render theory of strong interaction (quantum chromo dynamics – QCD) invariant with respect to time reversal or CP transformation. By introducing a new pseudoscalar field, the time reversal violating phases can be removed by rotating them into the complex phase of the new field.

If V is the VEV of the axion field, $\langle\phi_a\rangle = V\exp(i\theta)$ and axion mass is $m_a = m_\pi f_\pi/V$. Where, m_π is the pion mass, f_π the pion coupling. For $V \sim 10^8\,\mathrm{GeV}$, the mass works out be of the order of eV. Axions are the favoured candidate for CDM in galactic halos. As axions can couple to photons and charged matter they are expected to be copiously produced in cores of stars and especially in red giants and evolved stars.

One can estimate the rate of axion production as $\sim n_C^2 \sigma_a v_{th} \times$ volume of the core. Where, n_C is the central number density of nucleons and v_{th} is the thermal velocity. $\sigma_a \sim \sigma_\pi f_\pi^2/V^2, \sigma_s \sim 10\,\mathrm{mb}$. We get the axion production rate in the Sun as $\sim 10^{42}\,\mathrm{s^{-1}}\,(10^8\,\mathrm{GeV\,V^{-1}})^2$ giving solar axion flux on earth as $\sim 10^{14}\,\mathrm{s^{-1}\,cm^{-2}}$. (Sivaram 1985, 1987)

The solar axions would have a broad spectrum of energies centered around the Sun's interior temperature, i.e., 1 keV, i.e., if converted to photons would have their energies in the soft X-ray range.

16 Axions and CAST

The CERN Axion Solar Telescope (CAST) is at present most sensitive axion helioscope. It uses a prototype LHC magnet to reconvert solar axions to 'visible photons' with mean energy of $\sim$4 keV (9 T magnetic field). CAST took data with an evacuated conversion volume in 2004, being sensitive to axions with mass <0.01 eV. Final configuration: He^{-3}, buffer gas, probe axion parameter space, mass region: 0.4–1 eV.

CAST may also close the gap to existing HDM limits on axion to photon coupling strengths. Most sensitive detector of CAST is the X-ray telescope, providing a quantum efficiency of 95% in the region of interest for solar axions (1–7 keV). X-rays from axion to photon conversion are focused on a CCD chip into a focal point of size of a few mm^2. No axions have been detected so far.

17 Why is the Sun Getting More Luminous?

The Sun is becoming more luminous and after hundred million years it would be 1% brighter! What is the basis for this? This has to do with the increasing helium in the Sun's core. Every second, thermonuclear reactions produce 560 million tons of helium in the core. This raises the 'molecular weight' μ of the gas. This lowers the gas pressure given by: $P = \rho RT \mu^{-1}$. So to sustain the equilibrium, core temperature T has to become higher.

Thus for a change $\Delta\mu$, we have: $\Delta\mu/\mu \propto \Delta T/T$. In about a hundred million years, knowing the amount of helium already present in the core ($\sim$30%), we can easily show that $\Delta\mu/\mu \sim 0.003$. This implies ΔT increases by 0.3%. Since for the pp chain the average interaction rate depends on $T^{3.7}$, we get the power output $\sim 3.7\Delta T/T \sim 1\%$.

18 Solar Evolution from ZAMS

The luminosity of the Sun with respect to the composition is given by:

$$L \propto \frac{M^{5.3}\rho^{0.117}\mu^{7.5}}{\kappa_0}$$

Estimate for change in luminosity with time as composition changes is given by:

$$\frac{L(t)}{L(0)} = \left[\frac{\mu(t)}{\mu(0)}\right]^{7.5}, \quad \text{where}\, \mu(t) = \frac{4}{3+5X(t)}.$$

The metal content is small compared to H and He in the Sun. We know how *X(t)* changes with time:

$$\frac{dX(t)}{dt} = -\frac{L(t)}{MQ}, Q = 6 \times 10^{18}\, \text{ergs}\, g^{-1}$$

The time rate of change of composition and hence the luminosity is given by:

$$\frac{d\mu(t)}{dt} = \frac{5}{4}\mu^2(t)\frac{L(t)}{MQ} \text{ and } \frac{dL(t)}{dt} \approx \frac{\mu(0)}{MQ}\frac{L^{17/15+1}(t)}{L^{17/15+1}(0)}$$

Integrating and expressing in units of L_{Sun} we have:

$$\frac{L(t)}{L_{sun}} = \frac{L(0)}{L_{sun}}\left[1-0.3\frac{L(0)}{L_{sun}}\frac{t}{t_{sun}}\right]^{-15/17} (\text{where}, t = t_{sun}, \text{for}\, L = L_{sun})$$

Form the solution of the above equation, the luminosity on ZAMS, $L(0) \approx 0.75 L_{sun}$.

So in a billion years from now the Sun would be 12% more luminous. So three billion years ago Sun would have been 30% dimmer (paradox of the faint Sun etc!). However the B^8 branch reaction (giving rise to the high energy neutrinos) goes at least as T^{18}. So the fact that we observe half the (>5 MeV) neutrino flux in Kamiokande, etc. enable the core temperature of the Sun to be fixed to an accuracy of less than 0.3%!

Ironically, while we are so sure about the Sun's core temperature, Earth's core temperature is uncertain by more than 50%. (Estimated range between 4,000 and 10,000°!) Thanks to helioseismology and solar neutrinos, we know much more about the solar interior than 30 km below Earth's surface!

18.1 Five Phases of Sun's Evolution

MS: This phase spans eleven billion years for the Sun. Solar radius currently increases by 15 cm yr^{-1}. During MS, the diameter is expected to increase by 40% and the luminosity by 100%.

RG: During this phase there will be increase in stellar radius. Core contraction will occur and the shell hydrogen burning will take place. Temperature increases to 50 million kelvins. Increase in l by 2000. H burning shell leads to expansion and cooling of outer layers. Radius estimated to increase by factor of 200, may reach 0.8 earth orbit.

Central density will be of the order of $200\,\mathrm{kg\,cm^{-3}}$ $\sim 10^3\%$. Core temperature will be of the order of 200×10^6 K. Helium burning starts. The mass loss will be of the order of $0.3\,M_{Sun}$. RGB lasts 6×10^8 years (transition from MS to RG will take 7×10^8 years).

Core He burning: Core He flash. Degeneracy. Luminosity produced by He flash, raises T_C above 200×10^6 deg. Most of the energy expands inner layers, remove degeneracy. Surface luminosity decreases to $100\,L_{Sun}$, radius reduces by a factor of 5. Stable core He burning phase lasts $\sim 10^8$ years.

AGB phase: similar to RG. But is distinctive in having two burning shells, one of H, the other He. Surface inflates again to $>200\,R_{Sun}$. Can Earth be engulfed? Perhaps not! Because of mass loss + tidal effect.

During this phase there is high mass loss rate of the order of $10^{-4}\,M_{Sun}\,\mathrm{yr}^{-1}$. Ejection of entire envelope down to H burning shell will last $<$ few $\times\, 10^4$ years. Hot core of temperature $>10^5$ K becomes visible illuminating ejected envelope. The system becomes visible as a planetary nebula. This phase lasts $<10^7$ years.

WD: The remnant will be about $0.55\,M_{Sun}$. C–O core left after the envelope ejection. The cooling time is greater than 10^{10} years.

19 Fate of Other Stars

The fate of these stars depends on their mass. There can be three possible scenarios by which these stars can meet their fate.

- $>1.8\,M_{Sun}$: no core He flash.
- $<0.5\,M_{Sun}$: no core He burning. End up as He WD's.
- $>10\,M_{Sun}$: massive stars produce heavier elements. Explode as supernovae. Remnant neutron star or black holes.

20 Concluding Remarks

In this article, the Sun's place in the universe and its consequences has been discussed. Study of the various aspects of the physics and astronomy of the Sun, especially in recent years, has enabled us to understand how stars shine and evolve, how they could end their lives, etc. The physics involved in the various stages, especially the energetics of the nuclear reactions, the neutrino production and detection, the dying phases, the compact stellar remnants, etc. have been much clarified. The Sun has been a laboratory for understanding various atomic and nuclear processes and complex magneto- hydrodynamics and fluid phenomena. Various phenomena in particle physics like neutrino oscillations, experiments for detecting dark matter candidates like axions have been stimulated by trying to understand the solar interior. A lot of enigmas still remain to be understood about our nearest star, from the core to the corona! Current and future planned solar space probes could provide crucial new inputs. The future for solar physics looks as bright as the Sun!

References

Erdelyi, R. et al: Astron. Nachr. 726–734, (2007)
Mishurov, S.: Astron. Astrophys. **341**, 81, (2000)
Radick, J. et al: Astrophys. J. **118**, 239, (2001)
Sivaram, C.: Earth Moon Planets **37**, 155, (1987)
Sivaram, C.: Astrophys. Space Sci. **116**, 39, (1985)

Solar Interior

Ashok Ambastha

Abstract The solar interior is not visible by direct means, and until a few decades back its understanding was based only on the surface observations pertaining to its global properties (e.g., temperature, luminosity, radius, *etc*.). Solar neutrinos and global solar oscillations have provided more "direct" probes of the Sun's internal structure and dynamics. In these lectures we discuss some aspects of standard solar models, the neutrino problem and recent developments in helioseismology.

1 Introduction

The Sun, like all stars, is a self gravitating ball of gas, where two basic forces are at play: (1) gravity that causes stars to collapse, (2) pressure that causes stars to expand. There are stages in a star's life where gravity wins and the star begins to collapse, and also where pressure wins and the star begins to expand. However, the Sun is currently in a state of hydrostatic equilibrium where for the most part gravity and pressure balance each other. Although the outer layers of solar atmosphere are visible using various observational techniques, the large photospheric opacity does not allow revealing the inner layers of the Sun. Indirect inferences about the solar interior have been drawn mostly through mathematical models using the global and surface features of the Sun. More recently, solar neutrinos and solar oscillations have been providing increasingly important tool for detailed studies of the interior of the Sun.

2 Solar Modeling Procedure

A solar model is the solution to a set of equations describing the physical processes occurring within the Sun. These equations are often divided into two subsets: the basic structure equations and the chemical evolution equations. In general, the

A. Ambastha
Udaipur Solar Observatory, Physical Research Laboratory, Udaipur 313001, India
e-mail: ambastha@prl.res.in

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_2,

equations cannot be solved analytically and must be solved numerically. Therefore, a solar model is usually comprised of a set of tables describing the conditions (chemical composition, density, opacity, luminosity, mass, pressure, temperature, etc.) at different depths in the star. While modeling the interior of the Sun we have the advantage that we know a lot more about the Sun than other stars. But this also makes modeling more challenging, as any useful solar model must yield correct luminosity, radius, and age of the present Sun, i.e., mass $= 1.989 \times 10^{33}$ g, radius $= 6.959 \times 10^{10}$ cm, luminosity (energy output) $= 3.8418 \times 10^{33}$ erg s^{-1}, and age $= 4.57 \times 10^{9}$ years.

2.1 Basic Equations of Solar Structure

For most parts, stars are spherically symmetric, i.e., their internal structure is only a function of radius and not of latitude or longitude. This means that we can express the properties of stars using a set of 1-D equations, rather than a full set of 3-D equations. The main equations describing the star's hydrostatic equilibrium concern the following physical principles: (1) conservation of mass, (2) conservation of momentum, (3) conservation of energy, (4) transport of radiation, (5) nuclear reaction rates, and (6) change of abundances by various processes.

Conservation of mass: This is usually given in the form $\partial m/\partial r = 4\pi\rho r^2$. However, in case of the Sun it is more convenient to consider the mass m interior to a sphere of radius r as the independent variable, and we get

$$\frac{\partial r}{\partial m} = \frac{1}{4\pi\rho r^2}.$$

We have assumed here that the Sun is in nearly steady state, but in general $r = r(m;t)$ and thus we retain the partial derivatives. The reason to take m as the independent variable is that, except at the very beginning, the mass loss has been negligible and we know $m_\odot$ throughout the whole period to which our model will be applicable. On the other hand, we know the radius only for the present Sun, and the model has to provide it as a result.

Conservation of momentum: Consider the Sun in the hydrostatic equilibrium, we have $\partial P/\partial r = -\rho g$. Using conservation of mass and $g = -Gm/r^2$ we obtain

$$\frac{\partial P}{\partial m} = \frac{Gm}{4\pi r^4}.$$

It is to be noted that this equation does not describe the collapse in the protostar phase.

Energy balance: Let $L(m)$ be the luminosity generated inside the sphere of mass m, ε the energy generation per unit mass, and S the entropy per unit mass (i.e., the specific entropy). Then, we have

$$\frac{\partial L}{\partial m} = \varepsilon - T\frac{\partial S}{\partial t}.$$

During the main sequence evolution the interior of the Sun is very close to thermal equilibrium and the heating/cooling effect $\partial S/\partial t$ is small.

Equation of state: The pressure inside the star arises from momentum transfer by particles and photons. In the case of the Sun the pressure has two components, the gas pressure (p_G) and the radiation pressure (p_R). The radiation pressure is important only in the atmosphere and solar wind. For the present solar model we thus apply the perfect gas law, $p = \rho RT/\mu$, where, R is the gas constant and μ is the mean molecular weight (in a.m.u.). The constituents of the gas are usually denoted by X: fractional abundance of H, Y: fractional abundance of He and Z: fractional abundance of the heavier elements. The ionization of elements adds particles to the gas and reduces μ, e.g., the mean molecular weight of fully ionized electron–proton plasma is half of neutral hydrogen gas.

In order to integrate the above set of differential equations we need the boundary conditions. These are generally imposed at the center of the Sun at $r = 0$, $M = 0$, $dT/dr = 0$, $L(0) = 0$, and at the surface, $R = R_\odot$, $M = M_\odot$, $L = L_\odot$.

Equations of chemical evolution: The equations of chemical evolution depend solely upon which nuclear processes are occurring in the interior of the star; for example, in the Sun, the most important fusion process is the pp-chain, and to a much lesser extent, CNO cycle. In a simplistic model one could choose to keep track of five elements: hydrogen, two species of helium, carbon, and nitrogen.

2.2 Energy Production

Solar energy is produced in the central core mostly by hydrogen burning where prospects for thermonuclear fusion are more than met (Fig. 1). Of all nuclei, protons have the smallest charge, which is important to get two particles sufficiently close to each other by overcoming their electrostatic barrier of about 1 MeV. This is significant considering that the central temperature of the Sun is $\sim 1.5 \times 10^7$ K, i.e., 1.3 keV only. Thus very high density is required in order that sufficient number of close encounters can take place, which is available in the solar core where density is $\sim 160\,\text{gm}\,\text{cc}^{-1}$. There are two main reaction chains for stellar (and solar) fusion, and due to high H-abundance, and low nuclear charge, the most likely nuclear fusion process is pp-chain reaction. Around 99% of energy comes from the pp-chain and about 1% from the CNO cycle.

The pp-chain reaction is as follows:

$$^1\text{H} + {}^1\text{H} \Rightarrow \text{D} + \text{e}^+ + \nu + 1.44\,\text{Mev}$$

$$^2\text{D} + {}^1\text{H} \Rightarrow \text{He} + \gamma + 5.49\,\text{Mev}$$

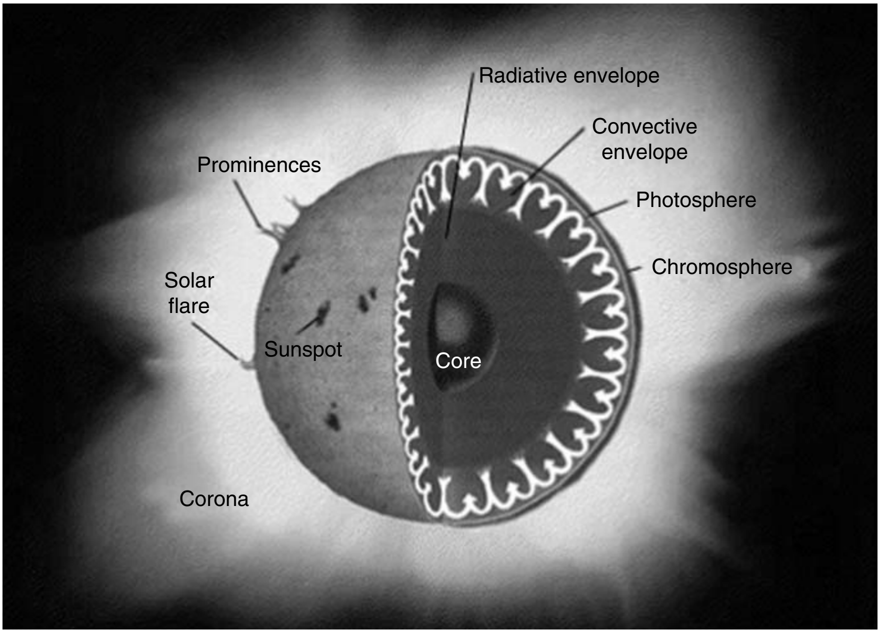


Fig. 1 Schematic diagram of solar structure showing the interior and the exterior

$$^{3}\mathrm{He} + {}^{3}\mathrm{He} \Rightarrow {}^{4}\mathrm{He} + {}^{1}\mathrm{H} + {}^{1}\mathrm{H} + 12.85\,\mathrm{Mev}$$

For each occurrence of 3rd reaction, first two must have occurred twice, so

$$4\,{}^{1}\mathrm{H} \Rightarrow {}^{4}\mathrm{He} + 2\mathrm{e}^{+} + 2\nu + 2\gamma + 26.7\,\mathrm{Mev},$$

i.e., 4 protons fuse to give one helium nucleus, 2 neutrinos and release of energy, carried by the γ-ray. The energy generation rate can be described as simple power law to more complicated expressions involving nuclear reaction rates.

2.3 Energy Transport Through the Solar Interior

The energy produced by the nuclear fusion is transported from the central core by γ-rays. Opacity of solar material due to the high density impedes the radiative flow of the energy generated in the core, and the γ-ray photons are continuously absorbed and re-emitted by the gas in the radiative zone. The processes that operate simultaneously are: (a) bound-bound transition, (b) bound-free transition (photo-ionization), (c) free-free transition, and (d) scattering. The outward energy diffusion takes around 1.7×10^{5} years to reach the base of the convection zone at about $0.71\,r_{\odot}$. At this distance the temperature falls from 15 million kelvins in the core to about 2 million kelvins. The radiative envelope is surrounded by this cooler and more opaque convective envelope. Therefore it becomes less efficient for energy to move by radiation outside the radiative zone, and heat energy starts to build up.

Convection occurs when the temperature gradient (the rate at which the temperature falls with radius) becomes larger than the adiabatic gradient (the rate at which the temperature would fall if a volume of material were moved higher without adding heat). Where this occurs, a packet of material moved upward will be warmer than its surroundings and will continue to rise. This process of transport by convection dominates above $0.75\, r_{\odot}$. Convection is a much faster way of energy transport than radiation in an opaque medium, and takes only about 10 days for the heated gas to rise through the convection zone. Convection cells nearer to the outside are smaller than the inner cells. The top of each cell is called a *granule*, which appears as tiny specks of light when seen through a telescope. The solar surface is effectively a black body that absorbs all energy coming from the convection zone and radiates it out at the temperature of 5778 K. The density is very low, i.e., only $0.0000002\,\mathrm{g\,cm^{-3}}$ (about 1/10,000th the density of air at sea level) at the visible surface.

3 Standard Solar Model

A standard solar model is defined as a model where the observed physical inputs (nuclear reaction rates, diffusion coefficients, opacity tables, equation of state) are not changed to bring the model in better agreement with the Sun. The agreement between the Sun and the model is an indication of how good are the input parameters. The major assumptions are the spherically symmetric Sun, negligible effect of distortions due to rotation/magnetic field/tidal forces, mechanical and thermal balance, standard nuclear and neutrino physics, uniform initial chemical composition, no significant mass loss nor any accretion, no mixing of nuclear reaction products outside the convection zone, gravitational settling of helium and heavy elements beneath the convection zone, and no transport of energy or momentum by waves. The Sun is assumed to be in steady state, i.e., it is neither expanding, nor contracting so that it is in hydrostatic and thermal equilibrium. The interior model is matched to outer atmospheric model (Vernazza et al. 1981). Figure 2 shows radial variations for some physical parameters obtained by one such "Standard Solar Model".

4 Solar Neutrinos

One of the most famous problems of solar physics has been the so-called solar neutrino deficiency, which means that the standard solar models predict a larger neutrino flux than observed. *"The use of a radically different observational probe may reveal wholly unexpected phenomena... perhaps, there is some great surprise in store for*

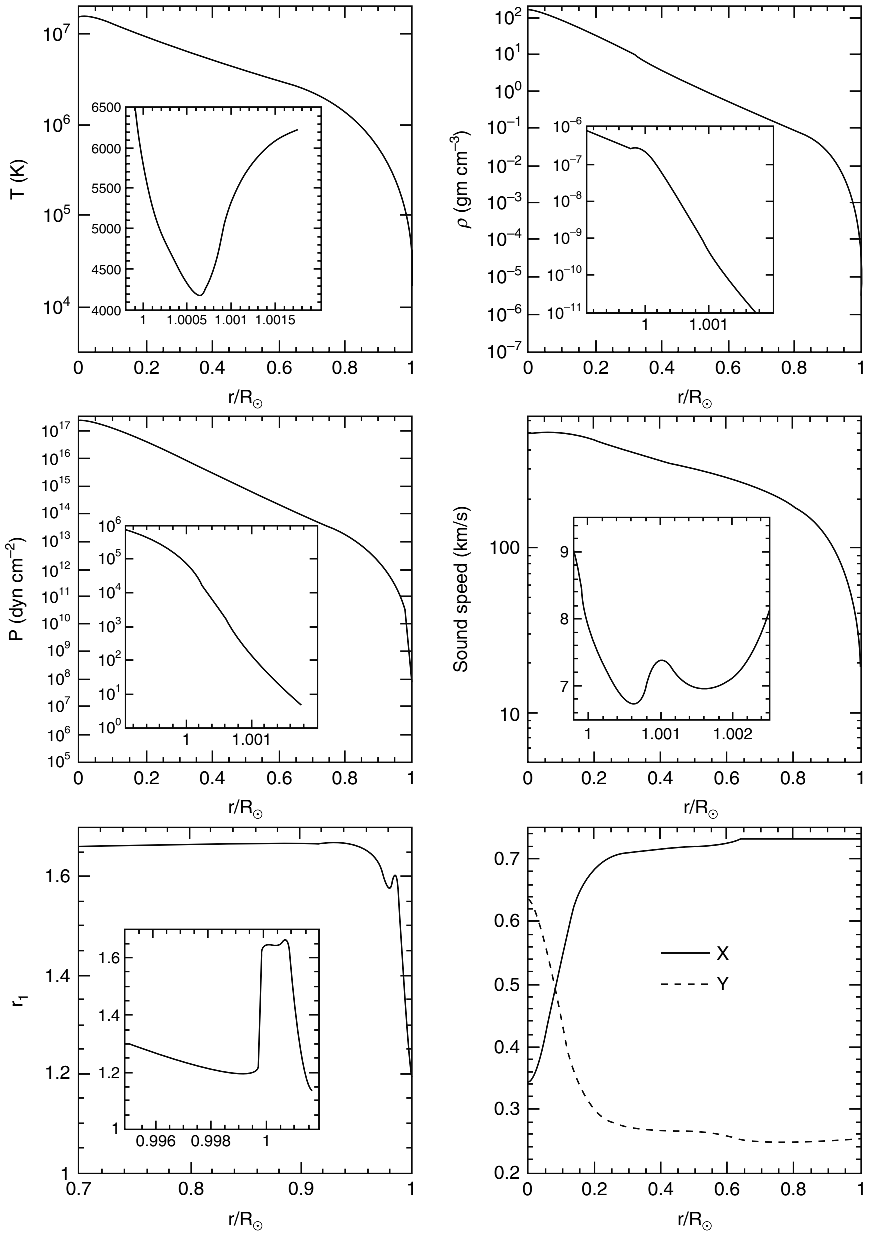


Fig. 2 Temperature, density, pressure, and sound speed decrease as r is increased, except some departures as seen near the surface (inset). Also shown are adiabatic index Γ_1 and H/He abundance profiles. The density drops from $20\,\mathrm{g\,cm^{-3}}$ (about the density of gold) down to only $0.2\,\mathrm{g\,cm^{-3}}$ (less than the density of water) from the bottom to the top of the radiative zone. The temperature falls from 7,000,000 K to about 2,000,000 K over the same distance (after Brun et al. 1999)

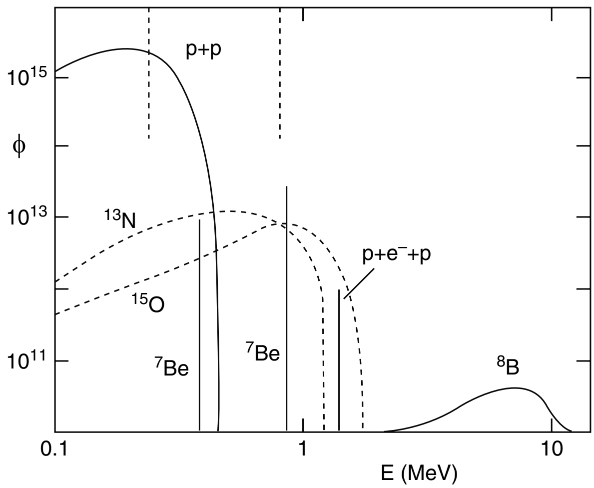


Fig. 3 Energy spectrum of the solar neutrino flux at 1 AU, (after Bahcall 1979). The dotted vertical lines mark the thresholds at 233 and 814 keV for the ^{71}Ga and ^{37}Cl experiments, respectively

us when the first experiment in neutrino astronomy is completed", said Bahcall in 1967. The motivation of the Chlorine experiment set up in Homestake Gold Mine by Ray Davies was to "*see*" into the interior of the Sun and thus verify directly the hypothesis of nuclear energy generation in stars. This was essentially a "neutrino detector" designed to detect solar neutrinos. Since neutrinos are very weakly interacting, such detectors are very large in volume, and are often built deep underground in order to reduce effects of cosmic rays and other background radiation. Before the detectors were built, solar models had predicted a production of 2×10^{38} neutrinos per second. Figure 3 shows the calculated neutrino fluxes at 1 AU. Here the energy of neutrinos is important, because it easier to detect higher-energy neutrinos than the main part of the spectrum.

The goals of solar neutrino experiments are essentially (a) to test physics of nuclear reactions operating in the solar core, (b) to confirm nature of thermonuclear energy generation in the Sun: pp-chain, CNO cycle, and (c) to study the properties of neutrinos.

There are three basic types of solar neutrino detectors: the ^{71}Ga experiment with the lowest energy threshold (233 keV), the ^{37}Cl experiments with threshold of 814 keV, and large water detectors having the highest threshold of about 5 MeV.

Radiochemical experiments – chlorine and gallium (charge current or CC reactions): The first solar neutrino experiment, Homestake gold mine in South Dakota beginning in 1967, used 615 tons of the ordinary cleaning fluid, tetrachloroethene, C_2Cl_4, whose chlorine is converted to argon in the first reaction given below:

$$^{37}\mathrm{Cl} + \nu_e \xrightarrow{\mathrm{CC}} {}^{37}\mathrm{Ar} + e^- \quad (\geq 0.814\,\mathrm{MeV})$$

$$^{71}\mathrm{Ga} + \nu_e \xrightarrow{\mathrm{CC}} {}^{71}\mathrm{Ge} + e^- \quad (\geq 0.233\,\mathrm{MeV})$$

Water detectors: Kamiokande water experiment started operating at Kamioka, Japan in 1987. Water detectors detect neutrinos through Cerenkov light from elastic ν-e^- scattering (*ES reaction*) if the recoil energy of the electron is at least 5 MeV. While the water tank can observe only the ^{8}B neutrinos, it is possible to determine their arrival direction and thus can identify which neutrinos come from the solar direction. SuperKamiokande (SK), that became operational in 1996, also uses water Cerenkov effect.

$$e^- + \nu_x \xrightarrow{\mathrm{ES}} e^- + \nu_x' \quad (\geq 5\,\mathrm{MeV})$$

Sudbury Neutrino Observatory (SNO) began collecting data in 1999. It used heavy water as the detecting medium for Cerenkov radiation that is emitted when an incoming neutrino creates a free electron or neutron (sensitive to both charge, or *CC*, and neutral current or *NC reaction*). The deuterium present in the heavy water is dissociated by a neutrino. All three flavors of neutrino ν_x (ν_e, ν_μ, ν_τ) participate equally in this process.

$$d + \nu_e \xrightarrow{\mathrm{CC}} p + p + e^-, \quad (\geq 5\,\mathrm{MeV})$$

$$d + \nu_x \xrightarrow{\mathrm{NC}} p + n + \nu_x', \quad (\geq 5\,\mathrm{MeV})$$

$$e^- + \nu_x \xrightarrow{\mathrm{ES}} e^- + \nu_x', \quad (\geq 5\,\mathrm{MeV})$$

Due to the extremely weak interaction between neutrinos and ordinary matter, a special unit for the solar neutrino flux, the solar neutrino unit (snu) is used, where 1 snu corresponds to the capture of one neutrino per second per 10^{36} target atoms. Typical predictions for observable fluxes for the gallium, chlorine, and water detectors and observational results are listed in Table 1. It shows the deficit of measured neutrino flux as compared to the predicted flux.

The neutrino deficit or the famous solar neutrino problem has been attributed to several different causes. That it could be an observation problem was a popular explanation as long as the Homestake observations were alone. But after the gallium and water detector results, this appeared very unlikely. For a long time it was thought that the origin of the problem would be in erroneous solar models. Quite a number

Table 1 Predictions for observable neutrino fluxes

Expt.	Chlorine	Gallium	SK	Sno
Threshold (MeV)	0.834	0.233	5	5
Predicted flux (in snu)	7.7 ± 1.1	130 ± 7	1.0 ± 0.15	
R	0.33 ± 0.03	0.55 ± 0.03	0.465 ± 0.015	0.36 ± 0.015

Here, R = (Measured neutrino flux)/(Predicted model neutrino flux) of attempts to correct the models have been made but without much success, as they lead to problems elsewhere in the models.

4.1 Non-Standard Solar Models

Reduction of the central temperature is one strategy to look for non-standard solar models, which would reduce the ^{8}B neutrino flux and help at least with the original chlorine experiment problem. This is a reasonable idea, as the ^{8}B neutrino rate is proportional to T_c^{18} where T_c is the temperature in the center. A lowering of the central temperature by only 7% would be sufficient, but this directly leads to problems with the ^{7}Be neutrino fluxes whose temperature dependence is proportional to T_c^8. Furthermore the lower-temperature solar models would face problems in view of other observations, in particular, the solar oscillations.

Reduction of the relative abundance of heavy elements (low Z) is another possibility, but this would reduce opacity κ and thus the temperature gradient. This does not seem a way out for solar-type stars in the first place and also contradicted by the solar oscillation observations. One more suggestion was a rapidly rotating core which would lower the thermal pressure. But this is again inconsistent with oscillation results that indicate that the core rotates at nearly the same angular speed as the surface (cf., Sect. 6.2). Furthermore, a strong internal magnetic field could lower the central temperature, but it should be very intense to have a significant, say 10%, contribution to the pressure of the core.

4.2 Origin of Neutrino Problem in Non-Solar Effects?

The fact that the gallium experiments also fall a factor of 2 short of the predicted neutrino flux indicated that the lowering of the temperature and shifting the neutrino peak to lower energies is not the right solution. More recently it has become evident that the solution lies in the physics of neutrinos. Certain non-standard elementary particle models predict that the neutrinos have finite masses. In such a case the neutrinos can oscillate between their three flavors. The nuclear reactions in the Sun produce electron neutrinos (ν_e) only and the predicted fluxes are for electron neutrinos. As these neutrinos travel through space, and if a large enough fraction of solar neutrinos would transform to other neutrino flavors before reaching the Earth, this would provide an elegant solution to the whole problem.

In 1998 the Superkamiokande observations indicated oscillations between muon neutrinos (ν_μ), produced by cosmic rays in the atmosphere of the Earth, and tau neutrinos (ν_τ). This did not solve the problem of solar neutrinos directly but it gave strong evidence that neutrinos are not mass-less particles. In 2001 the first results from the heavy water (D_2O) detector at the Sudbury Neutrino Observatory (SNO) indicated that the solar neutrino problem was coming to its final solution as this detector was able to distinguish between the total neutrino flux and the ν_e flux. The successful results were published in September 2003 (cf., http://www.sno.phy.queensu.ca/). The consistency of the standard solar model with solar oscillation observations and the failure of the non-standard solar models to solve the neutrino problem together with the new neutrino observations indicated that "the solar neutrino problem" was finally coming to be resolved. It also indicated that the solution lies in the properties of the neutrinos. In fact the failed attempts to find the solution within the solar models have contributed enormously to our detailed knowledge of the interior of the Sun.

5 Solar Oscillations

Surface of the Sun undergoes a series of mechanical vibrations of the gases which is observed as Doppler shifts of spectrum lines, oscillating with a period centered at 5-min (3 mHz). These oscillations were first reported by Leighton et al. (1962), using spectroheliograph recordings by simultaneously measuring the relative intensity in the red and blue wings of a spectral line. The exact nature of the oscillations was not understood till Ulrich (1970), Leibacher and Stein (1971) theoretically interpreted them as resonant, standing waves trapped below solar surface. This was observationally confirmed by Deubner (1975) by the $(k - \omega)$ diagnostic diagram which showed the oscillations to have a spectrum of discrete frequencies. A typical power spectrum of solar oscillations is shown in Fig. 4.

It has been demonstrated that the Sun oscillates in over 10^7 modes with different spatial patterns and temporal frequencies visible at solar surface. High frequency

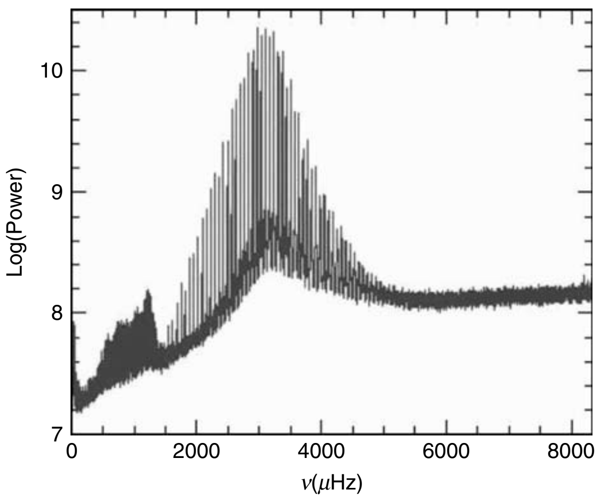


Fig. 4 Power spectrum of solar oscillations

Table 2 Velocity fields (line of sight component) at solar surface

Earth's rotation and orbital velocity	Solar rotation	Convective motions	Solar oscillations
$500\,\mathrm{m\,s^{-1}}$	$2\,\mathrm{km\,s^{-1}}$	$1\,\mathrm{km\,s^{-1}}$	$<1\,\mathrm{m\,s^{-1}}$

Table 3 Intensity variation at solar surface

Convection (granulation)	Sunspots	Solar oscillations
15%	factor of 10	10^{-6}

acoustic (p-) modes are driven by pressure forces, low frequency gravity (g-) modes are driven by buoyancy forces, and the fundamental or surface gravity (f-) modes are determined by surface gravity and are largely independent of stratification. The velocities and intensity variations associated with solar oscillations as compared to some other features are given in Tables 2 and 3.

Libbrecht et al. (1990) published a table of solar oscillation frequencies from observations at *Big Bear Solar Observatory* (BBSO) extending over summer months. This formed the basis of most early studies in helioseismology until 1996 when data from the ground-based *Global Oscillations Network Group* (GONG) and the space-borne *Solar and Heliospheric Observatory* (SOHO) became available. These observations have been providing accurately measured oscillation frequencies which stringently constrain the admissible solar models.

5.1 Helioseismology: The Science of the Ringing Sun

The science studying wave oscillations in the Sun is called *helioseismology* in analogy with terrestrial seismology to learn about Earth's interior by monitoring waves caused by earthquakes. Solar global modes of oscillations propagate in the solar interior, and temperature, composition, and velocities influence the oscillation characteristics, which yields insights into conditions in the solar interior. Therefore, precise identification and accurate measurement of oscillation modes is the mainstay of helioseismology.

The Sun act as a resonant cavity. In fact no physical boundaries exist inside the Sun, however, acoustic cavities exist due to density and temperature gradients that reflect or refract sound waves (Fig. 5). Acoustic waves are trapped in a region bounded on top by a large density drop near the surface, and on the bottom the increase in sound speed with depth refracts the downward propagating wave back toward the surface. Thus a standing wave is created.

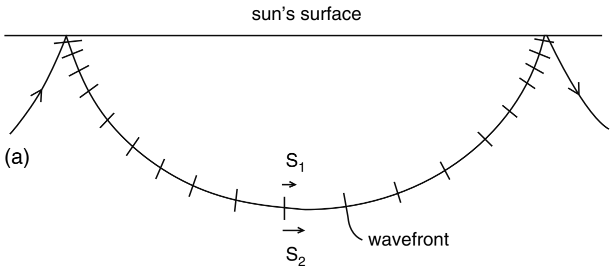


Fig. 5 Reflection and refraction of a wavefront

5.2 Solar Global Modes of Oscillation

Global helioseismology involves decomposition of time series of full-disk solar photospheric Doppler image data. For spherically symmetric Sun, global oscillations can be best given by spherical harmonic functions for an observational parameter, such as, velocity or intensity at the surface:

$$\tilde{\sigma}(r,\theta,\varphi,t) = \sum_{nlm} \sigma_n(r) Y_l^m(\theta,\varphi) \exp(i\omega t),$$

where Y_l^m are spherical harmonic functions of latitude θ and longitude φ, and r is the depth. A given mode with frequency ω is characterized by three integers, i.e., $l \geq 0$ (the degree; i.e., total no. of nodal planes slicing the solar surface), m (the azimuthal order; i.e., total number of planes cutting the equator perpendicularly $|m| \leq l$, and $n =$ the radial order ($-\infty < n < \infty$). For each pair l and m, there is a discrete spectrum of modes with distinct frequencies ω having different spatial structures represented by eigenfunctions $\sigma_n(r)$, oscillatory in space with nodes determined by the radial order n. Modes of positive n are defined to be p-modes, negative n as g-modes, and the $n = 0$ mode is called the f-mode. The degree l measures the horizontal component of the wavenumber, i.e., $k_h = \sqrt{l(l+1)}/r$ at radius r, and frequency $\nu = \omega/2\pi = 1/T$.

Amplitudes with depth of some acoustic and gravity modes from a theoretical model are shown in Fig. 6, where the eigenfunctions, say Δr, are scaled for clarity. Radial extension of modes is delimited to resonant cavity by two turning points, outside which the mode decays. Outer turning point is the same for all modes, while the inner turning point is different for every (l, n) pair.

5.3 Trapping of the Modes

The radial component of displacement can be described by a single second order differential equation (cf., Christensen-Dalsgaard 2003):

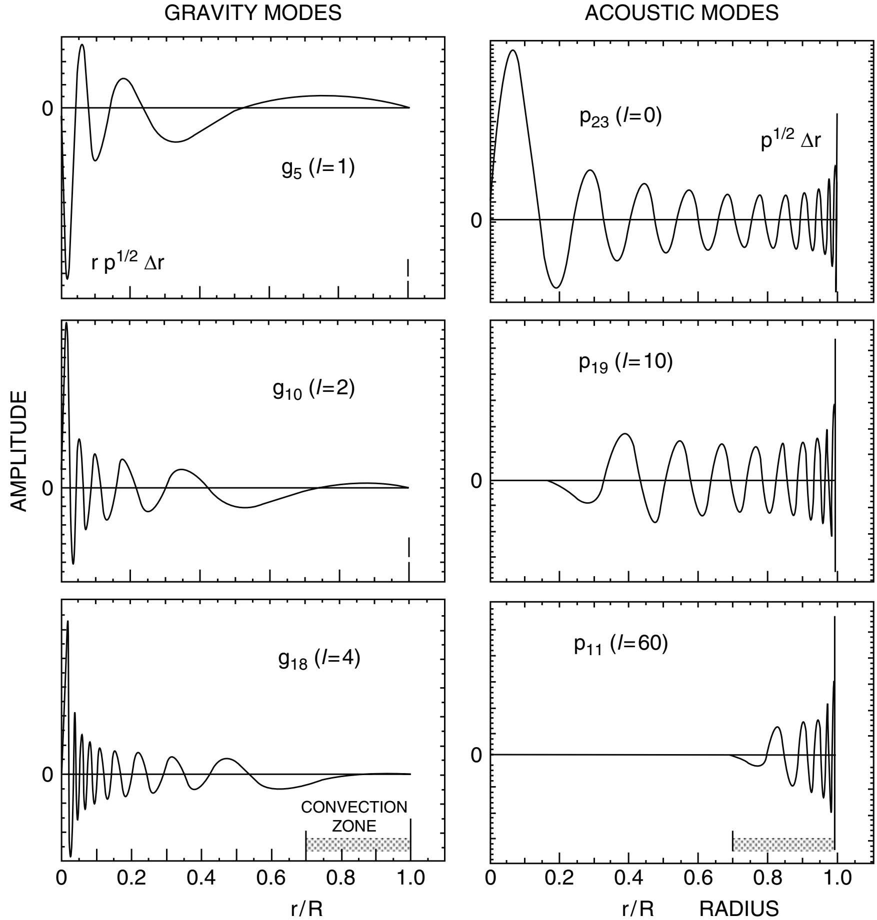


Fig. 6 The eigenfunctions associated with some g- and p-modes with depth. The three p-modes with frequency $\sim$3.3 mHz ($T \sim 5$ min) shown are obtained by decreasing n and increasing l. As l increases, these are increasingly confined to the surface. For the three g-modes with frequency ~ 0.1 mHz ($T \sim 160$ min), and increasing l, n gives structures confined towards the core

$$\frac{d^2\xi_r}{dr^2} = \frac{\omega^2}{c^2}\left(1 - \frac{N^2}{\omega^2}\right)\left(\frac{S_l^2}{\omega^2} - 1\right)\xi_r$$

This equation represents non-radial oscillations in a crude approximation, and describes the overall properties of oscillation modes, and gives a reasonable accurate determination of their frequencies. It is evident that the characteristic frequencies, acoustic or Lamb frequency (S_l) and the buoyancy or Brunt–Väisälä frequency (N) play a very important role in determining the behaviour of these modes. These frequencies are defined as:

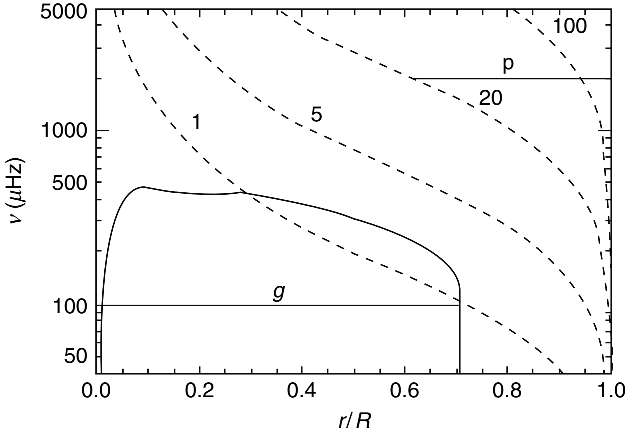


Fig. 7 The characteristic acoustic or Lamb frequency S_l (*dashed lines* for $l = 1, 5, 20$, and 100) and the Brunt–Väisälä frequency N (*solid lines*) against radius in a standard solar model. The heavy horizontal lines indicate the trapping regions for a g-mode with frequency $\nu = 100\,\mu$Hz and for a p-mode with degree $l = 20$ and $\nu = 2000\,\mu$Hz

$$S_l = \sqrt{l(l+1)} \quad c/r, \text{ and } N = \sqrt{g_0\left(\frac{1}{\Gamma_1}\frac{d\ln p_0}{dr} - \frac{d\ln\rho_0}{dr}\right)}$$

The Brunt–Väisälä frequency N is the frequency with which a small fluid element will oscillate about its equilibrium position when displaced, and it is independent of l. For $N^2 > 0$ the system is stable to small perturbations. On the other hand, the Lamb frequency is l dependant, and for $l = 0$, $S_l = 0$. S_l tends to infinity as r tends to zero and decreases monotonically towards the surface due to the decrease in c and increase in r. N^2 is negative in convection zone and positive in convectively stable region. These frequencies are illustrated in Fig. 7. The sharp maximum in N near the very centre is associated with the increase in the helium abundance in the region where nuclear burning is taking place.

5.4 Solar Dispersion Relation: The l–ν Diagram

Waves which resonate between the two turning points through a constructive interference form a mode. This happens only when there are an integral number of vertical wavelengths fitting in the cavity defined by the two points:

$$\int_{r_0}^{r_1} K_\nu(r)dr = n\pi + \varepsilon,$$

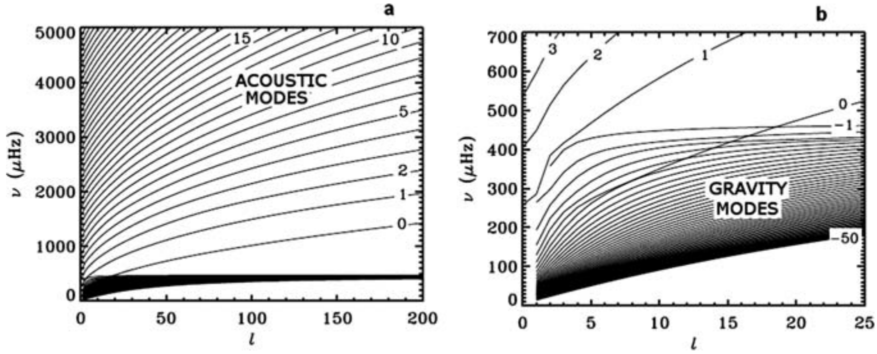


Fig. 8 Frequencies $\nu_{n,l}$ obtained from a theoretical solar model for acoustic (p-) and gravity (g-) modes. Loci of frequencies at fixed radial order n as l is varied are shown

where $K_v(r)$ the vertical wavenumber, and ε is a correction factor. The order n of the modes defines this dispersion relation between degree (through $K_v(r)$) and determines discrete frequency for every given l. Based on this relation, a diagram of frequencies as a function of degree can be plotted and ridges appear for every order n, the so called $l{-}\nu$ diagram (Fig. 8).

5.5 *Observational Requirements and the Instruments*

For λ6000 Å solar line (width ∼0.1 Å), a velocity corresponding to ∼1 m s^{-1} would shift it by ∼10^{-5} Å, so individual oscillation modes having amplitudes ∼0.1 m s^{-1} would require observational accuracy of parts per million. Maximum value of the frequencies depends on the time-interval Δt between observations, or cadence; so for a cadence of 1 min, the Nyquist frequency is $1/(2\Delta t)$ or 8.3 mHz. The frequency resolution of temporal spectrum is determined by the duration over which the observations are made; for example, we get $\Delta\nu = 1/86400 \approx 11.6\,\mu$Hz for a one day long data series.

The longer, continuous coverage of the Sun allows the mode frequencies to be determined with greater precision, which is important at low l region where the peaks in the spectrum are very narrow. There are three alternatives to obtain continuous observations over a long time period: (1) Observing from geographic South Pole, (2) Observing from a network of ground-based sites, such as, the Birmingham Solar Oscillation Network (BiSON), the International Research on Interior of Sun (IRIS), the Global Oscillation Network Group (GONG), and Taiwan Oscillation Network (TON), and (3) Observing from space-borne instrument in sun-lit (Lagrangian point), e.g., *Global Oscillations at Low Frequencies* (GOLF), and *Michelson Doppler Imager* (MDI) onboard the *SOlar and Heliospheric Observatory* (SOHO).

There are essentially two classes of observations: (a) un-imaged data in integrated light that give only the averaged velocity or intensity over the solar disk (this allows to detect only low degree $l \leq 4$ modes), and (b) imaged data over the solar disk with spatial resolution ($0 \leq l \leq N$, $N =$ number of pixels). Maximum value of l that can be studied depends on the spatial resolution of the observations. One can observe the oscillations in either the line-of-sight velocity via Doppler shift or in intensity. GONG and SOHO provide long time series imaged data. Operation of the GONG network of six ground-based observing stations started on May 7, 1995, at Learmonth (Australia), Udaipur (India), Tenerife (Canarias, Spain), Cerro-Tololo (Chile), Big Bear (USA), and Mauna Loa (Hawaii, USA). The initially used 256×256 CCD was upgraded in 2001 to a $1{,}024 \times 1{,}024$ CCD for higher spatial resolution to cover larger l-modes. Michelson Doppler Imager (MDI) on-board SOHO also has $1{,}024 \times 1{,}024$ CCD. It was launched in December 1995, and started observing on May 1, 1996. Contact with SOHO was lost temporarily in June 1998, but it was recovered back in February 1999. Frequencies of 400,000 modes have been determined for $l \leq 150$ and $1 \leq \nu \leq 4$ mHz using the GONG and SOHO observations.

5.6 The Analysis Techniques

Analysis of helioseismic data of oscillation frequencies is carried out by (a) forward method, where an equilibrium standard solar model is perturbed to obtain eigenfrequencies, which are then compared with observed mode frequencies, and (b) inverse method, where observed frequencies are used to infer the internal structure of the Sun. In forward modeling the depth-dependent coefficients of the equations describing the oscillations are computed from a solar model. Then these equations are solved subject to appropriate boundary conditions. The results are eigenfunctions and eigenfrequencies, i.e., the values of ω for which non-zero solutions exist. Now the solutions likely deviate somewhat from the observed frequencies. The direct approach is to make small corrections to those parameters of the solar model which have the largest uncertainties, e.g., corrections to the ideal gas law, the fractional abundance of helium and heavier elements $(Y; Z)$ in different layers of the Sun, the depth of base of the convection zone, and so on.

6 Inferences From the Global Helioseismic Data

Important inferences about the solar internal structure and dynamics have been obtained from the interpretation of the frequencies of solar global oscillation modes (Christensen-Dalsgaard 2002). The goal of the inverse problem is to derive unknown internal physical properties of the Sun, i.e., density, temperature, sound speed, from the observed frequencies. Two broad classes of inversion techniques are generally used: Regularised Least Squares (RLS) (cf., Antia and Basu 1994) and Optimally Localised Averages (OLA) (cf., Pijpers and Thompson 1992). Most inversion techniques use a reference solar model to calculate density and sound speed differences from the frequency differences. Hare and hound exercises provided the tests of inversions, where 'hare' constructs a solar model and calculates the frequencies of p-modes, which are perturbed by random errors and supplied to 'hounds'. The hounds calculate sound speed and density profiles using inversion techniques, and send these back to 'hare' for comparison with model.

6.1 Solar Radius, Density and Base of the Convection Zone

It is possible to measure the solar radius using the f mode in the $l-\nu$ diagram. It has no radial nodes, and has a simple dispersion relation, i.e., $\nu = \sqrt{gk_h}/2\pi$. The horizontal wavenumber is given by $k_h \approx ;\ \sqrt{l(l+1)}/r$, and f-mode frequencies are nearly independent of the internal structure of the Sun. Using $g = Gm_\odot/r^2$, one gets $\nu \propto r^{-3/2}$, which gives r as we can now determine frequencies with high precision. The problem with this method is that the optical depth which defines the surface must be defined precisely. Fixing the optical depth $\tau = 2/3$, the radius r is calculated to be 695.7 ± 0.026 Mm. The depth of the convection zone is another critical distance which is determined by the ratio of the mixing length and the pressure scale height. The p-mode ridges of high degree l are very sensitive to this ratio, which contribute to the determination of the base of convection zone. With improving solar models, the present estimate for its location is $r_{base} = (0.713 \pm 0.001)r_\odot$, i.e., (199700 ± 700) km from the surface of the Sun.

SOHO and GONG data analysis teams have obtained results for sound speed, density, and elemental abundances through inversions of measured frequency data (Fig. 9). Inversions of frequency differences does not show any change in the interior, except possibly in the HeII ionisation zone (Basu and Mandel 2004). There is also evidence that solar structure changes with the changes in solar activity.

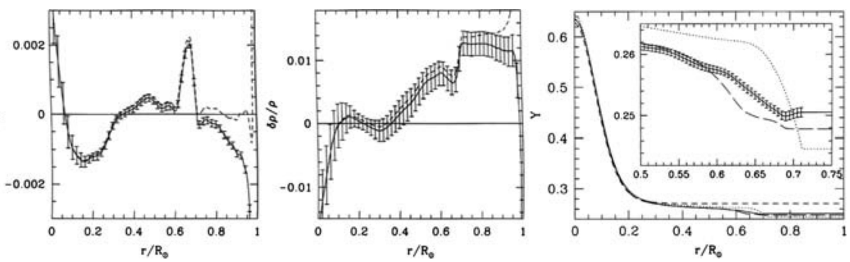


Fig. 9 Relative difference in (*left*) sound speed, (*middle*) density between the Sun and the standard solar model, and (*right*) fractional helium abundance; the inset – blow up close to convection zone base

6.2 Solar Internal Rotation and the "Tachocline"

An important achievement of helioseismology has been the inference of the large-scale rotation as a function of depth and latitude (Thompson et al. 2003). The inversion problem to determine the internal rotation rate of the Sun requires a very good frequency resolution, because the rotation rate is less than 0.5 μHz. If the Sun were to be spherically symmetric, there would be degeneracy in the azimuthal number m, but rotation breaks the degeneracy, which allows us to infer rotation rate $\Omega(r,\theta)$ as a function of depth. By measuring the splitting, we infer $\Omega(r)$ by means of inversion techniques:

$$D_{nlm} = \frac{\nu_{nlm} - \nu_{nl-m}}{2m} = \int_0^1 \int_0^1 dr\, d\cos\theta K_{nlm}(r,\theta)\Omega(r,\theta).$$

The internal rotational profiles shown in Fig. 10 (*left panel*) provides an important conclusion that differential rotation observed at the surface disappears below the base of the convection zone. The angular velocity is larger at the equator than at the poles throughout the convection zone, while the radiative interior rotates nearly uniformly. This important result also rules out a faster rotating solar core as envisaged by some proposals to mitigate the neutrino problem.

A shear layer, discovered between the radiative and convective zones (Fig. 10, *right panel*), is believed to be the seat of the solar dynamo (Gilman 2000). The changes in fluid flow velocities across the shear layer can stretch magnetic field lines of force and make them stronger. This change in flow velocity gives this layer

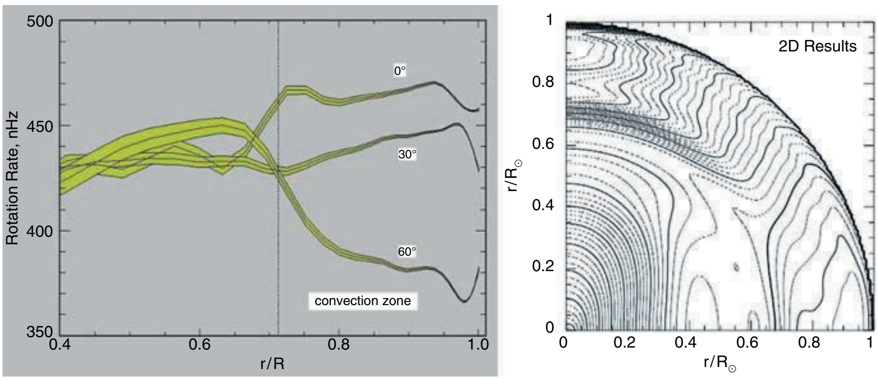


Fig. 10 (*left*) Solar internal rotation rate with radius at three latitudes, inferred from two months of MDI medium-l data (upto $l = 300$). (*right*) Contours of constant rotation rates at 5 nHz intervals by a 2-D inversion technique using GONG data. The base of convection zone is at r = 0.7 R, and the shear or interface layers are located just below the equatorial region and around the base of convection zone (*Tachocline*) (adopted from Antia et al. 1998)

its alternative name, the *tachocline*. There also appears to be sudden changes in chemical composition across this layer.

7 Local Helioseismology

Global helioseismology is complemented by local helioseismology which probes local perturbations in the interior close to the surface (cf., Duvall 1998). Local helioseismology provides a three-dimensional view of flows, magnetic structures, and their interactions in the solar interior. The main techniques of local helioseismology are (a) Fourier-Hankel spectral analysis (Braun et al. 1987), (b) ring-diagram analysis (Hill 1988), (c) time-distance helioseismology (Duvall et al. 1993), (d) helioseismic holography (Lindsey and Braun 1990), and (e) direct modeling (Woodard 2002).

Local helioseismology has been used to measure flows in the upper convection zone on a wide variety of scales, including differential rotation and meridional circulation, local flows around complexes of magnetic activity and sunspots, and convective flows (for a review, see Gizon and Birch 2005). Lindsey and Braun (2000) used helioseismic holography to provide images of active regions on the far-side of the Sun. This has importance in space weather predictions, as it allows about a week of warning before an active region emerges at the east limb. Daily far-side images obtained from MDI data are routinely available on the web at *http://soi.stanford.edu/data/farside.*

8 Summary

The evolution of the Sun and stars can be described by using a few basic equations. However, the models rely critically on the physics inputs such as opacities, nuclear reaction rates, abundances and so on. The structure of the present day Sun is now known very well from the recent developments in the tools of global and local helioseismology, and this has provided constraints on the solar models. Further, one of the major triumphs of the field was to show that the solar neutrino problem has a particle physics solution, and the problem did not lie with the solar models.

References

Antia, H.M., Basu, S.: *Astron. Astrophys. Suppl.* **107**, 421 (1994)
Antia, H.M., Basu, S., Chitre, S.M.: Mon. Not. R. Astron. Soc. **298**, 543 (1998)
Bahcall, J.N.: Space Sci. Rev. **24**, 227 (1979)
Basu, S., Mandel, A.: Astrophys. J. **116**, L155 (2004)
Braun, D.C., Duvall, T.L., Labonte, B.J.: Astrophys. J. Lett. **319**, L27 (1987)

Brun, A.S., Turck-Chieze, S., Zahn, J.P.: Astrophys. J. **525**, 1032 (1999)
Christensen-Dalsgaard, J.: Rev. Mod. Phys. **74**, 1073 (2002)
Christensen-Dalsgaard, J.: (2003): Lecture notes on Stellar Oscillations, 5th edn. (http://www.phys.au.dk/~jcd/oscilnotes/)
Deubner, F.L.: Astron. Astrophys. **44**, 371 (1975)
Duvall, T.L. Jr., Jefferies, S.M., Harvey, J.W., Pomerantz, M.A.: Nature **362**, 430 (1993)
Duvall, T.L. Jr.: Structure and dynamics of the interior of the Sun and Sun-like stars. In: Korzennik, S., Wilson, A. (eds.) Conf. Proc., SP-418, pp. 581. ESA (1998)
Gilman, P.A.: Solar Phys. **192**, 27 (2000)
Gizon, L., Birch, A.C.: (2005) http://solarphysics.livingreviews.org/Articles/lrsp-2005–6/
Hill, F.: Astrophys. J. **333**, 996 (1988)
Leibacher, J.W., Stein, R.F.: Astrophys. J. Lett. **7**, 191 (1971)
Leighton, R.B., Noyes, R.W., Simon, G.W.: Astrophys. J. **135**, 474 (1962)
Libbrecht, K.G., Woodard, M.F., Kaufman, J.M.: Astrophys. J. Suppl. **74**, 1129 (1990)
Lindsey, C., Braun, D.C.: Solar Phys. **126**, 101 (1990)
Lindsey, C., Braun, D.C.: Solar Phys. **192**, 261 (2000)
Pijpers, F.P., Thompson, M.J.: Astron. Astrophys. **262**, L33 (1992)
Thompson, M.J., Christensen-Dalsgaard, J., Miesch, M.S., Toomre, J.: Ann. Rev. Astron. Astrophys. **41**, 599 (2003)
Ulrich, R.K.: Astrophys. J. **162**, 993 (1970)
Vernazza, J.E., Avrett, E.H., Loeser, R.: Astrophys. J. Suppl. **45**, 635 (1981)
Woodard, M.F.: Astrophys. J. **565**, 634 (2002)

Dynamo Processes

Dibyendu Nandy

Abstract Magnetic fields play an important role in defining and modulating the space environment within the heliosphere. How these magnetic fields originate and evolve in stars such as the Sun or planets such as the Earth, is in itself a compelling question – the answer to which is of universal importance to astrophysics and space science. It is thought that magnetic fields are created through a dynamo process which involves complex, non-linear interactions within the magnetized plasma that is often encountered in astrophysical systems. In this chapter, after briefly discussing the importance of magnetic fields in the heliosphere, I provide a gentle introduction to concepts in magnetohydrodynamics, describe the mathematical foundation of mean-field dynamo theory, and explain the basic physical processes that constitute the dynamo mechanism. In doing this, I focus mainly on our star – the Sun – which is the primary source of heliospheric magnetic fields and their variability.

1 Introduction: Magnetic Fields in the Heliosphere

The heliosphere – or the sphere of influence of the Sun – encompasses the solar system including all its planets. The environment within this heliosphere is primarily governed by the radiative, particulate, and magnetic output of the Sun. Slow long-term variation in the Sun's radiative and magnetic output influences planetary atmospheres and climate. For example, a period of reduced solar activity between 1645 and 1715 AD, known as the Maunder minimum, coincided with a period of global cooling known as the *Little Ice Age*. On shorter timescales, explosive solar events such as flares and Coronal Mass Ejections (CMEs) hurl vast amounts of magnetized plasma into space; these adversely affect satellite operations, telecommunication facilities, oil-pipelines and air-traffic on polar routes on the one hand, but generates beautiful auroras on the other hand. Understanding solar variability and the changes that it induces in the heliosphere are therefore of particular importance

D. Nandy
Indian Institute of Science Education and Research, Kolkata, India
e-mail: dnandi@iiserkol.ac.in

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_3,

for planetary atmospheres, our technologies in space and on Earth and our life and society in general.

Much of the solar phenomena that defines and modulates the heliosphere can be traced back to the presence of magnetic fields in the Sun. These magnetic fields are ever-changing. On relatively short timescales, of seconds to days, sudden magnetic re-organization, mediated through magnetic reconnection leads to eruptive processes such as flares and CMEs. On moderate timescales, the Sun's magnetic output varies periodically, with an average period of 11 years. On relatively longer timescales of centuries to millennia, the magnetic (and the coupled radiative) output varies in a non-periodic irregular manner. It is widely believed that this solar magnetic variability, on timescales stretching from years to stellar evolutionary timescales (on the order of billions of years), has its origin in a dynamo mechanism in the Sun's interior. Other solar-like stars are also known to have variable magnetic activity – most likely due to the action of a similar dynamo mechanism in their interiors.

Some planets, such as the Earth, have magnetic fields. The Earth's magnetic field is much weaker compared to the Sun but it plays an important role in maintaining the magnetosphere – which shields us from harmful space radiation, plasma and energetic particles such as cosmic rays. Even though the Earth's magnetic field appears relatively non-varying, it is known to have reversed direction (flipping of the magnetic polarities) many times in the past. Moreover, if the Earth's magnetic fields were not being continually generated, it would possibly have decayed on a timescale of 20,000 years. This points out that a dynamo mechanism is working in the Earth's interior, creating and changing the geomagnetic field. It is quite likely that life as we know it, would not have been possible without the shielding and protective effects of the geomagnetic field which owes its existence to the geo-dynamo.

The earlier discussion illuminates the important role that dynamo processes play in the heliosphere, both in determining solar variability and in maintaining the geomagnetic field. Our focus here is on understanding the physical processes that underlie this dynamo mechanism. The rest of this chapter is organized as follows. In Sect. 2 I introduce the subject of magnetohydrodynamics and outline key physical concepts that are critical to understanding the dynamo mechanism. In Sect. 3 I lay the mathematical basis of dynamo theory through the dynamo equations and describe how they are derived. In Sect. 4 I focus on the solar dynamo as a representative example of dynamo action in the heliosphere. Finally, I end with some concluding remarks in Sect. 7.

2 Magnetohydrodynamics: Basic Theoretical Ideas

At about the same time that the magnetic nature of the sunspot cycle was being established through observations in the early twentieth century, efforts were on to construct a theory to describe the behavior of magnetic fields in a plasma. This theory, which came to be known as magnetohydrodynamics (MHD), was pioneered by

Hannes Alfvén – who was eventually awarded the Nobel Prize in 1970 for his contributions to MHD. Some basic concepts in MHD are crucial for gaining an insight into the dynamo mechanism and understanding many observable features of the solar magnetic cycle; it is therefore logical to start with these concepts. Assuming that readers have some background knowledge of electromagnetism, let us begin with the Maxwell's equations which describe electromagnetic fields and relate them to their sources. Two equations are of particular importance here:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{1}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}. \tag{2}$$

Here $\mathbf{E}$ and $\mathbf{B}$ denotes electric and magnetic fields, respectively, t denotes time, μ_0 the permeability of free space and $\mathbf{J}$ the current density. In (2) above I have neglected the displacement current term (assuming that we are dealing with a non-relativistic system). A further equation of relevance here is the one that expresses Ohm's law in a conductor (of conductivity σ) moving with a velocity v relative to the magnetic field:

$$\mathbf{J} = \sigma(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{3}$$

Substituting for $\mathbf{E}$ from (3) in (1), we have

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B} - \frac{\mathbf{J}}{\sigma}). \tag{4}$$

Now, substituting for $\mathbf{J}$ from (2) in (4) and using the relationship $\lambda = 1/\mu_0\sigma$ (where λ is the magnetic diffusivity), we get:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times B - \lambda \nabla \times \mathbf{B}). \tag{5}$$

If magnetic diffusivity is constant in space, (5) simplifies to:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \lambda \nabla^2 \mathbf{B}. \tag{6}$$

What we have ended up deriving here is known as the induction equation (expressed in different forms in (5) and (6)). This is one of the most fundamental equations in MHD and it describes the behavior of magnetic fields in a plasma system.

In deriving the induction equation, some assumptions are made related to the system in which this equation is applicable. These assumptions are:

- The plasma is a continuum – valid when the typical length scale of the system exceeds the ion gyro-radius.

- The plasma is a single fluid – valid if the system length scale is much larger than the Debye shielding length.
- Plasma is in thermodynamic equilibrium with distribution close to Maxwellian – valid if system timescale exceeds typical collision timescale and length scale exceeds mean free path.
- Relativistic effects are unimportant – holds if typical systems (flow) speeds are much less than the speed of light.
- Permeability, and in certain cases conductivity or diffusivity are isotropic.

Another important equation for the MHD system describes how the velocity field $\mathbf{v}$ evolves (the equation of motion):

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{1}{\rho}\nabla\left(p + \frac{B^2}{8\pi}\right) + \frac{(\mathbf{B} \cdot \nabla)\mathbf{B}}{4\pi\rho} + \mathbf{g} + \nu\, \nabla^2\mathbf{v}, \tag{7}$$

where ρ is the density, p the pressure, g the gravitational field and ν the kinematic viscosity of the plasma. Note that (7) is the familiar Navier-Stokes equation of hydrodynamics with two additional terms in it due to the inclusion of magnetic forces. The first additional term $B^2/8\pi$ is the pressure contribution from the magnetic field and the second term $(\mathbf{B} \cdot \nabla)\mathbf{B}/4\pi\rho$ signifies tension along the magnetic field lines.

The induction equation (5 or 6), the equation of motion (7), the energy equation for the system

$$\frac{\partial p}{\partial t} + (\mathbf{v} \cdot \nabla)p + \gamma p \nabla \cdot \mathbf{v} = Q, \tag{8}$$

where γ is the adiabatic coefficient and the term Q encompasses effects of heating, cooling and conduction (including ohmic heating), the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \tag{9}$$

and

$$\nabla \cdot \mathbf{B} = 0, \tag{10}$$

provides a complete description of the MHD system.

2.1 Magnetic Reynolds Number

The ratio of the first term to the second term on the R.H.S. of the induction equation (6) is

$$R_m = \frac{VB/L}{\lambda B/L^2} = \frac{VL}{\lambda} = \frac{\tau_\lambda}{\tau_V}. \tag{11}$$

The dimensionless number, R_m, is known as the magnetic Reynolds number and V, B, L are the typical values of velocity, magnetic field and length scale of the system under consideration (τ_λ and τ_V are typical diffusion and velocity timescales). One aspect which is immediately obvious is that R_m, which is proportional to L, will be larger by orders of magnitude in astrophysical systems relative to laboratory systems. Therefore magnetic fields in astrophysical systems behave very differently from their laboratory counterparts. For example, it is very difficult to build a laboratory model of a dynamo that is sustained through plasma motions acting to counter dissipation (through diffusion). However, significant progress have been made in the last decade in building laboratory dynamos and in the month of November, 1999, magnetic field self-excitation and a growing dynamo mode was observed for the very first time in any laboratory at the Riga Dynamo Facility in Latvia (Gailitis et al., 2000).

2.2 *Magnetic Diffusion*

If we consider the extreme limit, $R_m \ll 1$, then we can ignore the first term on the R.H.S. of (6). In that case we are left with:

$$\frac{\partial \mathbf{B}}{\partial t} = \lambda \nabla^2 \mathbf{B}. \tag{12}$$

This is the magnetic diffusion equation and expresses the fact that in the absence of any source or forcing term (involving plasma motions and associated energy), the initial magnetic field will simply diffuse away (with any inhomogeneity being smoothed out) with a characteristic timescale of $\tau_d = L^2/\lambda$ over a typical length scale of L.

2.3 *Concept of Flux Freezing*

If $R_m \gg 1$, as in most astrophysical systems, then the diffusion term in (6) is relatively unimportant compared to the preceding term. When one considers the *ideal MHD limit* (infinite conductivity and therefore very high R_m) and drops the diffusion term altogether, (6) reduces to:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}), \tag{13}$$

in which case it can be shown that (see e.g., Choudhuri 1998)

$$\frac{d}{dt} \int_S \mathbf{B} \cdot \mathbf{dS} = 0. \tag{14}$$

The above equation describes the following physical phenomenon: If the magnetic field vector in a plasma system satisfies (13), then the magnetic flux through any surface (say, S) constituting a part of the moving fluid will remain time-invariant when that fluid element is moving. In other words, magnetic fields remain frozen to the flow and moves along with it.

This is known as Alfvén's theorem of flux-freezing (Alfvén 1942) and can be appreciated by recognizing that if $R_m \gg 1$, as in most astrophysical systems, then magnetic fields attached to a moving fluid parcel do not have enough time to be dissipated or dispersed through diffusion (think $\tau_\lambda/\tau_V \gg 1$, where $\tau_v = L/v$ is the flow timescale). Therefore the frozen-in magnetic field moves with the fluid and in situations where the plasma-β parameter (ratio of gas to magnetic pressure) is high, allows the fluid to distort it. This is a fundamental concept in MHD and implies that any twisting or stretching motion in a magnetized plasma will result in the magnetic field being twisted or stretched.

2.4 Magnetic Buoyancy

Now let us look at the stability of horizontal magnetic flux tube in a system with gravity as in an astrophysical object. Suppose that the gas pressure outside the flux tube is $p_{external}$, the gas pressure inside is $p_{internal}$ and the flux tube has a magnetic field strength B and thus a magnetic pressure $B^2/8\pi$ associated with it. Then for pressure balance across the bounding surface of the flux tube:

$$p_{external} = p_{internal} + \frac{B^2}{8\pi}. \tag{15}$$

It follows from the above equation that

$$p_{internal} < p_{external}. \tag{16}$$

Since pressure $p = \rho R T$ (where R is the universal gas constant and T is the temperature) the above relation implies that for a magnetic flux tube

$$\rho_{internal} R T_{internal} < \rho_{external} R T_{external} \tag{17}$$

In situations where the plasma in the interior and exterior of the flux tube are in isothermal condition, this means the density inside the flux tube would be less than the density outside.

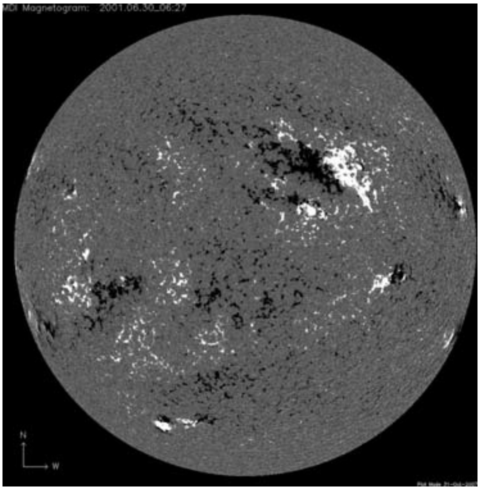


Fig. 1 A magnetogram image of the solar surface showing bipolar sunspot pairs. *White* denotes positive polarity while *black* denoted negative polarity sunspots. Note how the bipolar sunspot pairs have opposite polarity orientation in the two hemispheres with positive polarity sunspots appearing on the right hand side (i.e., on the leading side in the direction of rotation – *left* to *right*) above the equator and on the *left* (following side) below the equator. The small *black* and *white* patches and the *gray* region outside of sunspots, denote the much weaker, diffuse field on the surface. The formation of sunspots can be explained based on the concept of magnetic buoyancy

Therefore, any strong magnetic flux tube in the interior of astrophysical systems, would likely be less dense that its ambient surrounding. In the presence of a gravitation field directed downward, this would translate to the flux tube being buoyant (follows simply from Archimedes' principle). The flux tube would then tend to rise up against gravity. This concept of magnetic buoyancy, which was first worked out by Parker (1955a), explains the formation of bipolar sunspot pairs that are observed in the Sun (Fig. 1).

3 Formulating of the Kinematic Mean-Field Dynamo Problem

Plasma systems in astrophysical objects such as stars and planets often exist in a state of vigorous turbulence and host large scale flows. This is especially true for the those systems where convection, driven by temperature gradients, can support plasma motions. The kinetic energy of the plasma in such a region can presumably be used to feed a mechanism that converts it to magnetic energy. The subject that attempts to understand this mechanism and seeks theoretical explanations for the origin and evolution of astrophysical magnetic fields (or for that matter laboratory magnetic fields – as long as it is generated in a plasma), is known as dynamo theory.

To address the dynamo generation of magnetic fields one would have ideally liked to solve the complete set of MHD equations (5–10) simultaneously, to search for an answer to how interacting velocity and magnetic fields sustain each other and generate a global magnetic field. That, however, is an extremely formidable task that is highly computer intensive and relatively less suitable for a physical appreciation of individual processes at work in the system. We restrict ourselves to the more tractable problem of kinematic dynamo theory – where the velocity field is prescribed and one solves (5) or (6) for the magnetic field evolution.

The first significant effort in this subject turned out to be a negative result known as the *anti-dynamo theorem*. Cowling (1934) showed that an axisymmetric velocity field (the simplest possible flow field in a astrophysical body) cannot sustain an axisymmetric magnetic field (that is often observed). Arguably the most important step in astrophysical dynamo theory was taken by Parker (1955b) when he proposed that helical turbulent motions within rotating plasma systems (which are inherently non-axisymmetric) can sustain a magnetic field.

Parker's original dynamo formulation was focussed towards explaining the origin of the solar cycle and was largely intuitive. About a decade later, Steenbeck et al. (1966) put Parker's ideas on a firm mathematical footing – developing what is today known as the mean field theory of magnetohydrodynamics. Essentially they showed that a crucial term for dynamo action – the mean-electromotive-force (*mean e.m.f.* – essentially that process which converts the energy of plasma motions to magnetic energy) – is generated as a result of averaging over the interactions of the turbulent (or fluctuating) parts of the velocity and magnetic fields in a plasma.

In a turbulent system, the velocity and magnetic field can be written in terms of the average and fluctuating parts:

$$\mathbf{v} = \overline{\mathbf{v}} + \mathbf{v}', \quad \mathbf{B} = \overline{\mathbf{B}} + \mathbf{B}'. \tag{18}$$

Where the over-line indicates average and prime indicates fluctuating parts. Substituting (18) in the general form of the induction equation (5) and averaging the terms we get

$$\frac{\partial \overline{\mathbf{B}}}{\partial t} = \nabla \times (\overline{\mathbf{v}} \times \overline{\mathbf{B}}) + \nabla \times \mathcal{E} - \nabla \times (\lambda \nabla \times \overline{\mathbf{B}}), \tag{19}$$

considering that $\overline{\mathbf{v}'} = \overline{\mathbf{B}'} = 0$. Here $\mathcal{E}$ is the *mean e.m.f* term given by:

$$\mathcal{E} = \overline{\mathbf{v}' \times \mathbf{B}'}. \tag{20}$$

The *mean e.m.f* term can be perturbatively evaluated using a scheme known as the *first order smoothing approximation* (which requires a severe and possibly questionable truncation of the series expansion associated with the *mean e.m.f* term.). Assuming that the turbulence is isotropic, this leads to

$$\mathcal{E} = \alpha \overline{\mathbf{B}} - \beta \nabla \times \overline{\mathbf{B}}, \tag{21}$$

where

$$\alpha = -\frac{\tau}{3} \overline{\mathbf{v}' \cdot (\nabla \times \mathbf{v}')} \tag{22}$$

and

$$\beta = \frac{\tau}{3} \overline{\mathbf{v}' \cdot \mathbf{v}'}. \tag{23}$$

The term α, which is at the heart of the dynamo α-effect – represents the average helical motions present in the plasma, τ is the correlation time for turbulence and β signifies turbulent diffusivity. On substituting (21) for the *mean e.m.f.* in (19) and considering that the net magnetic diffusivity, η, is given by

$$\eta = \lambda + \beta, \tag{24}$$

we finally get

$$\frac{\partial \overline{\mathbf{B}}}{\partial t} = \nabla \times (\overline{\mathbf{v}} \times \overline{\mathbf{B}}) + \nabla \times (\alpha \overline{\mathbf{B}}) - \nabla \times (\eta \nabla \times \overline{\mathbf{B}}). \tag{25}$$

Usually the turbulent diffusivity β is much greater than λ and in all subsequent analysis from now on, I am going to denote the net magnetic diffusivity by the term η. Equation (25) is known as the dynamo equation and describes the dynamo generation of magnetic fields.

3.1 Evolution Equations for the Poloidal and Toroidal Fields

Most astrophysical objects of interest in dynamo theory, e.g., stars or planets are spherical systems, which by virtue of their rotation, also sustain mostly axisymmetric large-scale velocity and magnetic fields (although helical turbulent motions are not axisymmetric, their net effect, averaged over large-scales can result in the creation of an axisymmetric field). It is advantageous then to formulate the dynamo problem in terms of axisymmetric magnetic and velocity fields in spherical polar coordinates, expressed as

$$\mathbf{B} = B\,(r, \theta, t)\,\mathbf{e}_\phi + \nabla \times \left[A\,(r, \theta, t)\,\mathbf{e}_\phi\right], \tag{26}$$

$$\mathbf{v} = \mathbf{v}_p\,(r, \theta) + r \sin\theta \Omega\,(r, \theta)\,\mathbf{e}_\phi. \tag{27}$$

where B represents the toroidal component of the magnetic field (i.e., in the direction of rotation), $\mathbf{B}_p = \nabla \times (A\mathbf{e}_\phi)$ the poloidal component (in the $r - \theta$ meridional plane) and A the vector potential for the poloidal field; Ω is the (rotational) angular velocity (in the ϕ direction), and $\mathbf{v}_p = v_r\mathbf{e}_r + v_\theta\mathbf{e}_\theta$ is the meridional circulation (in the $r-\theta$ plane). We substitute (26) and (27) in the dynamo equation (25), to get the two coupled equations:

$$\frac{\partial A}{\partial t} + \frac{1}{s}(\mathbf{v}_p \cdot \nabla)(sA) = \eta \left(\nabla^2 - \frac{1}{s^2}\right) A + \alpha B, \tag{28}$$

$$\frac{\partial B}{\partial t} + \frac{1}{r}\left[\frac{\partial}{\partial r}(rv_r B) + \frac{\partial}{\partial \theta}(v_\theta B)\right] + \nabla\eta \times (\nabla \times B)$$
$$= \eta \left(\nabla^2 - \frac{1}{s^2}\right) B + s(\mathbf{B}_p \cdot \nabla)\Omega + \nabla \times (\alpha \mathbf{B}_p), \tag{29}$$

where $s = r\sin\theta$. Equations (28) and (29) are the evolution equations for the poloidal and toroidal components of the magnetic field, respectively, which have to be solved with the specification of boundary conditions appropriate for the system under consideration. Often an amplitude-limiting quenching factor, that signifies the non-linear feedback of the magnetic fields on the flows, is also included through a parametrization of the α-effect such as $\alpha \sim \alpha_0/[1 + (B/B_{quenching})^2]$, where α_0 is the amplitude of the α-effect and $B_{quenching}$ is the field strength at which non-linear feedbacks become important.

Let us now discuss specific terms of the toroidal and poloidal field evolution equations to gain an insight into what physical processes they represent. The first terms on the L.H.S. of (28–29) describe how the magnetic field evolves in time. This time evolution depends on the other (following) terms in the equations.

Advective Flux Transport by Meridional Circulation: The second terms on the L.H.S. of the above equations represent the transport of magnetic flux by meridional circulation, including stretching or compression effects due to expanding or contracting flow fields. In the high magnetic Reynolds number regime, where the magnetic flux in frozen in the plasma, these terms are quite important and magnetic field transport is dominated by plasma flows – characterizing a regime often referred to as the advection-dominated or circulation dominated regime.

Flux Transport Due to Diffusion Gradients: The third term on the L.H.S. of (29) involving the gradient of η represents a form of magnetic transport (also referred to as diamagnetic transport) in which the magnetic field is carried by a velocity field (the latter's magnitude being proportional to the gradient of diffusion) whose net effect is to carry magnetic fields from regions of strong diffusivity to regions of lower diffusivity. Obviously, in the absence of any diffusivity gradient (i.e., when η is constant in space), this term vanishes.

Decay and Dispersal by Magnetic Diffusion: The first terms on the R.H.S. of the above equations denote the effect of diffusivity on the magnetic field. Note that diffusion disperses the field in such a way as to smooth out inhomogeneity. Therefore diffusion does play a role in transporting magnetic fields in such systems. The relative importance of diffusive flux transport is measured in terms of the magnetic

Reynolds number again; when R_m is low (equivalently when diffusive timescales are shorter than the meridional flow timescale), diffusive dispersal dominates over advective flux transport. However, as opposed to flux transport by meridional circulation, diffusion also destroys the field (while being dispersed); this decay happens over a diffusion timescale, τ_d, which I have already discussed before.

Dynamo Source Terms: There are three important source terms in the above equations which represent the creation of magnetic field, without which the magnetic field of the system will simply decay away. These terms are described below.

The $\boldsymbol{\Omega}$-effect Toroidal Field Source: The second term on the R.H.S. of (29) represents the stretching of poloidal field by non-uniform rotation to create toroidal field. This process can be visualized in the following way. Imagine a pre-existing poloidal field (in the $r - \theta$ plane) that is frozen in a plasma system which is rotating at different speeds in the ϕ direction. Then that part of the poloidal field which is rotating faster will be stretched out of the $r - \theta$ plane and into the ϕ direction, therefore generating a new toroidal field component. This toroidal field generation process is commonly referred to as the Ω-effect because it owes its existence to the presence of non-uniform rotation.

Mean-Field $\boldsymbol{\alpha}$-effect Sources: The last terms on the R.H.S. of the above equations, involving α (derived from the *mean e.m.f.*), denotes the mean-field source terms. The α-effect term in (28) denotes the twisting of toroidal field out of the ϕ direction and into the $r - \theta$ plane creating a new poloidal component. While the α-effect term in (29) denotes the twisting of poloidal field out of the $r - \theta$ plane and into the ϕ direction therefore creating toroidal field. Obviously, the presence of helical motions is essential for these terms to exist.

3.2 Efficiency of the Dynamo Process: Dynamo Numbers

The efficiency of the dynamo mechanism can be estimated by the relative importance of the source and diffusive terms in the above equations. First lets consider the relative importance of the two source terms on the R.H.S. of (29) as compared to the diffusive term through a simple dimensional analysis. Considering that $\nabla \sim 1/L$ (where L is the typical length-scale of the system) the ratio of the Ω-effect source term (second term on the R.H.S. of (29)) to the diffusive term (first term on the R.H.S.) is

$$C_\Omega = \frac{(\Delta\Omega)L^2}{\eta} \tag{30}$$

and the ratio of the α-effect source term (third term on the R.H.S. of (29)) to the diffusive term (first term on the R.H.S.) is

$$C_\alpha = \frac{\alpha L}{\eta}. \tag{31}$$

Here, $\Delta\Omega$ represents the difference in angular velocity over the typical length-scale L, and α and η denote typical magnitudes of the α-effect and diffusivity, respectively, for the system.

When $C_\Omega \ll C_\alpha$ (weak differential rotation), the Ω-effect source term (due to differential rotation) in (29) can be neglected and the only remaining source terms in the poloidal and toroidal field evolution equations are both related to the α-effect and the resultant dynamo system is known as an α^2 dynamo. When C_Ω is comparable to C_α, we have to keep all the three source terms and the resultant dynamo system is known as an $\alpha^2\Omega$ dynamo. In the other extreme, when $C_\Omega \gg C_\alpha$ (strong differential rotation), we can neglect the α-effect source term in (29) (but not in (28), where it is still the only source term!) and the resultant system is known as an $\alpha\Omega$ dynamo. The Sun has strong differential rotation, and therefore, the solar dynamo is believed to be of the $\alpha\Omega$ type.

If we take the product of these two dimensionless numbers C_α and C_Ω, we get what is most commonly referred to as the dynamo number:

$$N_D = \frac{\alpha(\Delta\Omega)L^3}{\eta^2}. \tag{32}$$

The dynamo number compares the relative efficiency of the source terms to the diffusive terms in the dynamo equations. A high dynamo number ($N_D \gg 1$) for a specific system indicates that the dynamo process is efficient at creating magnetic fields in that system. On the other hand, a low dynamo number ($N_D < 1$) indicates that diffusion dominates over the source terms and therefore dynamo action is unlikely in that system.

4 Application to the Sun: The Solar Dynamo

The Sun's magnetic field and its variability is the primary determinant of the electromagnetic environment of the heliosphere. The behavior of solar magnetic fields are also well documented and the Sun's internal properties well constrained. It is prudent then to take the Sun's magnetic cycle as an example, and apply our knowledge of MHD dynamo theory to study it.

4.1 Solar Magnetic Fields

Sunspots have been observed for centuries, with telescopic observations initiated by Galileo Galilei in the early 1,600 s. The magnetic nature of sunspots have also been thoroughly explored since the discovery of strong magnetic fields, on the order of

1,000 Gauss (G), within sunspots by G.E. Hale in the early twentieth century (Hale 1908). Sunspots often appear in bipolar pairs (known as solar active regions) whose polarity orientation is opposite in the two hemispheres. These bipolar sunspot pairs have a systematic tilt that increases with latitude in both the hemispheres – a phenomenon termed as Joy's law (Hale et al. 1919). Many of these properties can be identified in Fig. 1. Years of sunspot observations have now firmly established that the sunspot cycle has an average period of 11 years, with large amplitude fluctuations. At the beginning of a new cycle, sunspots start appearing at about 40° latitude and with the progress of the cycle more and more sunspots appear at lower and lower latitudes. At the end of 11 years, a new cycle starts with bipolar sunspot pairs which have opposite magnetic polarity to that of the previous cycle, appearing again at about 40° latitudes. Thus, if one considers not only the number of sunspots but also their sign, the full cycle constitutes 22 years. Only after the development of the magnetograph by Babcock and Babcock (1955), it became possible to study the much weaker magnetic field that is distributed on the solar surface outside of sunspots. This field is of the order of 10 G and is concentrated in large unipolar patches (possibly within unresolved magnetic fibrils) that migrate poleward with the progress of the solar cycle (Bumba and Howard 1965; Howard and LaBonte 1981). This weak, diffuse field reverses its polarity (near the poles) at the time of solar maximum (the phase when the maximum number of sunspots are seen on the solar surface). It turns out that this polarity reversal also happens at intervals of 11 years and this shows that the cycle of the weak, diffuse field, is intimately connected to the sunspot cycle.

Figure 2 captures both the sunspot cycle as well as the cycle of the weak, diffuse field in one unified picture. The equatorward propagation of the sunspot formation belt as well as the poleward propagation of the weak, diffuse field is clearly discernable. Note also that the weak, diffuse field near the poles reverses its polarity at the time when the maximum number of sunspots are seen on the solar surface.

4.2 Large-Scale Solar Flows

Observations of small magnetic features on the solar surface show that they seem to be "carried" by a axisymmetric surface flow. An estimate of this surface flow velocity is possible by following the motions of these surface features and Doppler shift measurements. These estimates point out that there is a markedly poleward directed flow (from the equator) on the surface of the Sun. This largely axisymmetric flow in the meridian plane is referred to as the meridional circulation. With the development of helioseismic techniques such as ring-diagram analysis and time-distance helioseismology, it has became possible to study this flow more accurately. It is found that this poleward meridional flow pervades till at least, the outer 15% of the Sun (Giles et al. 1997). Various studies place the magnitude of this flow somewhere in between 10–25 $\mathrm{m\,s^{-1}}$. While, as yet, it has not been possible to observe a meridional flow in the solar interior, it follows from mass conservation that there should

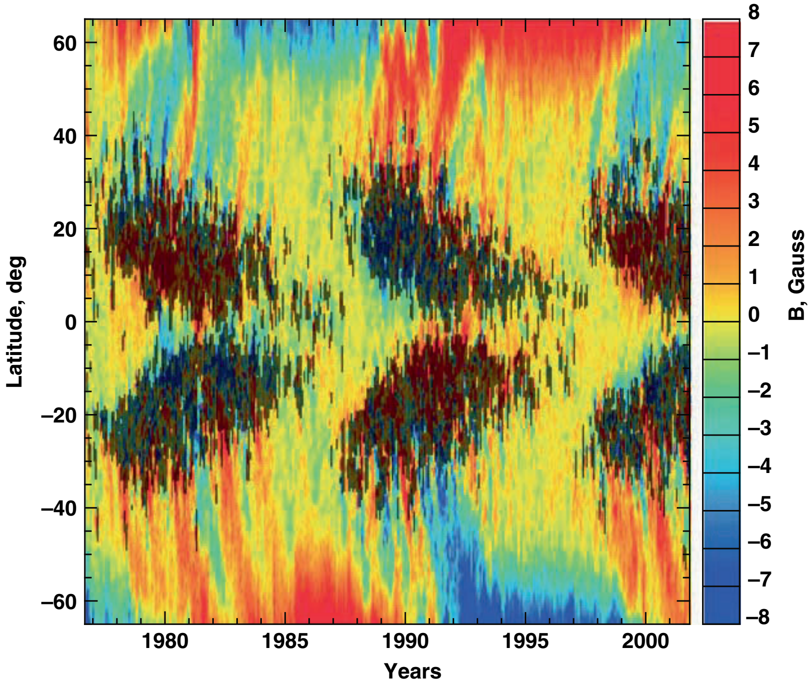


Fig. 2 Latitude versus time plot of sunspots as well as the weak, diffuse field. The time evolution of longitude-averaged weak, diffuse field is shown with colors depicting the magnitude of the field. Overlaid on this are vertical *black lines* which connect the latitudinal distribution of sunspots at a given time. This is often referred to as the sunspot butterfly diagram, as much an example of the wonderful symmetry sometimes found in nature, as of our fixation with something beautiful. The first plot, depicting the variation of sunspot eruption latitude with time in this manner, was constructed by Maunder (1904). *Image courtesy: Alexander G. Kosovichev*

be an equatorward counter-flow somewhere near the base of the solar convection zone (SCZ). Figure 3 shows a possible profile of this meridional circulation in the solar interior, although, at this writing, there is no consensus on the exact location of the equatorward counterflow.

It has been known for years that the Sun rotates differentially with the equator rotating faster than the poles. Helioseismology has now mapped the internal solar rotation as well (Charbonneau et al. 1999). The rotation varies with latitude in the upper 30% of the Sun – with the rotation rate increasing with decreasing latitude. Below the base of the SCZ this latitudinal variation in the rotation changes over into a radial variation over a thin region – termed as the tachocline. Beneath the tachocline, the radial variation falls off, with the radiative interior rotating more or less uniformly (see Fig. 3 for an analytic fit to the helioseismically determined solar internal rotation profile).

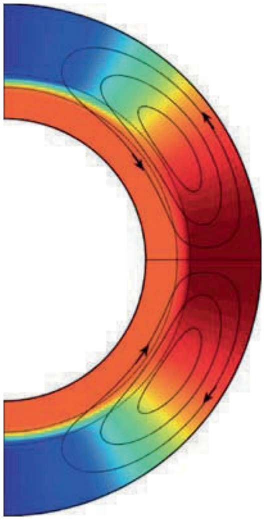


Fig. 3 Some representative streamlines of the solar meridional circulation are shown in *black* contours with *arrows* indicating the direction of flow. The background in color depicts an analytic fit to the solar internal rotation profile. *Deep red* indicates faster rotation, while *blue* indicates slower rotation. The two concentric *dashed lines* indicate the boundary of the tachocline – the region of strong radial shear in the rotation near the base of the solar convection zone. The domain shown is a meridional cut of the solar interior extending from $0.55R_{\odot}$ to $R_{\odot}$ (where $R_{\odot}$ is the solar radius) and from the north (*top*) to the south (*bottom*) pole

4.3 Understanding the Solar Cycle

Having described the observable properties of the solar cycle lets now summarize the current state of our understanding of the solar dynamo mechanism. For illustrative purposes, lets assume that we have a large-scale north-south oriented poloidal magnetic field in the solar interior. Due to the differential rotation, specifically the equator rotation faster than the poles, this poloidal field would be stretched out more near the equator (in the direction of rotation) creating a toroidal field. It is easy to see that this stretching would create opposite directed toroidal fields in the two hemispheres. With the discovery of the tachocline – the region of strong radial shear at the base of the SCZ, it is now fairly certain that the strong toroidal component of the solar magnetic fields are stored and amplified there. Given that the tachocline lies in a region of sub-adiabatic temperature gradient (also known as the overshoot layer), where magnetic buoyancy is suppressed, this storage of magnetic fields is possible there (Spiegel and Weiss 1980; van Ballegooijen 1982).

Very strong magnetic flux tubes, perhaps mediated via overshooting convection, can however come out of this "stable" layer. Once out in the convection zone, these magnetic flux tubes are subject to magnetic buoyancy (a concept I have outlined earlier in Sect. 2.4) and they rise up radially to erupt through the surface as bipolar sunspot pairs. Obviously in one of the spots the magnetic field is directed outward while in the other it is directed inward and thence their bipolar nature. Also, since the underlying toroidal field is directed in opposite directions in the two hemispheres, bipolar sunspot pairs have opposite orientations in the northern and southern hemispheres. Since the Sun is a rotating system, the buoyantly rising flux tubes are subject to the Coriolis force which results in the axis of the flux tubes acquiring a tilt which is manifested in the observed tilt of bipolar sunspot pairs (see Fig. 1).

It was initially thought that helical turbulent convection acts on the buoyantly rising toroidal flux tube to twist them into the $r - \theta$ plane re-generating the poloidal component of the magnetic field and this completing the dynamo chain (Parker 1955b). However, simulations of the buoyant rise of toroidal flux tubes show that the field strength of these flux tubes have to be on the order of 10^5 G at the base of the SCZ to match observable properties such as emergence latitudes (Choudhuri and Gilman 1987) and tilt angle distribution (D'Silva and Choudhuri 1993; Fan et al. 1993). This field strength is an order of magnitude stronger than the equipartition magnetic field strength (at which the energy in the fields and convection are in equipartition). This suggests that the mean-field dynamo α-effect powered by helical turbulent convection (discussed in Sect. 3) will get quenched and be ineffective in twisting such strong fields. Therefore alternative possibilities have to be explored. One of these alternative is an idea originally proposed by Babcock (1961) and Leighton (1969); the decay of tilted bipolar sunspot pairs, mediated via diffusion and meridional circulation would preferentially carry a net flux to the polar regions, reversing the older polar field and building up a new cycle poloidal field. This effect is actually observed, and has been extensively used in used in recent dynamo models to explain many observed features of the solar cycle (see e.g., Choudhuri et al. 1995; Durney 1995; Dikpati and Charbonneau 1999; Nandy and Choudhuri 2002; Chatterjee et al. 2004; Yeates et al. 2008).

These dynamo models based on the Babcock-Leighton idea solve the two coupled evolution equations for the poloidal and toroidal fields (28 and 29) in the $\alpha\Omega$ regime. However, both the mean-field source terms (the last terms in these equations) are discarded. Instead a phenomenological source-term that is concentrated in the surface layers, is used for the poloidal field equation. This term is parameterized to be mathematically similar to the αB term in the dynamo equations, but in spirit is very different as it does not involve averaging over small-scale helical turbulence, neither does it invoke the first-order smoothing approximation. A suitable computational algorithm for buoyancy is also prescribed that is used to mimic the buoyant eruption of toroidal flux tubes (Nandy and Choudhuri 2001). Meridional circulation plays an important role in these class of models in transporting the poloidal field first poleward and then downward into the tachocline where the strong toroidal fields are stored and amplified; the counterflow in the circulation is also responsible for the equatorward propagation of the sunspot formation belt. The meridional flow speed is believed to control the sunspot cycle period in these class of models (Charbonneau and Dikpati 2000; Nandy 2004; Hathaway et al. 2003; Yeates et al. 2008). In Fig. 4, I present a simulated butterfly diagram from one such dynamo model powered by the Babcock-Leighton idea for poloidal field generation. We see

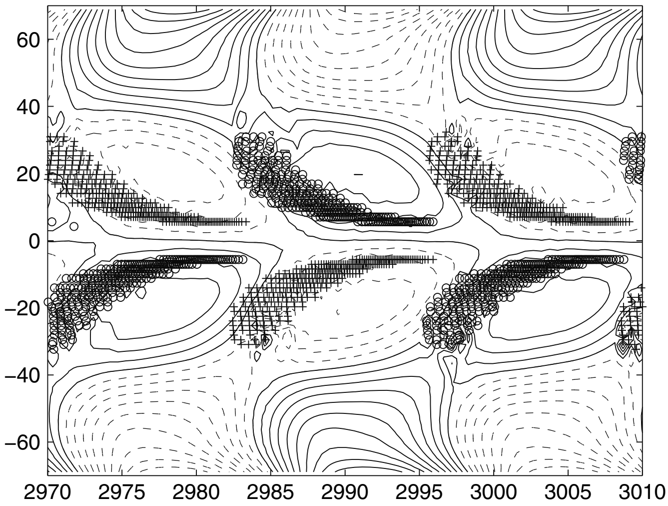


Fig. 4 Theoretically simulated solar butterfly diagram (y-axis denotes solar latitude and x-axis denotes time in years) from a recent dynamo model based on the Babcock-Leighton idea for poloidal field generation (from Chatterjee et al. 2004). The background shows contours of diffuse radial field. The *dashed* contours are for negative radial field, whereas the *solid* contours are for positive radial field. Sunspot eruption latitudes are denoted by symbols "o" and "+", indicating negative and positive toroidal field respectively. Note that many features of the observed solar butterfly diagram (Fig. 2) are well-reproduced by this model

that the large-scale features of the solar cycle are fairly well-reproduced in by this class of solar dynamo models.

5 Concluding Remarks

My endeavor here has been to give a flavor of ideas and concepts in dynamo theory. Given that this is directed towards students, who are possibly being exposed for the first time to dynamo theory, I have concentrated on highlighting important concepts in MHD. I have also laid down a pedagogical account of the development of ideas in this subject, stressing particulary on its mathematical formulation. Finally, I have focussed on the solar cycle as a representative case study to illustrate how the theoretical ideas come together to explain the solar dynamo mechanism. It has to be stressed however that this is just a limited account of the subject. Those desirous of a serious pursuit of dynamo theory are encouraged to consult other related works that give a more detailed account of this subject. Amongst these are the books "Magnetic Field Generation in Electrically Conducting Fluids" (Moffatt 1978), "Cosmical Magnetic Fields" (Parker 1979), "The Physics of Fluids and Plasmas: An Introduction for Astrophysicists" (Choudhuri 1998) and the recent reviews on the solar dynamo mechanism by Ossendrijver (2003) and Charbonneau (2005). This effort will be well served if it is only the beginning, and inspires a more careful and sustained effort at studying dynamo theory and perhaps researching some of its outstanding issues; there are many!

Acknowledgements I would like to thank the faculty at the Indian Institute of Astrophysics and at the Kodaikanal Solar Observatory for their hospitality during my visit there in the winter of 2007 to lecture at the IHY school. My research has been financially supported by the NASA Living With a Star program and the Ramanujan Fellowship of the Government of India.

References

Alfvén, H.: Ark. Mat. Astr. Fysik. **29B**, 2 (1942)
Babcock, H.W.: Astrophys. J. **133**, 572 (1961)
Babcock, H.W., Babcock, H.D.: Astrophys. J. **121**, 349 (1955)
Bumba, V., Howard, R.: Astrophys. J. **141**, 1502 (1965)
Charbonneau, P.: Living Rev. Sol. Phys. **2**, 2 (2005) http://solarphysics.livingreviews.org/Articles/lrsp-2005-2/
Charbonneau, P., et al.: Astrophys. J. **527**, 445 (1999)
Charbonneau, P., Dikpati, M.: Astrophys. J. **543**, 1027 (2000)
Choudhuri, A.R.: The Physics of Fluids and Plasmas: An Introduction for Astrophysicists. Cambridge University Press, Cambridge (1998)
Choudhuri, A.R., Gilman, P.A.: Astrophys. J. **316**, 788 (1987)
Choudhuri, A.R., Schüssler, M., Dikpati, M.: Astron. Astrophys. **303**, L29 (1995)
Chatterjee, P., Nandy, D., Choudhuri, A.R.: Astron. Astrophys. **427**, 1019 (2004)
Cowling, T.G.: Mon. Not. Royal Astron. Soc. **94**, 39 (1934)
Dikpati, M., Charbonneau, P.: Astrophys. J. **518**, 508 (1999)
D'Silva, S., Choudhuri, A.R.: Astron. Astrophys. **272**, 621 (1993)
Durney, B.R.: Solar Phys. **160**, 213 (1995)
Fan, Y., Fisher, G.H., DeLuca, E.E.: Astrophys. J. **405**, 390 (1993)
Gailitis, A., et al.: Phys. Rev. Lett. **84**, 4365 (2000)
Giles, P.M., et al.: Nature **390**, 52 (1997)
Hale, G.E.: Astrophys. J. **28**, 315 (1908)
Hale, G.E., Ellerman, F., Nicholson, S.B., Joy, A.H.: Astrophys. J. **49**, 153 (1919)
Hathaway, D.H., Nandy, D., Wilson, R.M., Reichmann, E.J.: Astrophys. J. **589**, 665 (2003)
Howard, R., LaBonte, B.J.: Solar Phys. **74**, 131 (1981)
Leighton, R.B.: Astrophys. J. **156**, 1 (1969)
Maunder, E.W.: Mon. Not. Royal Astron. Soc. **64**, 747 (1904)
Moffatt, H.K.: Magnetic Field Generation in Electrically Conducting Fluids. Cambridge University Press, Cambridge (1978)
Nandy, D.: In: Danesy, D. (ed.) Proceedings of SOHO 14/GONG 2004, Helio- and Asteroseismology: Towards a Golden Future, (ESA SP-599), 241 (2004)
Nandy, D., Choudhuri, A.R.: Astrophys. J. **551**, 576 (2001)
Nandy, D., Choudhuri, A.R.: Science **296**, 1671 (2002)
Ossendrijver, M.: Astron. Astrophys. Rev. **11**, 287 (2003)
Parker, E.N.: Astrophys. J. **121**, 491 (1955a)
Parker, E.N.: Astrophys. J. **122**, 293 (1955b)
Parker, E.N.: Cosmical Mangetic Fields. Clarendon Press, Oxford (1979)
Spiegel, E.A., Weiss, N.O.: Nature **287**, 616 (1980)
Steenbeck, M., Krause, F., Rädler, K.-H.: Z. Naturforsch. **21a**, 1285 (1966)
van Ballegooijen, A.A.: Astron. Astrophys. **113**, 99 (1982)
Yeates, A.R., Nandy, D., Mackay, D.H.: Astrophys. J. **673**, 544 (2008)

Large-Scale Solar Eruptions

Natchimuthukonar Gopalswamy

Abstract This chapter provides an over view of coronal mass ejections (CMEs) and the associated flares including statistical properties, associated phenomena (solar energetic particles, interplanetary shocks, geomagnetic storms), and their heliospheric consequences.

1 Introduction

A solar eruption can be defined as a transient ejection of material from the solar atmosphere. Transient eruptions occur from closed field regions of all sizes, starting from bright points to large active regions. In this chapter, we consider only those eruptions, which have observable manifestations in white light images obtained by coronagraphs. Coronagraphs are instruments that have an occulting disk that block the bright solar disk so the faint corona can be observed by means of Thomson scattered photospheric light. In coronagraph images, the large-scale eruptions are observed as moving bright features. Usually moving features with an angular width of a few degrees or more are known as coronal mass ejections (CMEs). Solar flares associated with CMEs are known as eruptive flares in contrast to compact flares, which are not associated with any mass motion. Flares represent electromagnetic emission observed prominently in soft X-rays, EUV, and H-alpha. Nonthermal emission is also observed over the entire radio spectrum (from kilometric to millimetric wavelengths), hard X-rays, and gamma rays. Flares cause enhanced ionization in the terrestrial ionosphere causing sudden ionospheric disturbances (SIDs) that seriously affect radio communication and navigation (Davies 1990). Energetic particles accelerated in flares and CME-driven shocks can be potentially harmful to space technology and humans in space. Energetic CMEs propagate far into the heliosphere causing observable effects along their path, sometimes all the way to the

N. Gopalswamy
NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
e-mail: nat.gopalswamy@nasa.gov

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_4,

heliospheric termination shock. When CMEs arrive at Earth, they can couple to the magnetosphere causing intense geomagnetic storms, which have implications to all layers of Earth's atmosphere and the ground. This chapter provides an overview of CMEs and flares, the two main aspects of large-scale solar eruptions. Several review articles exist on these topics (see the articles in the recent Geophysical Monographs Gopalswamy et al. 2006a; Song et al. 2001). A more detailed description of the eruption events can be found in (Gopalswamy 2007).

2 A Large-Scale Eruption Illustrated

CMEs have different manifestations depending on the wavelength of observation and the heliocentric distance. CMEs originate from regions on the Sun where the magnetic field lines connect opposite polarities. These regions are the active regions and filament regions, where the magnetic field strength is elevated with respect to the quiet Sun. Most active regions contain sunspots and also filaments. Quiescent filament regions have no sunspots. Figure 1 has a magnetogram and an H-alpha image showing active regions as compact magnetic regions and filament regions as diffuse magnetic regions. Both types of magnetic regions are potential sources of CMEs. Filaments consist of cool material (~8,000 K) compared to the hot (~2 MK) corona. Filaments are suspended in the corona and are roughly aligned with the polarity inversion line of the magnetic regions, while their ends are rooted in opposite

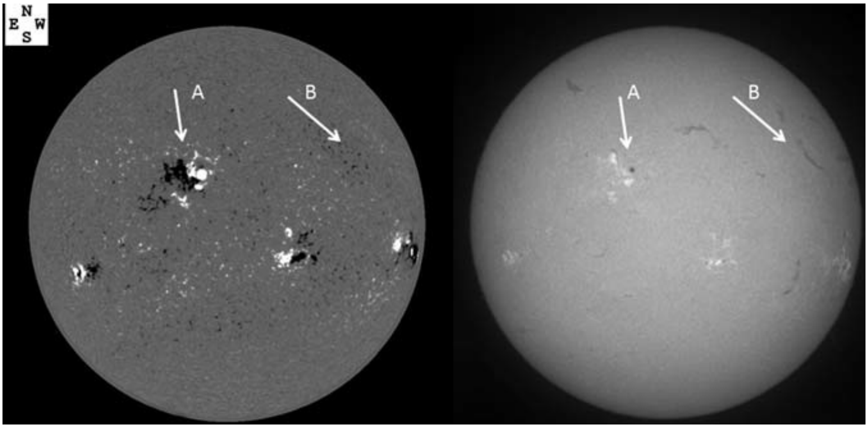


Fig. 1 A line of sight magnetogram from the Big Bear Solar Observatory taken at 15:21:03 UT (*left*) and an H-alpha image from the Kanzelhoehe observatory taken at 08:14:58 (*right*) both on 2005 May 13. In the magnetogram, white and black represent positive (north) and negative (south) magnetic polarities, respectively. The elongated dark features in the H-alpha image are the filaments. One of the active regions (AR 10759) is marked "A". One of the filaments is marked "B". Note that the active region magnetic fields are intense and the filaments are generally thin. In the filament regions, the magnetic field is also enhanced compared to the quiet Sun, bipolar, and more diffuse

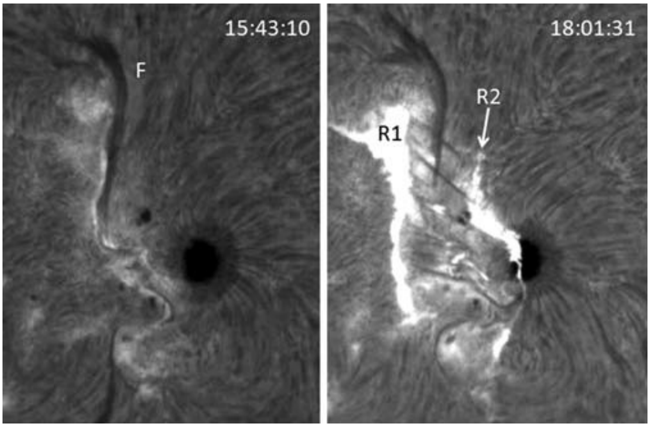


Fig. 2 Two H-alpha images of active region AR 10759 observed on 2005 May 13 at the Big Bear Solar Observatory (BBSO). Each image is of 331 × 351 pixels (pixel size = 0.6 arc sec). (*left*) before eruption and (*right*) after eruption. The elongated dark feature is the filament. Two sunspots of positive polarity can be seen to the west of the filament. In the right side image, part of the filament has disappeared due to eruption and two ribbons (R1, R2) on either side of the filament have appeared. The dark features connecting R1 and R2 are the post-flare loops at a temperature of ∼10,000 K. Courtesy: V. Yurchyshyn

polarity patches. In coronal images, hot loop structures can be seen overlying the filaments. During an eruption, the filament is ejected from the Sun (wholly or partially) including the surrounding coronal material. Thus a CME typically contains multi-thermal plasma. During the eruption process, the filament gets heated and hence may not remain cold for long in the interplanetary medium.

Figure 2 shows an eruption imaged in H-alpha wavelengths from the region A in Fig. 1. In the pre-eruption image, one can see the vertical dark filament (F) with two sunspots on the western side. After the eruption, one can see two vertical bright ribbons (R1, R2), characteristic of eruptive flares. The ribbons indicate energy input to the chromosphere, thought to be due to electrons from the energy release site just beneath the erupting filament in the corona. The ribbons represent the eruptive flare and the total area of the ribbons is used as a measure of the flare size. The ribbons mark the feet of an arcade of loops thought to be formed due to reconnection. Some of these loops can be seen as dark features connecting R1 and R2. These "post-eruption" loops are roughly perpendicular to neutral line, while the flare ribbons are roughly parallel to the neutral line. The whole arcade can be seen in coronal images obtained in extreme-ultraviolet, X-ray, and microwave wavelengths. Figure 3a shows the photospheric magnetogram and the coronal (potential) magnetic field lines extrapolated from the photospheric field in the active region that produced the eruption. In Fig. 3c one can see the entire arcade in a coronal image at 171 Å obtained by the TRACE satellite. Note the close resemblance between the potential field lines and the post-eruption arcade, in contrast to the pre-eruption coronal structures (Fig. 3b), which are almost aligned along the neutral line. Such

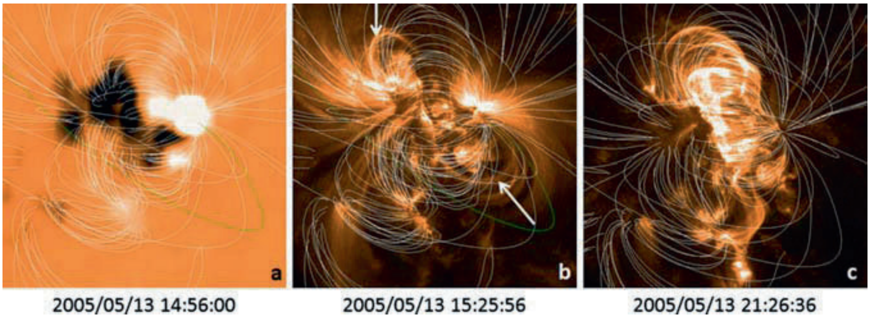


Fig. 3 Magnetic structure in AR 10759 revealed by (**a**) low-resolution photospheric magnetogram at 14:56:00 UT before the eruption (*white* is positive and *dark* is negative) with computed potential field lines (*thin white lines* are closed and the *green line* is open); (**b**) TRACE 171 Å image at 15:25:56 UT showing non-potential coronal structures (two highly sheared coronal loops are pointed by *arrows* with foot-points on either side of the neutral line; (**c**) TRACE 171 Å image of the post-eruption arcade at 21:26:36 UT. Computed potential field lines are also shown superposed. Courtesy: K. Schrijver and M. DeRosa

non-potential structures represent storage of magnetic energy in the corona, which is released during the eruption in the form of thermal energy (flare heating) and kinetic energy (CME). Figure 4 shows a large-scale disturbance in the corona surrounding the active region revealed by EUV images obtained by the Solar and Heliospheric Observatory (SOHO). Such disturbances are coronal waves (MHD fast mode waves or shocks depending on the speed) surrounding the erupting CME. The EUV disturbance has a size of $\sim$0.4 $R_\odot$ at 16:37 UT. The disturbance spreads over the entire solar disk by 17:07 UT as can be seen in Fig. 5. Detection of such large-scale EUV disturbances has become a standard way of identifying the solar source of CMEs. The white light CME in Fig. 5 taken about 15 min after the EUV image clearly shows the close connection between the EUV disturbance and the CME in white light. The soft X-ray flare in Fig. 5 also shows the close connection between flares and CMEs. The current paradigm for CMEs is that the flare reconnection process creates a flux rope (or builds upon a pre-existing one), whose outward motion causes the disturbance seen in white light as the CME.

The CME first appeared at a heliocentric distance of $\sim$4.6 $R_\odot$ just above the occulting disk (which has a radius of 3.7 $R_\odot$) at 17:12 UT (see Fig. 6). The disturbance can be seen simultaneously at the north and south edges of the occulting disk, suggesting that it was elongated in the north-south direction as is clear from the white light image taken at 17:42 UT (Fig. 6). Between 17:12 and 17:42 UT, the leading edge of the white light feature moved a distance of $\sim$4.4 $R_\odot$, indicating a speed of $\sim$8.8 $R_\odot$/h or $\sim$1,700 km s^{-1}. Since the eruption occurred on the frontside of the Sun that faces Earth, the CME should reach Earth (which is $\sim$214 $R_\odot$ away from the Sun) in a day or so. In fact, the leading part of the CME-driven shock arrived at Earth (see Fig. 7a) on 2005 May 15 at 02:19 UT, which is a little more than a day because CMEs undergo deceleration due to interaction with the solar wind. The speed of the CME heading towards Earth cannot be measured accurately because of the occulting

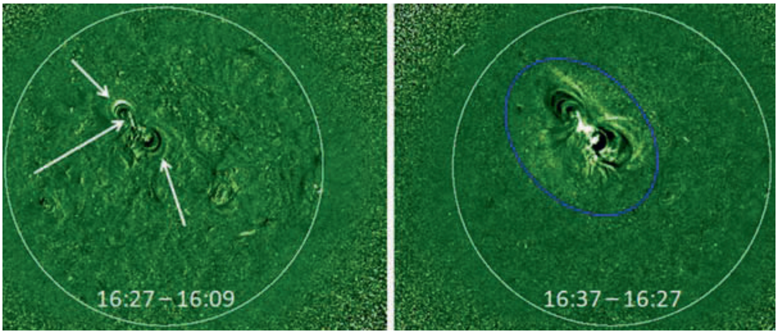


Fig. 4 EUV images at 195 Å obtained by the Solar and Heliospheric Observatory (SOHO) showing changes in the sheared loops (*left* at 16:27 UT) and large-scale diffuse disturbance surrounding the active region (*right* at 16:37 UT). These are running difference images (previous image subtracted from the current image to reveal changes taking place during the interval between the two images). In the left image, the three *arrows* point to significant changes: the central brightening near the neutral line (corresponding to the beginning of the H-alpha ribbons) and the expansion of the sheared loops noted in Fig. 3b. In the right image, the post-eruption arcade near the neutral line has expanded; the large-scale disturbance is shown encircled

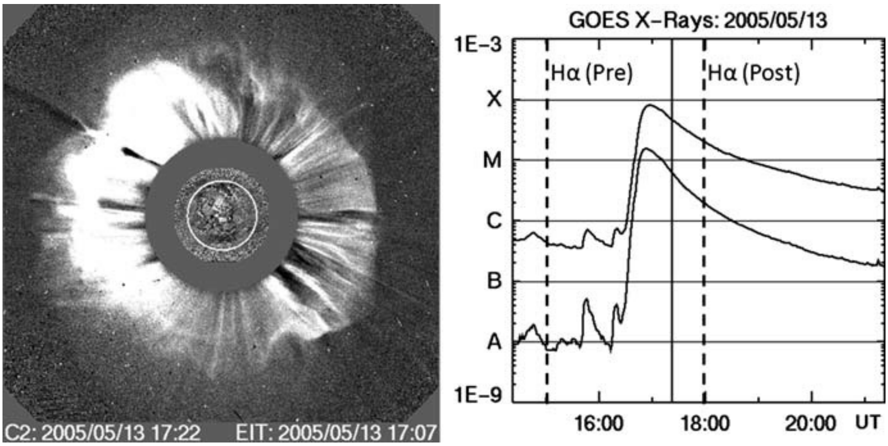


Fig. 5 (*left*) White light CME detected by SOHO in its C2 telescope of the Large Angle and Spectrometric Coronagraph (LASCO) on 2005 May 13 at 17:22 UT. Superposed on this image is a EUV difference image from SOHO/EIT at 17:07 UT. Note that the EUV disturbance has spread over the entire solar disk compared to the image at 16:37 shown in Fig. 4. (*right*) Soft X-ray light curve in two energy channels (1–8 Å upper curve; 0.4–5 Å lower curve). The soft X-ray detector has no spatial resolution, but most of the flare emission comes only from AR 10759. The time of the LASCO image is marked by the solid vertical line. The two dashed lines correspond to the pre and post-eruption H-alpha pictures shown in Fig. 2

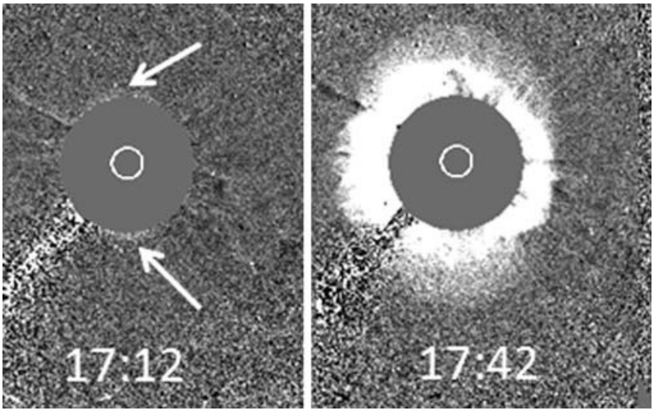


Fig. 6 White light CME detected by SOHO in its C3 telescope of LASCO. These are difference images with previous frame subtracted. In the left frame the CME appears above the occulting disk (pointed by *arrows*). In the right frame, the CME has expanded significantly. The occulting disk is employed by the coronagraph to block the photospheric light so weak coronal features can be imaged. The white circle at the center of the images represents the optical disk of the Sun

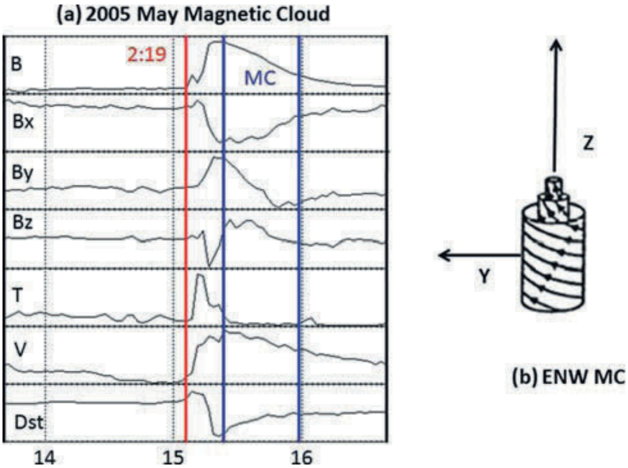


Fig. 7 (**a**) Plots of the magnetic field magnitude (B), its three components (Bx, By, Bz) in Geocentric Solar Magnetospheric (GSM) coordinates, temperature, solar wind speed, and the Dst index (all in arbitrary units). The origin of the X-axis roughly corresponds to the beginning of activities in AR 10759. The vertical line at 02:19 UT on 2005 May 15 is the time of arrival of the shock at the ACE spacecraft at the L1 point. The vertical line at 11:00 UT the same day is the arrival time of the magnetic cloud that was driving the CME. The Dst index shows an intense geomagnetic storm, with an intensity of −263 nT. All the physical quantities change drastically at the arrival of the shock and magnetic cloud. (**b**) Flux-rope structure of the ICME as derived from the By, and Bz components. The Bz component in the MC always points to the north, while the By component rotates smoothly from the east (E) to west (W). Thus the MC is classified as a ENW cloud

disk, which introduces additional uncertainty in the CME speed. Unlike the 2-dimensional images of CMEs near the Sun, the in-situ measurements are only along a single trajectory through the CME as it moves past the observing spacecraft. The measurements shown in Fig. 7a indicate that the disturbance lasted for a period of $\sim$1 day. The interplanetary CME (ICME) had a flux rope structure with its axis in the Z-direction, but it was pointing to the north (see Fig. 7b). The shock sheath ahead of the flux rope had an intense south-pointing field, which merged with Earth's magnetic field and caused a very intense geomagnetic storm (measured by the Dst (Disturbance Storm time) index, which reached -263 nT). A CME such as the one on 2005 May 13, which results in a Dst index ≤ -50 nT is said to be geoeffective.

The shock at 1 AU is directly detected by spacecraft in the solar wind. A shock near the Sun has to be inferred from remote observations such as at radio wavelengths and from energetic particles. When the CME speed near the Sun exceeds the coronal Alfven speed, it can drive a shock. The shock accelerates electrons and ions (solar energetic particles or SEPs for short). The accelerated electrons produce radio emission by interacting with the ambient corona. Accelerated ions and electrons also propagate to the detectors in near-Earth space within tens of minutes. Shocks continue to accelerate SEPs as they propagate into the IP medium, so they are continuously detected until the shock reaches the observing spacecraft, when a sudden increase in SEP intensity is observed. The intensity increase is referred to as an energetic storm particle (ESP) event because a geomagnetic storm usually follows due to the sheath and/or ICME behind the shock. In the 2005 May 13 event, a metric type II burst was observed at 16:38 UT, coincident with the large-scale EUV disturbance. A more intense type II was observed by the Wind/WAVES experiment in the IP medium starting around 17:00 UT.

Figure 8 shows the SEP event as detected by the Geostationary Operational Environment Satellite (GOES) and the radio dynamic spectrum (frequency vs. time plot of the radio intensity) obtained by the Wind/WAVES experiment. Both the SEPs and the type II burst are thought to be due to the CME-driven shock. From Fig. 8, we see that SEPs are observed more than an hour after the CME onset near the surface. However, when we account for the propagation time of 10 MeV protons, the SEP release time is $\sim$17:18 UT, when the CME was at a heliocentric distance of $\sim 7\,R_\odot$. Note that SEPs are observed when the shock is near the Sun until its arrival at the spacecraft at 02:19 UT on May 15, marked by the ESP event. During the ESP event, the proton intensity in the 10 MeV channel jumps from 100 to 3,000 pfu. The SEP intensity drops significantly in the sheath and Magnetic Cloud (MC) regions.

The type II burst occurs at the local plasma frequency (fp) or it harmonic (2fp). Assuming the radio emission to be at the harmonic, one can estimate the plasma density (n) at the heliocentric distance of the CME since $\mathrm{fp} = 9 \times 10^{-3}\mathrm{n}^{1/2}$, where fp is in MHz and n is in cm^{-3}. For example, the type II burst at 17:42 UT occurs at 0.6 MHz, so $\mathrm{fp} = 0.3$ MHz and $\mathrm{n} \sim 1.1 \times 10^3\,\mathrm{cm}^{-3}$. At this time, the CME is at a height of $\sim 9.0\,R_\odot$. The actual height of the CME can be estimated to be $\sim 11\,R_\odot$ assuming that the CME is cone-shaped. As the CME propagates away from the Sun, the shock ahead of it passes through layers of decreasing plasma density and hence producing radio emission at decreasing frequency. Thus the type II burst can be

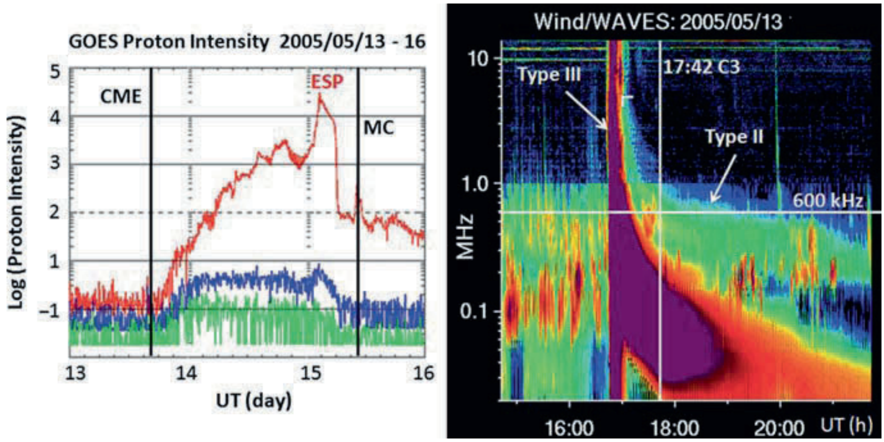


Fig. 8 (*left*) Intensity plots of SEPs (protons) during the 2005 May 13 CME in three energy channels >10 MeV (*red*), >50 MeV (*blue*), and >100 MeV (*green*) as detected by the GOES satellite. The SEP intensity is in particle flux units (pfu); 1 pfu $=$ 1 proton $(\mathrm{cm}^2\,\mathrm{s}\,\mathrm{sr})^{-1}$. The sharp spike in the SEP intensity (>10 MeV channel) is the ESP event. Approximate onset time of the CME at the Sun (16:37 UT on May 13) and near Earth (MC arrival at 11 UT on May 15) are marked by the vertical black lines. (*right*) Dynamic spectrum obtained by the Radio and Plasma Wave (WAVES) experiment on board the Wind spacecraft. The type III burst is indicative of electron beams from the eruption site. The type II burst is due to nonthermal electrons accelerated at the CME-driven shock. The type II radio emission occurs at ~600 KHz when the CME was last seen in the coronagraph image at 17:42 UT (indicated by the vertical line)

used to track the CME-driven shock beyond the coronagraphic field of view. For the 2005 May 13 CME, the type II radio emission was observed all the way to the Wind spacecraft (where the local plasma frequency was ~30 kHz).

The 2005 May 13 event discussed above is a well-observed CME with all the associated phenomena such as a flare (H-alpha and soft X-ray), filament eruption, EUV wave, white-light CME, type II radio burst (metric and IP), SEP, ICME with shock, and geomagnetic storm. The white-light CME is also a halo CME in that the CME appears to surround the occulting disk in sky-plane projection (Howard et al. 1982). CMEs like this make significant impact on the heliosphere and hence are important in deciding space weather. Space weather refers to the conditions in the space environment that is hazardous to space or ground based technological systems or to human health or life. SEPs, geomagnetic storms, and ionospheric storms are some of the space weather effects directly linked to CMEs.

3 CME Properties

CMEs have been studied extensively using data from several spaceborne coronagraphs since the early 1970s and the ground-based Mauna Loa K-coronameter (see Gopalswamy 2004; Kahler 2006 for recent reviews.) The statistical properties of

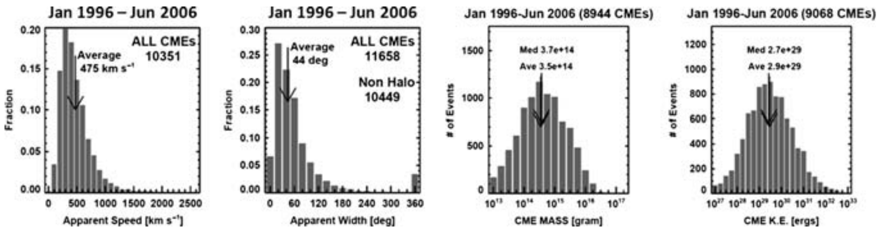


Fig. 9 Speed, width, mass and kinetic energy distributions of CMEs observed over most of solar cycle 23 (1996–2006)

CMEs based on SOHO observations can be summarized as follows: (1) the CME speed varies over two orders of magnitude from $\sim$20 to more than 3,000 km s^{-1}, with an average value of 475 km s^{-1}. (2) The CME angular width ranges from <5° to 360° (halo CMEs). The average width of CMEs with width <120° is $\sim$44°. Almost 90% of the CMEs have width <120°, and only $\sim$3.5% have an apparent width of 360°. About 11% of CMEs have width ≥120°. The speed and width distributions of CMEs are shown in Fig. 9. (3) CMEs with above-average speeds decelerate due to coronal drag, while the lower speed ones accelerate. CMEs with speeds close to the average speed do not have observable acceleration. (4) CME mass ranges from a few times 10^{12} g to more than 10^{16} g with an average value of 3.4×10^{14} g (Gopalswamy 2004). The corresponding kinetic energies range from $\sim 10^{27}$ erg to more than 10^{33} erg, with an average value of 2.9×10^{29} erg (see Fig. 9). (5) Within the field of view of the Large Angle and Spectrometric Coronagraph (LASCO), CMEs occur with an average rate of <0.5 per day (solar minimum) to >6 per day (solar maximum). High-latitude (>60°) CMEs contribute to the CME rate significantly during solar maximum. (6) The average speed of CMEs (averaged over Carrington Rotation periods) increases by a factor of 2 from $\sim$250 km s^{-1} during solar minimum to $\sim$550 km s^{-1} during solar maximum. (7) CMEs contain coronal material at a temperature of $\sim$ a few MK in the outer structure with cool prominence material ($\sim$8,000 K) in the core. (8) About a third of the ICMEs observed at 1 AU have flux rope structure (magnetic clouds). Non-cloud ICMEs generally originate at larger distances from the Sun center. (9) All CMEs are associated with flares, but there are many flares not associated with CMEs. (10) Type II radio bursts are signatures of electron acceleration by CME-driven shocks (Gopalswamy 2006). (11) Large SEP events are always associated with energetic CMEs.

4 CMEs and Flares

Mass motion in the form of H-alpha ejecta during flares has been identified as one of the key aspects of flare – CME association because $\sim$90% of such flares are associated with CMEs (see, e.g., Kahler 1992; Munro et al. 1979). On the other

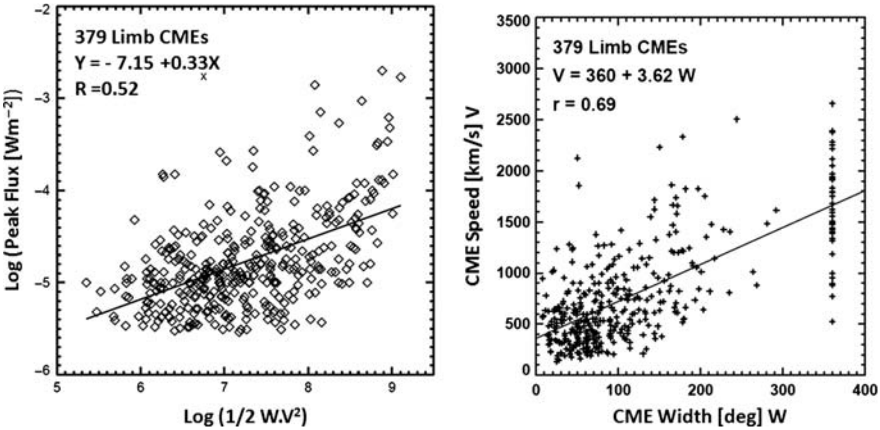


Fig. 10 Scatter plot between (*left*) peak soft X-ray flux and proxy kinetic energy and (*right*) CMEs width and speed. The data are for limb eruptions central meridian distance ($\geq 60°$)

hand not all flares are associated with CMEs. In fact, even some X-class flares are not associated with CMEs. Like CMEs, flares also occur from closed field regions, so, the released energy must go entirely into heating in flares without CMEs. There is some evidence that flares without CMEs are generally much hotter: the average temperature of flares without CMEs was found to be 18.4 MK compared to 11.7 MK for flares with CMEs (Kay et al. 2003).

For flares with CMEs, there is a reasonable correlation between flare size in soft X-rays and the CME kinetic energy (Hundhausen 1999; Moon et al. 2002). The correlation coefficient is only 0.53, suggesting that the partition between heating and mass motion may not be uniform in all eruptions. Figure 10a shows a similar plot between peak soft X-ray flux and the quantity 0.5 WV^2, where W is the CME width and V is the CME speed for a set of 342 CMEs associated with flare originating close to the limb. The limb events were chosen to avoid projection effects in the measured speeds. We take $^1/_2$WV^2 to be the kinetic energy of CMEs because the CME mass has been shown to be proportional to the CME angular width (Gopalswamy et al. 2005a). The correlation coefficient is also very similar to previous values. Figure 10b shows a reasonable correlation between CME width and speed. However, all flares are not associated with CMEs. In flares without CMEs, all the released energy must go into heating. On the other hand, in flares with CMEs, the free energy is partitioned between heating (soft X-ray flares) and mass motion (CMEs). A recent study of flares with and without CMEs involving parameters such as peak flux, fluence, and duration of soft X-ray flares found that they all followed a power law (Yashiro et al. 2006). However, the power law index for the CME-associated flares was >2, while that for CMEless flares was <2. This suggests that in flares without CMEs, the free energy goes entirely into heating the corona (Hudson 1991).

Studies on temporal correspondence between CMEs and flares have concluded that CME onset typically precedes the associated X-ray flare onset by several minutes. On the other hand, tiny flare-like brightenings and filament activation is known before the lift off of a filament (Gopalswamy et al. 2006b). In the 2005 May 13 example discussed above, the soft X-ray flare was reported to start at 16:13 UT. H-alpha movie shows filament activation more than 10 min before. The filament activation may indicate some kind of instability in the system. On the other hand the small-scale brightenings may push the system towards eruption. Thus, the flare-CME relationship is generally complex.

The spatial relation between flares and CMEs has also been studied by many authors (Harrison 2006; Kahler et al. 1989; Yashiro et al. 2008). The general consensus in the pre-SOHO era was that the flare was located anywhere under the span of the CME. However, using the extensive SOHO data base on CMEs, it was recently found that the CME leading edge is typically located radially above the flare location (Yashiro et al. 2008). They studied a set of 496 flare-CME pairs with solar sources away from the disk center (central meridian distance (CMD) $> 45°$). They found that the offset between the flare position and the central position angle (CPA) of the CME has a Gaussian distribution centered on zero (see Fig. 11). This finding calls for a closer flare-CME relationship as implied by the CSHKP eruption model (Yashiro et al. 2008). In fact, it has become common practice to identify the solar source of CMEs as the flare location derived from H-alpha or soft X-ray imaging observations.

Another important near-surface activity often associated with CMEs is prominence eruption (Gopalswamy et al. 2003; Munro et al. 1979; St. Cyr and Webb 1991). Sometimes CMEs are classified as flare-related and prominence-eruption related. This classification, however, is not very sharp because flares are also associated with prominence eruptions as demonstrated in Sect. 2. In fact, flares and filament eruptions correspond to the heating and mass motion aspects of eruptions, both of which are present in all CMEs. This classification may be meaningful for CMEs from

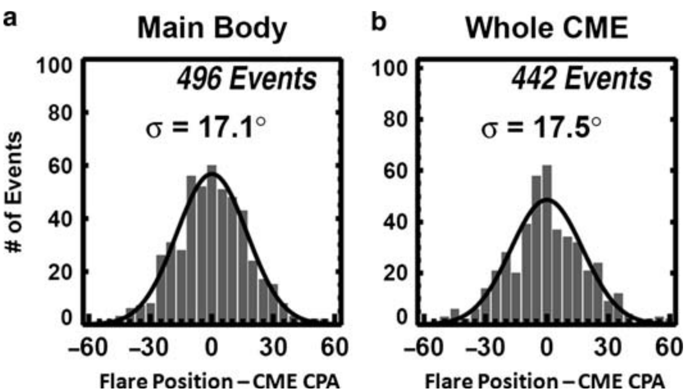


Fig. 11 Distributions of the offset of flare positions with respect to the central position angle (CPA) of the associated CMEs. The standard deviation (σ) obtained by fitting a Gaussian to the distribution (*solid curve*) is shown on the plot. The main body (probably the flux rope) and the whole CME (including the diffuse outer part) show similar relationship

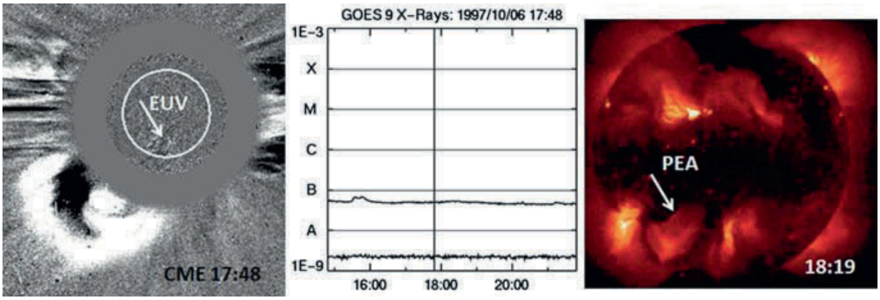


Fig. 12 (*left*) The CME on 1997 October 6 in the LASCO/C2 field of view with the EUV disturbance pointed by *arrow*. (*middle*) GOES soft X-ray plot showing no significant enhancement above the B-level background at the time of the CME marked by the vertical line. (*right*) Yohkoh Soft X-ray Telescope (SXT) image showing the post-eruption arcade (PEA) in the quiet Sun region

quiescent filament regions, but even there, one can see large-scale post-eruption arcades in soft X-ray and EUV images, but they are not intense enough to be above the background in GOES soft X-ray plots. Figure 12 shows an example. The large CME on 1997 October 6 originated from a quiescent filament region (the eruption can be seen as a EUV disturbance in the EIT image superposed on the LASCO image). The GOES X-ray background was very low, but the eruption does not show up in GOES soft X-rays. However, a large-scale post-eruption arcade (PEA) can be seen in the soft X-ray image taken at 18:19 UT. The arcade lasted for several hours.

5 CMEs, Shocks, Type II Bursts, and SEPs

Shocks have been inferred from metric type II radio bursts, which have been studied for more than six decades (see Gopalswamy 2006 for a review). There was a lot of controversy regarding the origin of shock waves (Cliver et al. 1999) responsible for the metric type II bursts (flare blast waves or CME-driven shocks). The root cause of the controversy was the observation that a third of metric type II bursts was not associated with CMEs. However, it was found that most of the solar eruptions associated with CMEless type II bursts originated close to the disk center. Coronagraphic observations favor CMEs originating from close to the limb. Only very energetic eruptions from close to the disk center will appear as halo or partial halo CMEs above the occulting disk (Cane et al. 2007). For example, when solar sources close to the limb are considered, all the metric type II bursts are associated with CMEs (Gopalswamy et al. 2005a; Munro et al. 1979). Therefore, all the metric type II bursts may be due to CME-driven shocks. Type II bursts at decameter-hectometric (DH) wavelengths originating from the near-Sun interplanetary medium (beyond about 3 $R_\odot$) are all associated with fast and wide CMEs even if the eruption occurs near the disk center. Some type II bursts occur at various bands from metric (near Sun) to kilometric (near Earth) wavelengths. CMEs associated with these mkm type II bursts have the highest speed and width. There is a progressive increase in speed and fraction of halo CMEs as one goes from CMEs associated with metric, DH, and mkm type II bursts. The number of type II bursts also decreases in the same order suggesting that only the most energetic of CMEs can produce type II bursts throughout the inner heliosphere. Another interesting result is that the speed and fraction of halos of CMEs associated with large SEP events is nearly identical to those of CMEs associated with mkm type II bursts. The reason is that the same intense shock accelerates electrons (observed as type II bursts) and ions (observed as SEP events in the interplanetary medium) (Gopalswamy et al. 2008).

White-light observations of shocks are rare, mainly because the shocks are very thin and the standoff distance between the flux rope and the shock is very small close to the Sun. Faint disturbances at the flanks of halo CMEs have been identified as shocks (Sheeley et al. 2000). A few other CMEs with possible shocks ahead of them have also been reported (Vourlidas et al. 2003). Figure 13 shows another example in which the leading diffuse feature can be identified as a shock. This event has a distinct flux rope and CME core, so the outer edge of the diffuse feature can be identified as a shock. In direct images, the diffuse feature was not visible, but a kink (marked S in Fig. 13) in the streamer indicates a disturbance. In the difference image, the kink location is seen as the intersection of the diffuse feature with the streamer. The diffuse feature itself can be identified as the shock sheath.

The close physical relationship among large flares (M and X class), fast and wide coronal mass ejections, DH type II radio bursts (including mkm type II bursts), CME-driven shocks observed *in situ*, and large SEP events is revealed in Fig. 14. We have plotted the number of events under each category summed over Carrington rotation periods. The numbers are close to one another having a similar variation with the solar cycle. The rates peak in the solar maximum phase (1999–2002), but there are occasional large spikes, which are due to some active regions that are prolific producers of eruptions in quick succession. There are many major flares, which are not associated with CMEs or type II bursts and hence the higher rate for M- and X-class flares. All the other events are physically related: FW CMEs

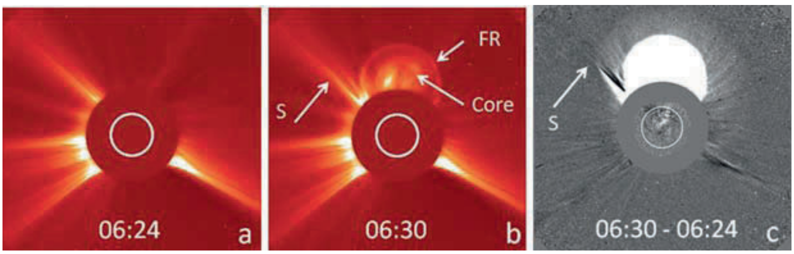


Fig. 13 Possible shock observed in a SOHO/LASCO/C2 image on 2005 January 15. (**a**) pre-event LASCO image at 06:24 UT. (**b**) LASCO image at 06:30 UT with CME appearing above the north-west limb with a flux rope (FR) structure and a bright core; a kink is obvious in the streamer at the location marked by "S". (**c**) Difference image (06:30 minus 06:24), which shows the coronal change over a period of 6 min; this image shows the diffuse structure (S), which is likely to be the shock. The diffuse structure envelopes the flux rope and is more extended than the flux rope. A EUV difference image is superposed on the LASCO difference image to show that the solar source is located in the northern hemisphere of the Sun

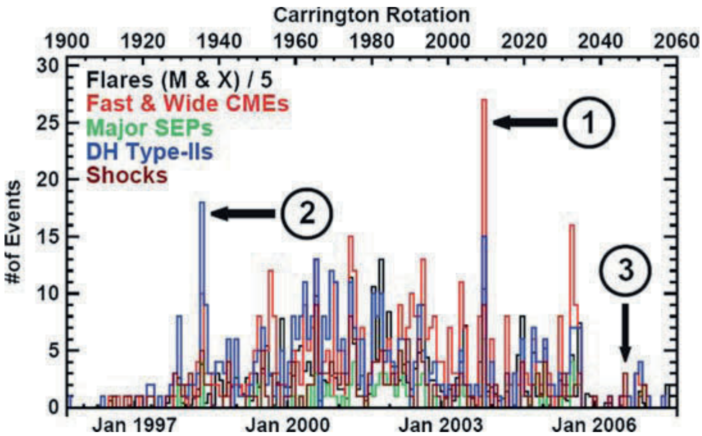


Fig. 14 Numbers of major flares (divided by a factor of 5 to fit the scale), fast and wide (FW) CMEs, DH type II bursts (including mkm type II bursts), large SEP events, and in-situ shocks plotted as a function of time. Large SEP events are those with proton intensity in the >10 MeV energy channel exceeding 10 pfu; pfu is the particle flux unit defined as the number of particles/(cm^2 s sr). The numbers are binned over Carrington Rotation periods ($\sim$27 days). Examples of FW CMEs without type II bursts and SEPs (1), DH type II without FW CMEs (2), and interplanetary shocks without DH type II bursts and SEPs (3) are indicated by *arrows*

drive shocks, which accelerate electrons (producing type II radio bursts) and ions (observed in situ as SEP events). The shocks are also detected in situ when they survive the Sun-Earth transit, which can last anywhere from <1 day to >3 days. Figure 14 reveals some exceptions: (1) FW CMEs without type II bursts, (2) DH type II bursts without FW CMEs, and (3) in-situ shocks not accompanied by other types of energetic events. FW CMEs without type II bursts indicate that occasionally the Alfven speed in the ambient medium can be very high (up to 1,600 km s^{-1}). DH type II bursts (and sometimes SEP events) associated with CMEs as slow as $\sim$400 km s^{-1} indicate that very low coronal Alfven speed ($<$400 km s^{-1}) can prevail in the corona. Shocks without associated type II burst or SEP event generally form at large distances from the Sun and are extremely weak. These are likely to be "subcritical" shocks, which do not accelerate particles in significant numbers (Mann et al. 2003).

It is clear that every large SEP event is associated with a unique CME that is highly energetic. In the above discussion, we have tacitly assumed that the SEPs are accelerated by the CME-driven shock supported by the direct evidence provided by the ESP events. The flare reconnection site is another place where particles are accelerated. Hard X-ray bursts and type III radio bursts during flares without CMEs provide strong evidence for particle acceleration by a non-shock process. Historically, SEPs produced in flares and propagating into the IP medium are referred to as "impulsive events", which are of very short duration and small intensity compared to the large SEP events (also known as the gradual events). Similarly the gradual SEP events were also studied only from the point of view of the associated CMEs (and shocks). Since each of the CMEs associated large SEP events are also associated with a major flare, one naturally expects contributions from the flares. Flare particles in CME events are also expected to be of higher intensity because of the larger flare size compared to those associated with impulsive events. While there is no question that the two mechanisms operate in nature, but what was missed is the fact that they operate simultaneously in every CME event. There may even be interaction between the two sources in that the flare particles may pass through the shock region and get further accelerated. There is an ongoing debate on the relative contribution from the flare and shock sources (Cane et al. 2007; Tylka and Lee 2006).

6 CMEs and Geomagnetic Storms

Geomagnetic storms occur when an interplanetary structure such as an ICME containing southward magnetic field arrive at Earth's magnetosphere and reconnects with the geomagnetic field. The storm strength is measured by one of the indices such as the Dst index. A white-light CME observed near the Sun may result in a shock, sheath, and the driving ICME at 1 AU. The shock and the ICME front boundary delineate the sheath region. A CME can be geoeffective due to a south-pointing magnetic field component in the sheath region or in the ICME (ejecta) region. When the ICME is an MC, then one can infer which section of the MC is geoeffective based on the magnetic structure of the MC (flux rope orientation with respect to the ecliptic plane and its sense of rotation). Although most of the large geomagnetic storms are caused by magnetic clouds, non-cloud ejecta can also be geoeffective. Depending on the magnetic structure in the sheath and cloud portions, the following situations are encountered: (1) sheath and MC are geoeffective, (2) sheath alone is geoeffective, (3) MCs alone are geoeffective, and (4) none are geoeffective. The last situation arises when neither the sheath nor the MC contains south-pointing magnetic field component. When both sheath and MC are geoeffective, the Dst profile can be complex depending on the location of the south-pointing field in the sheath and cloud portions.

6.1 Halo CMEs

Halo CMEs constitute a special population selected by the nature of coronagraphic observations: slow and narrow CMEs originating from close to the Sun center are not usually seen in coronagraph images. Typically, fast (CME speed $\geq$900 km s^{-1}) and wide CMEs (width $\geq$60°) appear as halos. Occasionally some halos are slow. Figure 15 shows that the average speed of disk halos (with the central meridian distance (CMD) of the solar source $\leq$45°) is $\sim$933 km s^{-1}, which is close to that of the backside halos (977 km s^{-1}). However, the average speed of limb halos (45°$<$CMD$\leq$90°) is $\sim$1,548 km s^{-1}. There are two reasons for the higher speed

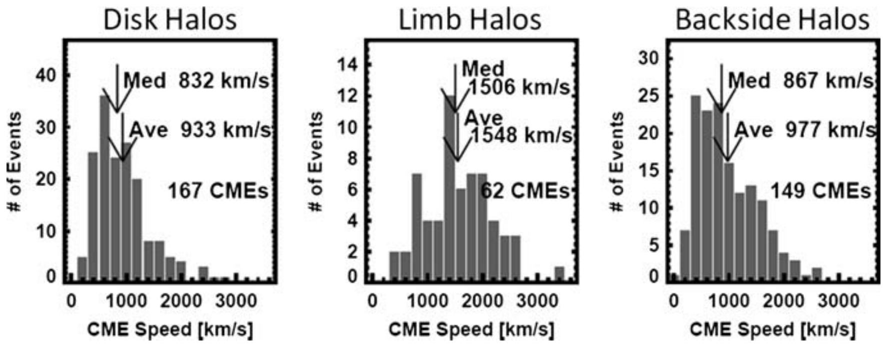


Fig. 15 Speed distributions for (*left*) disk halos (CMD $\leq 45°$), (*middle*) limb halos ($45° <$ CMD $\leq 90°$), and (*right*) backside halos (CMD $> 90°$)

of limb halos: (1) the disk halos originate close to the disk center, so their speeds are severely affected by projection effects in contrast to the limb halos, which are not significantly affected by projection effects; (2) halo CMEs by definition have to surround the occulting disk, so a CME originating close to the limb should travel very fast to be observed above the opposite limb. Thus only the fastest limb CMEs can become halos. The backside halos are far behind the limb and hence are subject to projection effects similar to the disk halos. The average speed of all halos is $\sim$1,000 km s^{-1}, which is more than two times the average speed of all CMEs. For such speeds, the average width of the CMEs can be inferred as $\sim$180° from Fig. 10b. Thus the halo CMEs are generally more energetic. This is further confirmed by the fact that the average size of soft X-ray flares associated with halo CMEs is an order of magnitude larger than the average size of all the flares observed in solar cycle 23 (Gopalswamy et al. 2007). As Fig. 10 shows, higher flare intensity implies higher kinetic energy for the associated CMEs. Higher kinetic energy enables the CMEs to propagate far into the interplanetary medium. From coronagraphic observations alone, it is difficult to say whether a halo CME is frontsided or not. In Fig. 5 the halo CME is front sided (i.e., Earth-directed) as can be seen in the superposed EUV images. We need to combine coronagraphic observations with coronal images obtained in EUV, soft X-ray, microwave, or H-alpha wavelengths to identify disk activity.

6.2 Geoeffective CMEs

The source locations of halo CMEs are compared with those of ICMEs in Fig. 16. We have plotted the solar sources MCs and non-cloud ICMEs separately. MCs have higher magnetic field, smooth rotation, and low proton temperature (indicating flux rope structure Burlaga et al. 1982). Non-cloud ICMEs do not appear as flux ropes in in-situ observations. The events in Fig. 16 are the most energetic events in the

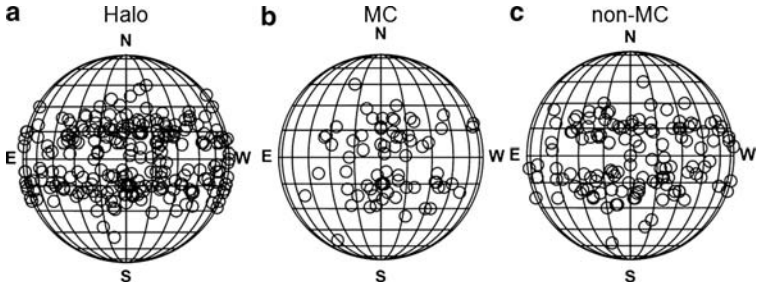


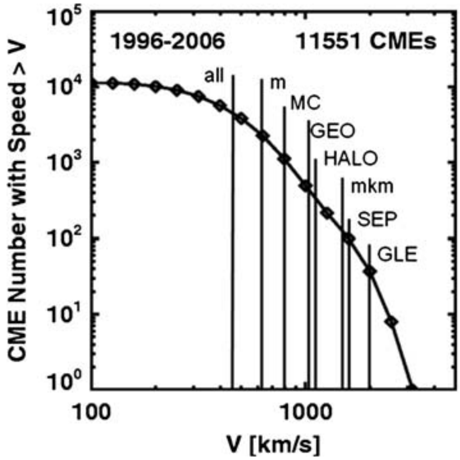
Fig. 16 Solar sources of halo CMEs (**a**) and shock-driving ICMEs (**b**)

heliosphere and they are due to energetic CMEs from the Sun. The close similarity between halo CME sources and the sources of ICMEs indicate that there is high degree of overlap between halo CMEs and the CMEs that become ICMEs. Those ICMEs that have south-pointing magnetic field in the ejecta part or sheath part produce geomagnetic storms. The product of the speed and magnetic field strength in ICMEs has been shown to have a good correlation with the intensity of geomagnetic storms. In a step further, it was recently shown that CME speed and its product with the magnetic field in the associated ICME have also equally good correlation with the Dst index (Gopalswamy 2008). This raises the possibility of storm prediction based on remote-sensing measurements of CMEs if one can estimate the magnetic field strength and orientation in CMEs when they are still near the Sun.

7 Summary

We have made a significant progress in understanding the origin, propagation and heliospheric consequences of solar eruptions, thanks to the continuous data acquired by the SOHO mission on CMEs and their solar sources. We are also able to understand the kinematic and magnetic connection between the ICMEs observed in the IP medium and the white-light CMEs remote sensed by coronagraphs. Concentrated efforts by NASA's Living with a Star (LWS) program and the US National Space Weather program have promoted efforts to find the chain of connectivity from the Sun all the way to the geospace and ground.

The whole solar cycle worth of CME data from SOHO have taught us that only a small subset of energetic CMEs are responsible for severe heliospheric consequences. This can be easily understood from Fig. 17, which shows the cumulative number of CMEs above a certain speed V. We have marked the average values of various subsets: (1) CMEs associated with metric type II radio bursts (m). (2) CMEs associated with magnetic clouds (MC). (3) CMEs that cause major geomagnetic storms (GEO). (4) Halo CMEs that appear to surround the occulting disk of the coronagraph (HALO). (5) CMEs associated with type II bursts having emission components at all wavelengths (mkm). (6) CMEs producing large solar



CME population	<V> km/s
All CMEs	471
metric II CMEs	610
MC CMEs	774
Geoeffective CMEs	1042
Halo CMEs	1052
mkm II CMEs	1500
SEP CMEs	1600
GLE CMEs	2000

Fig. 17 Cumulative distribution of CME speeds with the average speeds (<V>) of various special populations marked (*left*) and listed (*right*)

energetic particle (SEP) events. (7) CMEs associated with the ground level enhancement (GLE) in SEPs, which are a subset of SEP events, but associated with the fastest population of CMEs. These are rare events with the average CME speed of $\sim$2,000 km s^{-1} (Gopalswamy et al. 2005b). The three fastest populations (mkm, SEP, GLE) all drive shocks and accelerate particles. The three populations with intermediate speed (MC, GEO, and HALO) represent CMEs that directly impact Earth's magnetosphere. The m population also drives shocks, but very close to the Sun. CMEs faster than 1,000 km s^{-1} contain GEO, HALO, SEP-related, mkm type II related, and GLE-related CMEs. There are $\sim$1,000 CMEs over the solar cycle, which have speeds exceeding $\sim$1,000 km s^{-1}. In other words, $\sim$10% of the fastest CMEs make significant impact on the heliosphere.

References

Burlaga, L.F., Klein, L., Sheeley, N.R. Jr., Michels, D.J., Howard, R.A., Koomen, M.J., Schwenn, R., Rosenbauer, H.: Geophys. Res. Lett. **9**, 1317 (1982)
Cane, H., Richardson, I., VonRosenvinge, T.: Space Sci. Rev. **130**, 301 (2007)
Cliver, E.W., Webb, D.F., Howard, R.A.: Solar Phys. **187**, 89 (1999)
Davis, K.: Ionospheric Radio, p. 580. Peter Peregrinus, London (1990)
Gopalswamy, N.: In: Poletto, G., Suess, S. (eds.) The Sun and the Heliosphere as an Integrated System, p. 201. Springer, New York (2004)
Gopalswamy, N.: In: Gopalswamy, N., Mewaldt, R., Torsti, J. (eds.) Solar Eruptions and Energetic Particles. Geophysical Monograph Series vol. 165, p. 207. AGU, Washington DC. (2006)
Gopalswamy, N.: In: Hasan, S.S., Banerjee, D. (eds.) Kodai School on Solar Physics, vol. 919, p. 275. AIP conference Proceedings. AIP, New York (2007)
Gopalswamy, N.: J. Atmos. Sol. Terr. Phys. **70**, 2078 (2008)

Gopalswamy, N., Shimojo, M., Lu, W., Yashiro, S., Shibasaki, K., Howard, R.A.: Astrophys. J. **586**, 562 (2003)
Gopalswamy, N., Aguilar-Rodriguez, E., Yashiro, S., Nunes, S., Kaiser, M.L., Howard, R.A.: J. Geophys. Res. **110**, 12S07 (2005a)
Gopalswamy, N., Xie, H., Usoskin, I.: In: Acharya B.S., Gupta S., Jagadeesan P., Jain A., Karthikeyan S., Morris S., Tonwar S. (eds.) Proceedings of the 29th International Cosmic Ray Conference. vol. 1, p.169. Tata Institute of Fundamental Research, Mumbai (2005b)
Gopalswamy, N., Mewaldt R., Torsti J. (eds.): Geophysical Monograph Series, vol. 165, p. 385. AGU, Washington DC (2006a)
Gopalswamy, N., Mikic, Z., Maia, D., Alexander, D., Cremades, H., Kaufmann, P., Tripathi, D., Wang, Y.-M.: Space Sci. Rev. **123**, 303 (2006b)
Gopalswamy, N., Yashiro, S., Akiyama, S.: J. Geophys. Res. **112**, A06112 (2007)
Gopalswamy, N., Yashiro, S., Xie, H., Akiyama, S., Aguilar-Rodriguez, E., Kaiser, M.L., Howard, R.A., Bougeret, J.-L.: Astrophys. J. **674**, 560 (2008)
Harrison, R.A.: In: Gopalswamy, N., Mewaldt, R., Torsti, J. (eds.) Solar Eruptions and Energetic Particles. Geophysical Monograph Series, vol. 165, p. 73. American Geophysical Union, Washington DC (2006)
Howard, R.A., Michels, D.J., Sheeley, N.R. Jr., Koomen, M.J.: Astrophys. J. **263**, L101 (1982)
Hudson, H.S.: Solar Phys. **133**, 357 (1991)
Hundhausen, A.: In: Strong K.T., Saba J.L.R., Haisch B.M., Schmelz J.T. (eds.) The Many Faces of the Sun: A Summary of the Results from NASA's Solar Maximum Mission, p. 143. Springer, New York. (1999)
Kahler, S.W., Sheeley, N.R. Jr., Ligget, M.: Astrophys. J. **344**, 1026 (1989)
Kahler, S.W.: Ann. Rev. Astron. Astrophys. **30**, 113 (1992)
Kahler, S. W.: In: Gopalswamy, N., Mewaldt R., Torsti J.: (eds.) Solar Eruptions and Energetic Particles. Geophysical Monograph Series, vol. 165, p. 21. American Geophysical Union, Washington DC (2006)
Kay, H.R.M., Harra, L.K., Matthews, S.A., Culhane, J.L., Green, L.M.: Astron. Astrophys. **400**, 779 (2003)
Mann, G., Klassen, A., Aurass, H., Classen, H.-T.: Astron. Astrophys. **400**, 329 (2003)
Moon, Y.-J., Choe, G.S., Wang, H., Park, Y.D., Gopalswamy, N., Yang, G., Yashiro, S.: Astrophys. J. **581**, 694 (2002)
Munro, R.H., Gosling, J.T., Hildner, E., MacQueen, R.M., Poland, A.I., Ross, C.L., Solar Phys. **61**, 201 (1979)
Sheeley, N.R. Jr., Hakala, W.N, Wang, Y.-M.: J. Geophys. Res. **105**, 5081 (2000)
Song, P., Singer, H.J., Siscoe, G.L.: (eds.) Space Weather. Geophysical Monograph Series, vol. 125, p. 440. American Geophysical Union, Washington DC (2001)
St. Cyr, O.C., Webb, D.F., Solar Phys. **136**, 379 (1991).
Tylka, A., Lee, M.A.: Astrophys. J. **646**, 1319 (2006)
Vourlidas, A., Wu, S.T., Wang, A.H., Subramanian, P., Howard, R.A.: Astrophys. J. **598**, 1392 (2003)
Yashiro, S., Akiyama, S., Gopalswamy, N., Howard, R.A.: Astrophys. J. **650**, L143 (2006)
Yashiro, S., Michalek, G., Akiyama, S., Gopalswamy, N., Howard, R. A.: Astrophys. J. **673**, 1174 (2008)

Solar Energetic Particles: Acceleration and Observations

Takashi Sako

Abstract Research of solar energetic particles (SEPs) is important in understanding particle acceleration, transport and interactions taking place in the universe. The importance of space weather to modern human life is also increasing. In this lecture, I introduce a selected subset of SEP observations together with observation techniques and future plans. The aim is to connect these SEP observations with associated particle acceleration mechanisms and the subsequent transport and interaction processes. Because the observational properties are determined by different processes, a wide range of observations is necessary in order to fully understand the phenomena taking place. I will also give an overview of the role of the SEP studies in general astrophysics.

1 What are Solar Energetic Particles?

We know that the Sun is shining steadily in visible light, but it is more dynamic in shorter wavelengths, i.e. at higher energies. Dynamic solar activity sometimes produces violent eruptions that can release huge amounts of energy during short time intervals. In these eruptions, such as solar flares and coronal mass ejections, particles can be accelerated to relativistic energies. These particles and their high energy secondaries are called solar energetic particles, SEPs in short. Particle acceleration is a common process taking place all over the universe, e.g. in supernova remnants, active galactic nuclei, gamma-ray bursts, and so on (Axford 1981). However, details of the mechanism of particle acceleration are not yet fully understood, especially those of ion acceleration, even though ions constitute a major part of cosmic-rays. Because the Sun is one of the efficient astronomical ion accelerators, it is a natural laboratory to study particle acceleration processes. Although the maximum energy achieved by acceleration near the Sun is far below that of the cosmic-rays, we know

T. Sako
Solar-Terrestrial Environment Laboratory, Nagoya University, Nagoya 464-8601, Japan
e-mail: sako@stelab.nagoya-u.ac.jp

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_5,

when and where the particles are accelerated. Together with simultaneous high-resolution observations over a wide range of wavelengths and particle species, the study of SEPs offer great advantages in the investigation of particle acceleration processes.

SEPs can have important effects on the present day society. Our modern life strongly depends on satellites for weather forecasts, car navigation, TV transmission and so on, even though we may not be aware of that. SEPs hitting electronics of these satellites can produce errors in their functions or can damage them. Energetic particles are also dangerous for human health if they hit the human body. Usually we are protected by the thick absorbing layer of the Earth's atmosphere, but astronauts and airplane crews can be directly irradiated by particles. Therefore, the radiation protection of satellite electronics and humans is important, and it is one of the pressing needs for space weather forecasting. Good understanding of the basic processes of particle acceleration and transport is necessary for predicting the arrival of SEPs.

Both charged and neutral particles carry information about their acceleration site. Although they arrive at Earth via different paths and have gone through different processes. In order to extract from the observational data the information about particle acceleration, we need to understand the processes that are illustrated in Fig. 1. In other words, the aim of SEPs studies is to understand these processes in a consistent manner. Because any observation method has advantages and disadvantages, we need to use various angles of approach. In this lecture, I first introduce various SEP observations together with the associated measurement technique. Then I discuss what happens to the particles from the time they are accelerated until they are observed, i.e. I describe acceleration, transport and interaction processes.

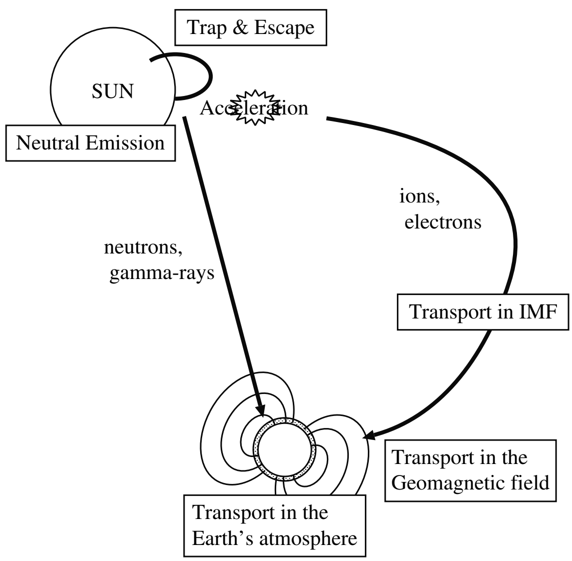


Fig. 1 A schematic view of the processes SEPs can experience

2 Observations of SEPs

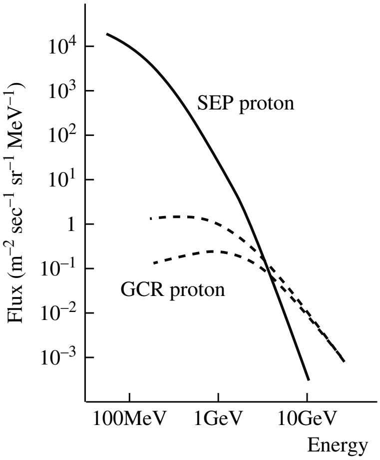
Because SEPs can be absorbed in the Earth's atmosphere ($\sim$1,000 g cm^{-2} or 10 m water equivalent at the sea level), observations have to be made both in space and on the ground, depending on particle species and energy. Electrons and gamma-rays can be observed only in space. Also ions with energies below 100 MeV can be observed only from space because the atmosphere strongly attenuates the particle flux (Shibata 1994). On the other hand, ions above 100 MeV can be observed by ground based detectors. In this section, I introduce a selected subset of observations: ground-based observations of ions (presumably protons) and neutrons, and spaceborne observations of gamma-rays.

2.1 Proton Observations

Events where solar particles are observed by ground-based detectors are called ground level enhancements (GLEs). GLE events are usually identified with a world-wide network of neutron monitors (NMs). Over more than 50 NM observatories are distributed throughout the world to monitor the long-term intensity variation of galactic cosmic-rays. Because the particles entering the atmosphere experience inelastic collisions several times, surviving particles at the ground level are mainly neutrons. NMs can detect them with a high efficiency (Clem and Dorman 2000; Hatton 1971). As will be discussed in Sect. 3.2, protons travel along the field lines of the interplanetary magnetic field (IMF). The particle transport is a diffusive process. In the early phase of a GLE, particles arrive at Earth from the direction of IMF connected to the Sun, while particle stream becomes more isotropic in the later phase. The detailed conditions are different from event to event. Due to the geomagnetic field, each NM depending on the observation site has a varying sensitivity to particle arrival directions and rigidities. This means that the network of the NMs can act as a single powerful detector that is sensitive to the direction distribution (anisotropy) and energy spectrum of primary protons in interplanetary space. Bieber et al. have organized NMs into such a network ("*the Spaceship Earth*") and measured the properties of protons (Bieber 2004). The energy spectrum is observed sometimes to continue up to 10 GeV, but at the same time there are indications of spectral roll-off in the highest energy meaning that the particle acceleration is limited to energies below the roll-off. An energy spectrum derived by Bombardieri et al. (2007) in the 15 April 2001 GLE is shown in Fig. 2.

Observations of GLEs with non-NM type detectors are also carried out, e.g. by Solar Neutron Telescopes (SNTs) based on the plastic scintillators (Muraki et al. 2007), Milagro which uses a large water pool designed for gamma-ray astronomy (Milagro et al. 2007), and the URAGAN muon hodoscope (Timashkov et al. 2007). These new experiments have several detection channels with a different sensitivity in energy and direction so that observations at a single station can give some information on the energy spectrum and anisotropy. Continuous monitoring in the low

Fig. 2 Energy spectrum of the SEPs observed in the 15 April 2001 GLE (*solid line*) together with the galactic cosmic-ray proton spectra (*dashed lines*). SEP particles sometimes overwhelm the ubiquitous GCR flux below 10 GeV. Two dashed lines indicate the GCR fluxes modulated by the different level of solar activity (Shikaze et al. 2007)



energy range by the *GOES* satellite is also important in order to cover a wide range of energies.

2.2 *Neutron Observations*

Because of its electrical neutrality, neutrons are not affected by magnetic fields. That frees us from the problem of particle transport in the interplanetary magnetic field and geomagnetic field. On the other hand, because any acceleration process works only on charged particles, neutrons are emitted as secondary particles through certain interactions as introduced in Sect. 3.3. So far, a dozen solar neutron events have been recorded mainly by NMs. To detect a relatively low flux of secondary neutrons effectively, a dedicated network of the SNTs is also in operation (Muraki et al. 2007). The SNTs are installed on high mountain tops at low latitudes to avoid the large attenuation of the particle flux in the Earth's atmosphere.

Because the neutrons fly along direct paths from the emission site to Earth, we can derive their emission time profile, if we can measure the velocity of neutrons that is a function of energy. The SNTs, which are designed to measure the energy of particles, can determine this emission profile. Sako et al. reported a long-lived emission of neutrons during an event that occurred on 7 September 2005 based on the observations of SNTs (Sako et al. 2006). The energy spectra of solar neutrons are well described by a power law up to the GeV energy (Watanabe et al. 2006), although in the case of the 7 September 2005 event, a cut-off at 500 MeV was observed (Sako et al. 2006).

To avoid the atmospheric attenuation, spaceborne observations are ideal, even though the instruments do not have a wide effective area and are unable to measure high energies. Although, the first detection of solar neutrons was made by an instrument on board the *Solar Maximum Mission*/Gamma-Ray Spectrometer (Chupp et al.

1982). So far the observations of solar neutrons in space have been made by gamma-ray detectors. A dedicated detector of solar neutrons proposed by Imaida et al. was launched in 2009 (Imaida et al. 1999).

2.3 Gamma-Ray Observations

All observations of solar gamma-rays are made in space. As discussed in Sect. 3.3, gamma-rays are emitted in several different processes, so their detection requires different observation techniques. Below 20 MeV, the aim is to measure the gamma-ray emission lines from excited nuclei or electron-positron annihilation. For measurements in this energy range, the excellent energy resolution of semi-conductor detectors is required. Currently these observations are made by the *Reuven Ramaty High-Energy Solar Spectroscopic Imager (RHESSI)* (Lin et al. 2002). The *RHESSI* satellite has succeeded in the first imaging of the gamma-ray emission site in this energy range. Hurford et al. reported that the ions and electrons lost energy on the solar surface at slightly different places (Hurford et al. 2003).

At higher energies, gamma-rays are emitted via the decay of π^0 mesons. In this energy range, a wide dynamic range and a large effective area are more important than the energy resolution. *SMM*/GRS, *CORONAS-F*/SONG are specially dedicated for such observations. Other detectors not dedicated for solar observations such as the *Compton Gamma-Ray Observatory*/Energetic Gamma-Ray Telescope Experiment also have a chance to make important observations (Kanbach et al. 1993).

The ion induced gamma-rays are overlaid with a continuum of the electron induced bremsstrahlung radiation as shown in Fig. 3 (Ramaty and Mandzhavidze 2001).

As is the case of neutrons, also gamma-rays arrive along direct paths from the emission site. As they travel with the speed of light, their arrival times do not show an energy dependent time dispersion like in the case of neutrons. The time profile of gamma-rays is a good estimator for the time when ions interact with the solar atmosphere (Watanabe et al. 2006). On the other hand, because the effective energy of the parent ions that excite nuclei is below 100 MeV, it is uncertain if the time indicator is valid for the relativistic ($\sim$1 GeV) ions.

3 Acceleration, Emission and Transport of SEPs

3.1 Acceleration of SEPs

Acceleration mechanisms near the Sun are not well known. The suggested theoretical models include (a) 1st order Fermi acceleration (or shock acceleration), (b) 2nd order Fermi acceleration (stochastic acceleration), and (c) DC electric field acceleration. The idea of shock acceleration is widely used to explain particle acceleration

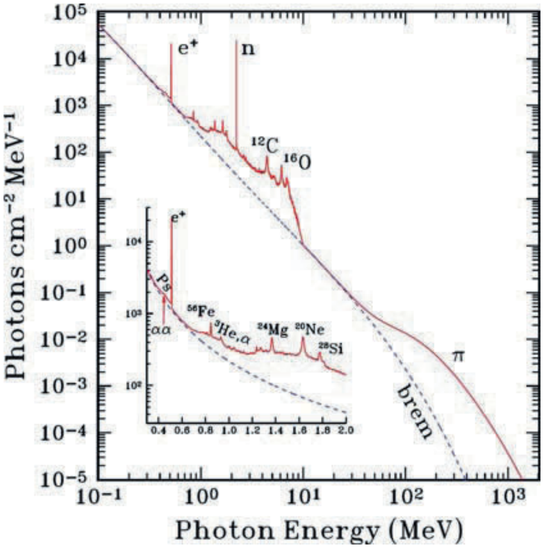


Fig. 3 Theoretical solar flare gamma-ray spectrum. Characteristic features due to the line emissions are seen below 20 MeV overlaid with continuous bremsstrahlung emission. Continuum above 100 MeV shows a distinctive feature due to the π^0 meson decay above bremsstrahlung

in various astrophysical objects such as supernova remnants. A theoretical spectral form, proposed by Ellison and Ramaty, models the shock spectrum as a power law with an exponential cut-off (Ellison and Ramaty 1985). A theoretical spectral form for the stochastic acceleration was suggested by Ramaty and Murphy (1987). Bombardieri et al. have compared an observed GLE spectrum with those predicted by the two models, and they concluded that both mechanisms worked at different stages of a single event (Bombardieri et al. 2006). However, we have not reached a general conclusion that could explain all SEP events.

3.2 Transport of SEPs

As moving charged particles feel the Lorentz force in the magnetic field, they cannot follow direct paths to Earth. The gyro-radius of a 100 MeV proton in the magnetic field of 3×10^{-5} Gauss, that is typical value near the Earth's orbit, is 5×10^8 m. Because this is far smaller than 1 astronomical unit, SEPs travel along field lines that form a large-scale magnetic field called the Parker spiral. The magnetic field line crossing at Earth originates near the western limb of the Sun, therefore, the charged particles emitted in this part of the Sun arrive at Earth more efficiently and more rapidly. Particles emitted along the direction of the field line (cosine of the pitch angle $\sim$1) can arrive at the time $t = L/v$ after their release at the Sun (here, L is the length of the magnetic field line and v is the velocity of the particle). By measuring the arrival times of first particles, the emission time can be estimated

(Saiz et al. 2005). At later times, particles with various pitch angles arrive, making such an analysis impossible. Though the prompt component comes along the direction of the magnetic field line, the later component is more isotropic. Changes in the pitch angle of particles (pitch angle scattering) means that particle transport is a diffusive process that changes the direction distribution of particles into more isotropic. Particle transport in the IMF together with the geomagnetic field effects is well studied and applied to the analysis of GLEs, for example, by Saiz et al. (2008) and Bombardieri et al. (2006).

3.3 Emission of Neutral particles

Neutral particles are emitted in interactions between charged particles and the solar atmosphere. Particles moving down from the acceleration site towards the solar surface efficiently produce these neutral particle emissions. The connection of the solar surface to the acceleration site is thought to be via loop-like structures of solar magnetic field. In general, the loop has stronger magnetic field at its feet, therefore particles with small pitch angles are reflected (mirrored) and trapped in the loop. Particles penetrating deeper into the solar atmosphere or trapped in the loop for longer time periods are more likely to interact with the atmospheric elements. High energy neutrons are directly produced in the inelastic interactions. High energy π^0 mesons are also produced in these inelastic interactions, and they immediately decay into high-energy gamma-rays. Low energy neutrons produced in scattering events are thermalized with a time scale of 100 sec. Thermalized neutrons are then captured by hydrogen nuclei producing excited deuterium nuclei. Finally the deuterium emits a 2.223 MeV line gamma-ray delayed by 100 sec relative to the other emission. Nuclei like carbon and nitrogen are also excited by charged particles, and they emit line gamma-rays without a delay. These complicated processes can be modeled and modeling results can be applied to the observed neutron and gamma-ray profiles (Hua et al. 1987; Murphy et al. 2007).

Through the above description of the involved processes it is clear that the spectra of neutrons and π^0 decay gamma-rays reflect the spectrum of the parent ions. Although the intensity of the line gamma-rays carries only integrated information on the spectrum of parent ions, comparison of the intensities in different lines can give a good estimate of that spectrum (Hua et al. 1987).

4 Summary

Though I could not fully cover the topics on electrons and ions except proton, they are also equally important for the study of SEPs. In the observations of ions, not only the atomic number but also their charge state gives us significant information. The study of SEPs presents us with a challenge to understand the various observations and trace them back to their origin at the particle acceleration site. Most likely particle acceleration events at the Sun have some variations. Therefore, systematic studies to classify them are important. The study of heliospheric phenomena seems sometimes quite complicated compared with the other astrophysical studies, because our observations are far more detailed and far more precise. On the other hand, SEP studies can contribute to the knowledge of general astrophysics, because only SEP studies are based on in-situ observations.

Finally, I want to discuss shortly about the highest energy particles. Figure 4 is the famous Hillas diagram showing possible highest energies for various celestial objects (Hillas 1984). The idea is that when the gyro-radius of particles reaches the size of the object, no more acceleration is expected, because such particles can escape from the object. Typical size versus magnetic field strength for various objects is plotted together with the corresponding maximum particle energy.

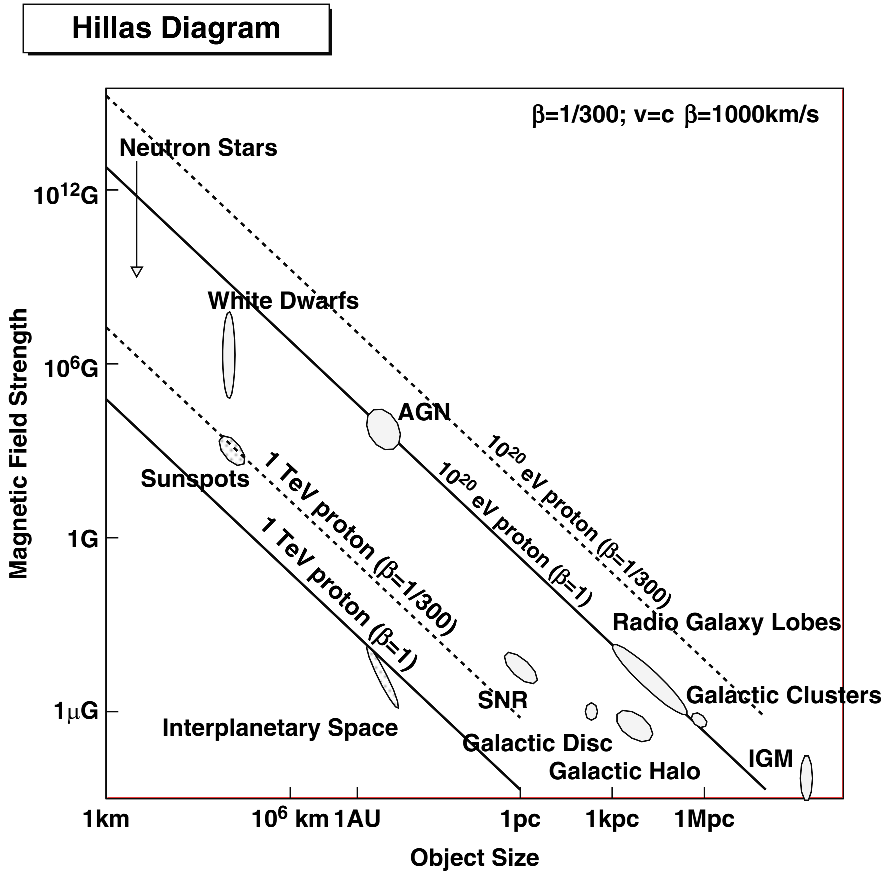


Fig. 4 The Hillas diagram showing the typical magnetic field strengths and sizes of various celestial objects. The maximum achievable energies (1 TeV and 10^{20} eV) are also drawn in the figure

Though the plot was originally introduced to estimate the possible sources of the highest-energy cosmic-rays at 10^{20} eV, it also predicts the highest energy of SEPs. Under certain possible conditions, heliospheric phenomena are capable to accelerate particles up to TeV energy. So far, we know that the energy of SEPs reaches up to 10 GeV. Uncovering this 2 orders of magnitude difference is one of the interesting challenges in SEP studies, and it has also significance for general particle acceleration studies.

Acknowledgements The author is grateful to Professor T. Terasawa for giving me an opportunity to give a lecture at the Kodai IHY school. Dr. K. Watanabe has provided some useful information about the gamma-ray and neutron observations used in the lecture and this text.

References

Axford, W.I.: Proc. 17th Int. Cosmic Ray Conf., Paris, **12**, 155 (1981)
Shibata, S.: J. Geophys. Res. **99**(A4), 6651–6665 (1994)
Clem, J.M., Dorman, L.I.: Space Sci. Rev. **93**, 335–359 (2000)
Hatton, C.J.: "The neutron monitor". In: Wilson, J.G., Wouthuysen, S.A. (eds.), Progress in Elementary Particle and Cosmic Ray Physics, vol. 10. North Holland, Amsterdam (1971)
Bieber, J.W.: Astrophys. J. **601**, L103–L106 (2004)
Bombardieri, D.J., et al.: Astrophys. J. **665**, 813–823 (2007)
Muraki, Y., et al.: Astropart. Phys. **29**, 229–242 (2008)
Morgan, T., et al.: 30th Int. Cosmic Ray Conf., Merida, (2007)
Timashkov, D.A., et al.: 30th Int. Cosmic Ray Conf., Merida, (2007)
Sako, T., et al.: Astrophys. J. **651**, L69–L72 (2006)
Watanabe, K., et al.: Adv. Space Res. **38**, 425–430 (2006)
Chupp, E.L., et al.: Astrophys. J. **263**, L95–L98 (1982)
Imaida, I., et al.: Nucl. Instrum. Methods Phys. Res. **A421**, 99–112 (1999)
Lin, R.P., et al.: Solar Phys. **210**, 3–32 (2002)
Hurford, G.J., et al.: Astrophys. J. **595**, L77–L80 (2003)
Shikaze, Y., et al.: Astropart. Phys. **28**, 154–167 (2007)
Kanbach, G., et al.: Astron. Astrophys. Sup. **97**, 349–353 (1993)
Ramaty, R., Mandzhavidze, N.: "Solar Flares: Gamma Rays", Encyclopedia of Astronomy and Astrophysics, Edited by Paul Murdin, article 2292. DOI: 10.1888/0333750888/2292, Bristol: Institute of Physics Publishing, pp 1–4, (2001)
Ellison, D.C., Ramaty, R.: Astrophys. J. **298**, 400–408 (1985)
Ramaty, R., Murphy, R.J.: Space Sci. Rev. **45**, 213 (1987)
Bombardieri, D.J., et al.: Astrophys. J. **644**, 565–574 (2006)
Saiz, A., et al.: Astrophys. J. **626**, 1131–1137 (2005)
Saiz, A., et al.: Astrophys. J. **672**, 650–658 (2008)
Hua, X.M., et al.: Solar Phys. **107**, 351–383 (1987)
Murphy, R.J., et al.: Astrophys. J. Suppl. **168**, 167–194 (2007)
Hillas, A.M.: Ann. Rev. Astron. Astrophys. **22**, 425–444 (1984)

The Solar Wind and Its Interaction with the Interstellar Medium

John D. Richardson

Abstract The solar wind is a magnetized plasma of ions and electrons which flows outward from the Sun. This chapter begins with a brief history of the discovery of the solar wind. Solar wind properties at 1 AU are discussed as well as how these properties change over a solar cycle. The solar wind evolves as it moves outward through the solar system and interacts with the interstellar neutrals which slow and heat the wind. The solar wind runs into a variety of obstacles as it moves outward, the non-conducting Moon and asteroids, comets and unmagnetized planets where the atmosphere forms a conducting barrier, and magnetized planets with large magnetospheres. Finally the solar wind reaches the interstellar medium, going through the termination shock and then diverting down the tail of the heliosphere.

1 Introduction and Basic Concepts

The solar wind is a stream of ions and electrons which flows outward from the Sun. This chapter discusses briefly the history of the solar wind, the source of the solar wind, the evolution of the solar wind with distance, solar cycle changes in the solar wind, and how the solar wind interacts with obstacles such as planets, comets, the interstellar medium, and other structures within the solar wind. The International Geophysical Year (IGY) in 1957 occurred at the dawn of the space age; the subsequent 50 years of observations have provided a good understanding of many aspects of the solar wind. In the current International Heliophysical Year (IHY), the solar wind interaction with the local interstellar medium (LISM) was studied in situ for the first time. In the next 10 years the Voyagers should leave the solar plasma and make the first direct measurements of the LISM.

Figure 1 shows a schematic diagram of the heliosphere. The color table in the top panel shows the plasma temperature and in the bottom panel shows the H density. The arrows in the top panel show the flow of the solar wind and the interstellar

J. D. Richardson
Center for Space Research, Massachusetts Institute of Technology, Cambridge, USA
e-mail: jdr@space.mit.edu

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_6,

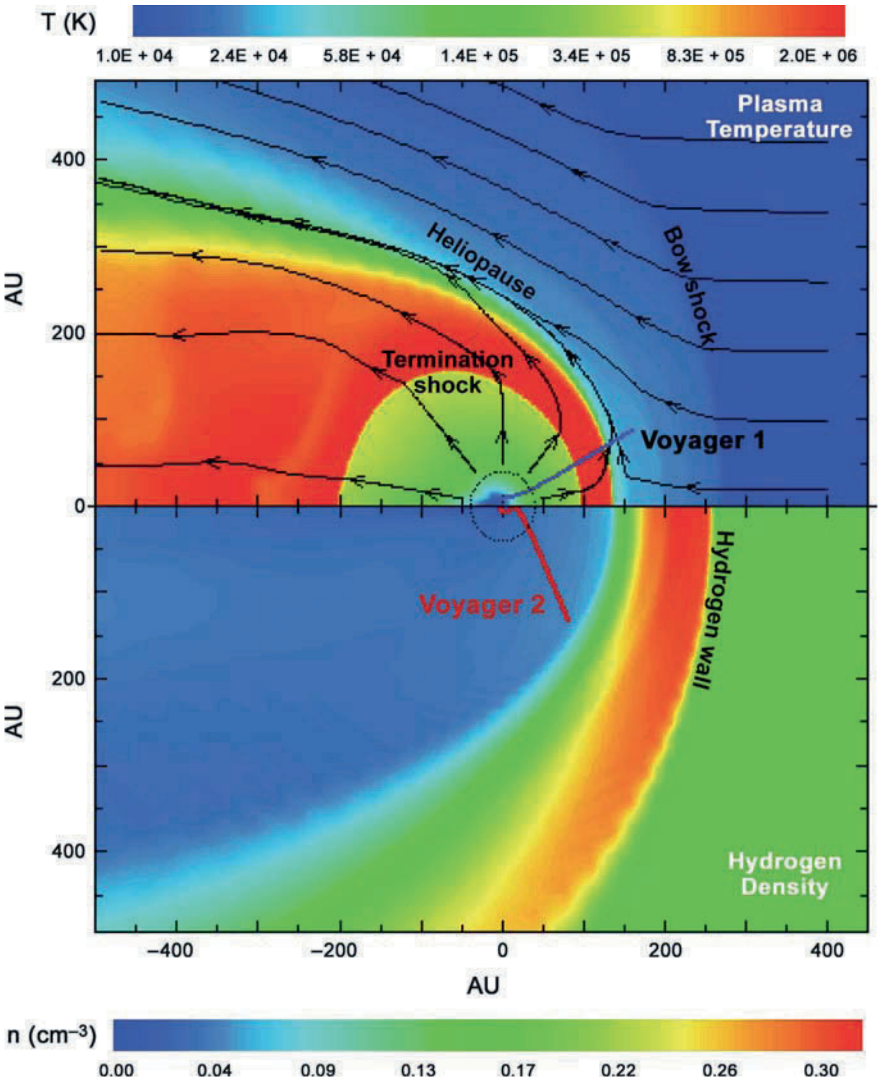


Fig. 1 A model showing the equatorial heliosphere from a plasma (*top*) and neutral (*bottom*) perspective. The *color bar* on the *top panel* shows the plasma temperature and the *arrows* show the plasma flow. The termination shock, heliopause, and bow shock are labeled. The *color bar* on the *bottom panel* shows the H density; the hydrogen wall in front of the heliopause is labeled and the trajectories of the Voyager spacecraft are shown. Figure courtesy of H. Müller

plasma. The solar wind flows radially outward from the Sun until it reaches the termination shock, where it starts to turn and move down the heliotail in the direction of the LISM flow. The region of shocked solar wind plasma is the heliosheath. The LISM is deflected at the bow shock and moves around the heliosphere. The heliopause is the boundary between the solar wind plasma in the heliosphere and the LISM plasma. This chapter discusses recent results on the interaction between the solar wind and the LISM.

A few basic physics principles which are common themes of solar wind studies are discussed here. The first is the concept of a frozen-in magnetic field; simply, this means that the plasma and the magnetic field move together. If an ion or electron moves across a magnetic field line, it generates an electric field $\mathbf{E} = -\mathbf{V} \times \mathbf{B}$ which causes it to gyrate around the magnetic field line instead of moving past it. Since this electric field force is perpendicular to $\mathbf{B}$, plasma is free to stream along the magnetic field lines, but to first order the plasma must stay on the same magnetic field line. This constraint applies only to charged particles; neutral particles are not affected by the magnetic field and can flow across field lines.

The second concept is that of charge exchange and pickup ions. Charge exchange is a major mode of interaction between ions and neutrals. When an ion and neutral collide, the neutral may lose an electron to the ion. The ion becomes a neutral and continues to move in the same direction since it is not bound by the magnetic field. The new ion (the former neutral) is now bound by the magnetic field and is accelerated to the speed of the plasma. It has an initial thermal energy equal to that of the plasma flow energy. The energy for this heating and acceleration comes from the plasma flow energy, so the plasma slows down.

2 History

One of the first clues that the Sun had an out-flowing wind came from observations of solar eclipses. When the solar disk was blocked by the Moon, asymmetric streamers of material were observed. Changes in the Earth's magnetic field provided another piece of evidence; fluctuations were observed in compass needles which were called magnetic storms. These were linked to the Sun when it was noticed that the magnetic storm frequency was related to the sunspot number. In 1859, Carrington (1860) observed a large solar flare and realized it was related to a subsequent magnetic storm. With the invention of the spectroheliograph, many such correlations were observed. Chapman and Ferraro (1930) proposed that the Sun periodically ejects huge clouds of plasma which produce magnetic storms when they reach Earth.

Comets provide another clue about the nature of the solar wind. Dust and neutral gas surrounding the comet are pushed outward from the Sun by radiation pressure. A second tail is often observed at a small angle to the first. Hoffmeister (1943) in Germany and later Biermann (1951) proposed that the Sun emits a steady stream of particles, a "solar corpuscular radiation", which pushes the ions. This flow, now known as the solar wind, is a magnetized plasma; ions from comets are frozen onto the solar wind magnetic field and carried out with the solar wind. Since the solar wind flow is often not radial, a second tail is created.

The Sun's outer layer, or corona, is held down by the Sun's gravity. The corona is very hot, with a temperature of a million degrees, so the thermal particle motion is very fast. The fastest of these particles escape the Sun and form the solar wind. Parker (1958) developed an equation describing the solar wind escape and found that one viable solution was a plasma outflow which became supersonic, flowed continuously outward from the Sun, and asymptotically approached a constant speed. Other solutions are also possible; one is a subsonic solar breeze which decreases in speed with distance. Mankind's entry into the space age resolved the question of which solution was correct. The solar wind was supersonic, but highly variable (Gringauz 1961; Neugebauer and Snyder 1962).

3 Solar Wind Basics

The Sun has a magnetic field and rotates every 27 days; the field lines emanating from the Sun are connected to the solar surface and also rotate. The field is frozen-in to the outward flowing solar wind and thus carried outward with the plasma. The magnetic field lines are stretched by the outflow of the solar wind (Fig. 2) so that the heliospheric current sheet (HCS) forms. The HCS separates field lines from north and south of the Sun's magnetic equator; the field lines in these two (north and south) sectors are in opposite directions, outward in one hemisphere and inward in the other. Where the inward and outward field adjoin the change in field direction drives a current, thus the name current sheet. Since the feet of the field lines are anchored to the Sun and rotate with the Sun, magnetic field lines form spiral pattern, called the Parker spiral. The tilt of the solar dipole causes the HCS to move up and down in heliolatitude, forming a wavy surface which has been compared to a ballerina skirt.

We have had solar wind observations for almost 50 years, with fairly continuous solar wind monitoring since the early 1970s. Table 1 shows average, median,

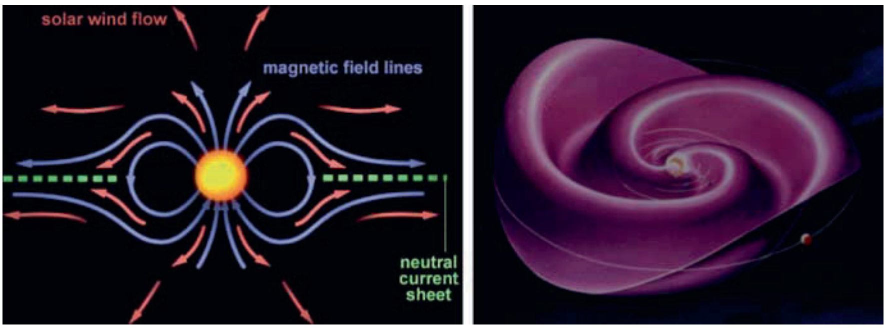


Fig. 2 As the solar wind moves outward, the magnetic field is stretched so that a current sheet forms between the inward and the outward field lines (*left panel*). The rotation of the Sun causes the field lines to spiral outward and the tilt of the solar wind field causes the current sheet to move up and down (*right panel*)

Table 1 Solar wind parameters: the average, median, minimum, and maximum values of the solar wind speed, density, temperature, magnetic field magnitude, and alpha/proton ratio observed by wind

	Average	Median	Minimum	Maximum
Speed (km/s)	440	415	260	2140
Density (cm^{-3})	7.3	5.7	0.1	135
Temperature (K)	83,000	69,000	6,000	3,800,000
B(nT)	6.7	5.9	0.26	72
He^{++}/H^{+} (%)	4.0	4.1	0.1	30

maximum and minimum values measured by the Wind spacecraft near Earth from 1994 to 2007. The extreme variation of the solar wind is demonstrated by Table 1, with changes of almost a factor of ten in speed, a factor of 1,000 in density, 500 in temperature, and 300 in magnetic field strength. Most solar wind ions are protons, but He^{++} is also observed with density ratios which range from near 0 to 30%. For a solar wind speed of 440 km/s, the average Parker (magnetic field) spiral angle at Earth would be about 45°; peaks in the distribution of east/west magnetic field angles are observed at this angle as predicted.

This already fairly complex picture is further complicated both by long-term changes which occur over a solar cycle and short-term transients such as those which drive magnetic storms. The 11-year solar cycle has been observed in sunspot numbers for centuries. By definition, the sunspot number peaks at solar maximum and is lowest at solar minimum. The magnetic field strength varies by about a factor of two over the solar cycle, peaking near solar maximum. The direction of the Sun's magnetic dipole field changes every solar maximum. At solar minimum, the solar magnetic field is roughly dipolar and the current sheet is closely aligned with the solar equator, whereas at solar maximum the field is highly tilted and has strong non-dipolar components. The configuration of the solar wind also changes dramatically. At solar maximum, the solar wind speed is slow (400 km/s) with average densities of 7 cm^{-3} at all heliolatitudes. At solar minimum, coronal holes, regions of open magnetic field lines, cover the solar poles and extend to low heliolatitudes. Solar plasma can flow out along these open field lines at high speeds, 700–800 km/s with densities of 3–4 cm^{-3}. Near the equator the flow is slow and fast like the solar maximum solar wind. Thus the heliolatitudinal gradient in solar wind speed and density is large at solar minimum. Between solar maximum and minimum the solar wind evolves between these two states. Figure 3 illustrates these differences; the left panel shows data from the Ulysses spacecraft at solar minimum with high solar wind speeds above 30° heliolatitude and low speeds nearer the equator. The right panel shows that near solar maximum speeds are much more variable with only small regions of high-speed flow.

The amount of He^{++} also varies over a solar cycle, with more He at solar maximum than at solar minimum (Ogilvie and Hirshberg 1974; Feldman et al. 1978; Aellig et al. 2001). Wind observations show that the He/H ratio varies linearly with speed during solar minimum but is not a function of speed during solar maximum.

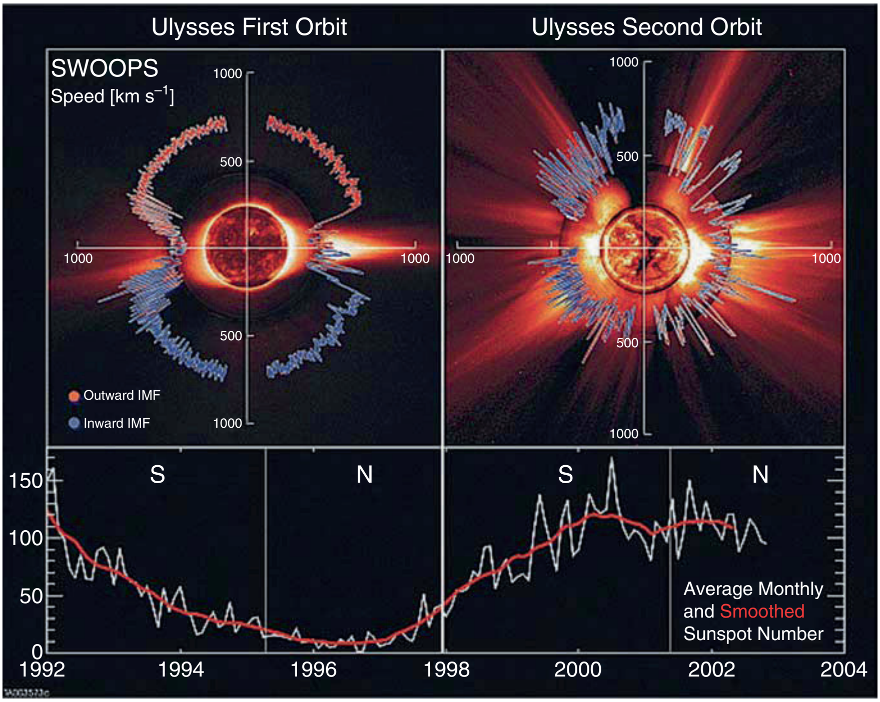


Fig. 3 The solar wind as a function of latitude at solar minimum (left) and solar maximum (right) observed by Ulysses. The bottom panel shows the sunspot numbers. At solar minimum, flow is slow at the equator and fast over the poles. At solar maximum, the flow is slow except for in the narrow regions corresponding to ICMEs. Figure from McComas et al. (2003)

Figure 4 shows histograms of the He/H abundance at solar minimum, solar maximum, and in between for the slow solar wind. At solar minimum the solar wind has relatively low He/H abundances. Between solar maximum and minimum a bi-modal distribution with both low and high He/H values is observed. At solar maximum, most of the solar wind has high He/H ratios but a small peak is still present at very small ratios. These data suggest that a constant low-speed, low He/H solar wind source is always present near the HCS which dominates observations near solar minimum when the HCS tilt is small but makes up a small part of observations when the tilt is large (which results in Wind observing less flow from near the HCS). These two peaks in the He/H ratio suggest that the slow solar wind has two separate sources.

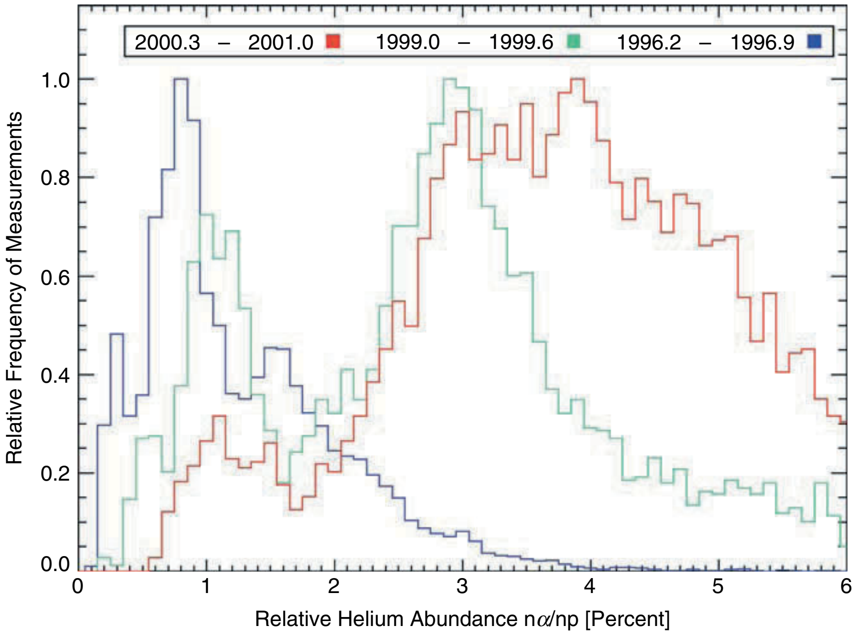


Fig. 4 Ratio of He/H in the slow (<400 km/s) solar wind for three phases of the solar cycle. He/H is smallest near solar minimum and largest near solar maximum. A peak at 1% is observed in all three profiles. Figure courtesy of J. Kasper

4 Radial Evolution of the Solar Wind

The solar wind evolves with distance. Figure 5 shows the solar wind proton speed, density, and thermal temperature variation with distance observed by Voyager 2. Superposed on the V2 speed profile are the speeds observed at 1 AU. Inside 30 AU, the speeds at Earth and those at V2 are similar. The solar wind parameters exhibit substantial variation, but to first order the speed is constant, the density decreases as R^{-2}, and the temperature decreases out to 20–25 AU and then increases. The gradient in speed with heliolatitude at solar minimum causes the deviation of solar wind speeds at Earth and V2 in 1986–1987 and 1995–1998. In 1986–1987, V2 is at a lower average heliolatitude than Earth and observes lower speeds while in 1995–1998 V2 is at higher heliolatitude than Earth and observes much higher speeds. The large speed decrease in 2007 occurred when V2 crossed the termination shock. In addition to the solar cycle changes, speed variations with a 1.3-year period were observed from 1987 to 1998 (Richardson et al. 1994). This variation in speed was observed throughout the heliosphere (Gazis 1995) and has been an occasional feature observed in historic solar wind data (Silverman 1992). A similar period has been observed in convection patterns in the Sun and may be related (Howe et al. 2000).

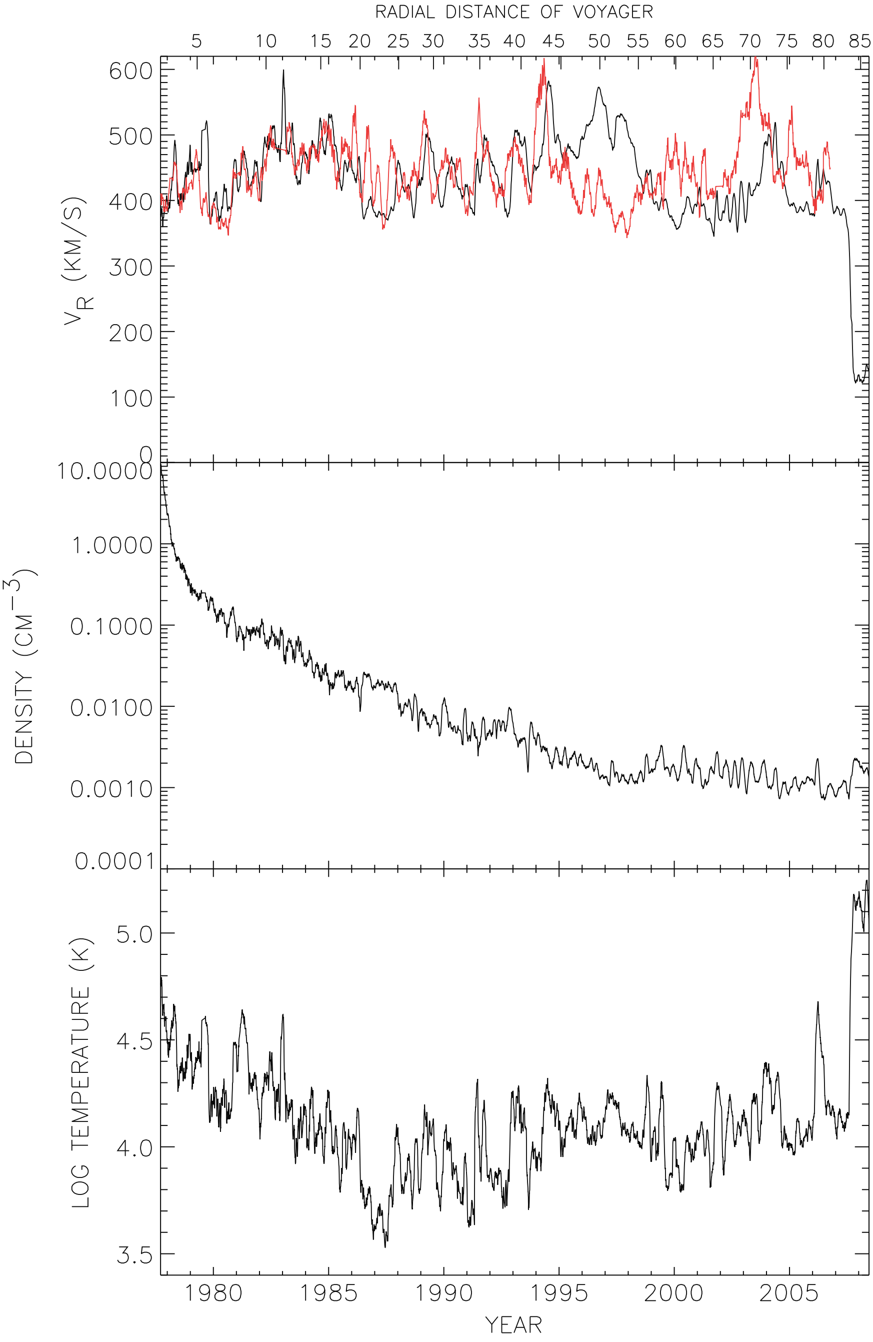


Fig. 5 Fifty-day running averages of the solar wind speed, density, and temperature versus radial distance observed by Voyager 2. The speeds observed at 1 AU are superposed. The density falls as R^{-2} and the temperatures decrease to 20–25 AU, then increase. The TS crossing in 2007 is marked by a large decrease in speed and an increase in density and temperature

The interplanetary coronal mass ejections (ICMEs, see chapter by Gopalswamy) prevalent near and after solar maximum have strong effects on the solar wind evolution. When a series of ICMEs occurs, the first one runs into the ambient solar wind and is slowed. Subsequent ICMEs overtake and merge with the first ICME. By the time these ICMEs reach the outer heliosphere many may combine to form a merged interaction region (MIR), a region of increased speed, density, dynamic pressure, and magnetic field (Burlaga 1995). In 2002–2005, V2 observed a series of these structures, roughly 2 per year lasting 30–60 days, with factor of 10 increases in the dynamic pressure. These large increases in dynamic pressure can drive the termination shock (TS) outward by 3–4 AU (Zank and Müller 2003).

5 Obstacles in the Solar Wind

The solar wind does not move outward unimpeded. In the heliosphere many small obstacles such as the planets, moons, comets, dust, and neutral atoms affect the local solar wind. But most of the solar wind is deflected down the heliotail and eventually merges with the flow of the interstellar medium. We describe these different interactions.

The simplest interaction is with non-conducting obstacles with no atmospheres like the Moon and most asteroids. The magnetic field lines move through these bodies unimpeded but the plasma cannot; it runs into the surface and is absorbed. A plasma void is formed behind the Moon, which results in a pressure gradient force in the solar wind towards the depleted region. But the plasma can only move along magnetic field lines so, as shown in Fig. 6, the plasma streams into the depletion region from two directions. On one side the pressure force results in an increase in the ambient speed and on the other side it results in a decrease.

Conducting obstacles have a more complex interaction since the magnetic field cannot pass through them but must move around them. Planets with magnetic fields have magnetospheres, regions dominated by the planet's magnetic field, which are 1–100 times the size of the planet. The size of the magnetosphere is determined by pressure balance between the solar wind dynamic pressure and the combined pressure of the planet's magnetic field and plasma. Since the solar wind pressure varies with time, so does the magnetosphere size. The solar wind flows around these magnetospheres and the planetary magnetic field forms a long tail, the magnetotail, in the downstream region.

Planets and comets which do not have magnetic fields but which do have atmospheres (Venus, Mars, Titan) also form conducting obstacles. The upper layers of the atmosphere are ionized by solar photons and by charge exchange with the solar wind and form an ionized plasma layer called the ionosphere. The solar wind magnetic field compresses this plasma (since the field cannot move through a plasma) until the plasma pressure equals the solar wind pressure, a boundary called the ionopause at planets and the cometopause at comets. The solar wind must flow around these boundaries of the conducting plasmas.

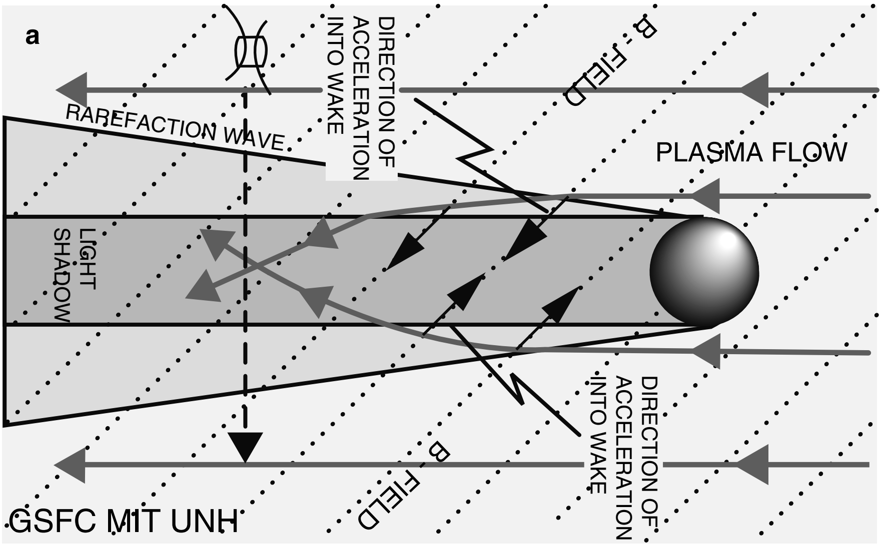


Fig. 6 The solar wind is absorbed by the Moon, leaving a vacuum in the Moon's wake which is filled by plasma streaming along the field lines. From Ogilvie et al. (1996)

The solar wind is supersonic, thus it must go through a shock before it can divert around an obstacle. The fastest wave speed in the solar wind is the fast mode speed, which is the square root of the sum of the squares of the sound speed and the Alfvèn speed. At a fast mode shock, the density, temperature, and magnetic field increase. Two types of fast mode shocks are present in the solar wind; those in front of planets and comets are called reverse shocks since these shocks are standing waves which, in the solar wind frame, move in the reverse direction to the solar wind flow. At a reverse shock, the speed decreases, the density and temperature increase, and the flow direction changes downstream as the solar wind starts to flow around the magnetosphere. The shock upstream of the obstacle is called a bow shock and the region of solar wind downstream of the shock is called the magnetosheath. Sheath regions have highly variable plasma and field parameters.

The interstellar medium is a very large conducting obstacle, a magnetized plasma in which roughly 1/3 of the gas is ionized. Figure 1 shows that the solar wind makes a large bubble of solar material in the LISM. The boundary of the solar wind is determined by where the solar wind dynamic pressure balances the dynamic, thermal, and magnetic pressure of the LISM. This boundary is called the heliopause. Analogous to a magnetopause, inside the heliopause all the plasma is solar and outside all is from the LISM. The solar wind must turn and flow in the LISM direction, so a reverse shock, called the termination shock, forms at which the density, temperature and magnetic field increase and the speed decreases. At the TS the flow starts to turn so as to eventually flow down the heliotail.

We mentioned above that the Sun is much more active at solar maximum; much of this activity is in the form of coronal mass ejections (CMEs) which are huge eruptions of material outward from the Sun. These large plasma structures propagate outward through the solar wind where they are called interplanetary CMEs, or ICMEs. ICMEs are ejected with a wide distribution of speeds; fast ICMES have speeds of several thousand km/s near the Sun. The fast ICMEs run into the solar wind ahead; again, since the flow is supersonic, a shock forms so that the leading solar wind can flow around the faster ICME plasma. In this case the shock moves ahead of the ICME plasma and is a fast forward shock; the density, temperature, field, and speed all increase at the shock. The ambient solar wind turns at the shock so that it can flow around the ICME (in the ICME frame).

A similar situation to the ICME interaction develops between solar minimum and solar maximum when coronal holes and their fast solar wind extend to low latitudes. As the Sun rotates, fast and slow streams are alternately emitted from the Sun. The fast wind catches up to the slow wind and again a shock must form, but in this case two shocks form. A fast shock propagates upstream into the slow wind and a reverse shock propagates into the fast wind. These features are called forward-reverse shock pairs.

The interstellar medium starts to affect the solar wind well before the TS. Although the ions cannot pass through the heliopause, the neutrals in the LISM are not bound by magnetic forces and move into the heliosphere. The helium reaches Earth essentially unchanged from the LISM except for the acceleration of the Sun's gravity. The LISM H is coupled to the protons in the heliosheath via charge exchange and thus is slowed, heated, and diverted in flow direction compared to the He. A similar slowdown occurs upstream of the heliopause nose and forms the region of enhanced H known as the hydrogen wall, shown in Fig. 1.

These neutrals are ionized, the He inside 1 AU and the H outside 5 AU. The pickup ions formed from these neutrals are accelerated up to the solar wind speed and have an initial temperature equal to the solar wind energy, about 1 keV. Figure 7 shows the density of the various species in the vicinity of the heliosphere. The LISM ions increase in density at the bow shock and do not cross the heliopause (HP). The solar wind ions decrease in density as R^{-2}, increase at the TS, and do not cross the HP. The LISM neutrals enter the heliosphere and are slowly lost via charge exchange. The neutral LISM is the densest component in the heliosphere outside about 10 AU. The pickup ions are the ionized LISM neutrals. Their density decreases with distance, then jumps at the TS. By the TS, almost 20% of the ions are pickup ions. Since the pickup ions are hot, they dominate the solar wind thermal pressure outside about 30 AU. Thus even before the TS the LISM has significant impact on the inner heliosphere through the neutral component.

Figure 5 compares the solar wind speeds observed at 1 AU with those observed by V2. Outside 30–40 AU the V2 speeds are systematically lower than those at 1 AU. The speed decrease at V2 is a result of the ionization of the LISM neutrals and their subsequent acceleration to the solar wind speed at the expense of the flow energy in the thermal solar wind. The ratio between the solar wind slowdown and the pickup ion density is $N_{PI}/N_T = 7/6\Delta V/V$ (Richardson et al. 1995; Lee 1995), where ΔV

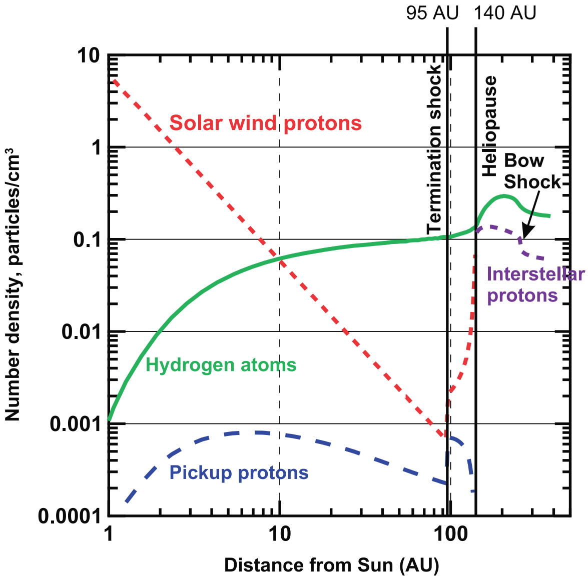


Fig. 7 The density of the plasma and neutral components of the solar wind and LISM. Figure Courtesy of V. Izmodenov

is the solar wind speed decrease, V is the speed of the solar wind, N_{PI} is the pickup ion density, and N_T is the total ion density. Observations show that the slowdown is about 17% near the TS (Richardson et al. 2008a), so roughly 30% of the solar wind flow energy has been converted to particle energy before effects of the TS are observed.

The increase in temperature outside 20–30 AU in Fig. 5 is also caused by the pickup ions. They are formed with a ring distribution (all their energy is perpendicular to the magnetic field) which is unstable. This instability generates magnetic fluctuations which heat the thermal ions; only about 5% of the energy generated by relaxing the ring distribution is needed to provide the observed solar wind heating (Smith et al. 2006; Isenberg et al. 2005).

We mentioned above that the TS location results from a balance between the solar wind and LISM pressures. But until 2004, we did not know the LISM pressure and so we did not know the TS location. The first sign that Voyager was approaching the TS was the observation of particles streaming along magnetic field lines from the TS to V1 starting at 85 AU. Since the TS is not circular, some parts of a field line may cross the TS while others are in the solar wind. The particles heated at the shock can stream along the magnetic field lines into the heliosphere; this region of streaming particles is called the termination foreshock and is analogous to foreshock regions observed upstream of planetary bow shocks. The particle intensities vary as the connection of the field lines to the TS changes. These particles were observed at V1 roughly 2 1/2 years before the TS crossing.

In December 2004, V1 crossed the TS at 94 AU (Decker et al. 2005; Burlaga et al. 2005; Stone et al. 2005); this event revealed the scale size of the heliosphere. Unfortunately the plasma instrument on V1 is not working so the solar wind ions were not observed. Also, the TS crossing occurred in a data gap. V2 started observing the foreshock particles at 75 AU, 10 AU closer than V1, and these particles were streaming in the opposite direction from those at V1 (Decker et al. 2006). These observations support the hypothesis that the TS is blunt or flattened at the nose. V1 and V2 are on opposite sides of the nose, so particles coming to V1 and V2 from the nose arrive from opposite directions (Jokipii et al. 2004). Thus the foreshock particles told us about the shape of the heliosphere.

The difference of the foreshock boundary in the V1 and V2 directions suggested that the heliosphere might be asymmetric, with the TS and HP closer in the south (where V2 is) than the north (where V1 is). V2 crossed the TS in August 2007 at 84 AU, 10 AU closer than at V1 (Decker et al. 2008; Burlaga et al. 2008; Stone et al. 2008; Richardson et al. 2008b). The location of the TS depends on the solar wind dynamic pressure which varies with time. Over a solar cycle average solar wind dynamic pressure varies by a factor of 2 (Lazarus and McNutt 1990), causing the TS to move in and out over a range of 10–14 AU (Karmesin et al. 1995; Wang and Belcher 1998). If we account for the pressure change in the solar wind from 2004 to 2007, the TS asymmetry is 7–8 AU with the TS closer in the V2 than the V1 direction. One possible explanation for this asymmetry is shown in Fig. 8; if the solar wind magnetic field is tilted from the direction of the LISM flow then the field

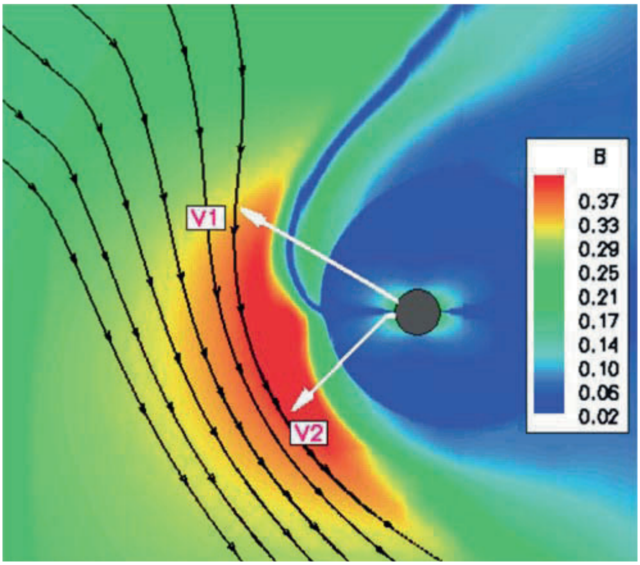


Fig. 8 A model showing the formation of an asymmetric heliosphere caused by LISM field lines draping around and compressing the southern part of the heliosphere. Figure courtesy of M. Opher

is more compressed (and thus stronger) on one side of the heliosphere, leading to an asymmetry (Linde et al. 1998; Ratkiewicz et al. 1998; Opher et al. 2006, Opher et al. 2007; Pogorelov et al. 2004, Pogorelov and Zank 2006).

The V2 TS crossing provided the first plasma data near the TS and the first time we had data at the TS itself. The TS differed from other shocks in the solar system in a few significant ways. The decrease in speed began 0.7 AU upstream of the TS; the solar wind speed dropped from 400 to 300 km/s at the TS in three steps. At other shocks the speed decrease has all occurred at the shock. The shock was relatively weak; the density and field jumps were a factor of 2 compared to a factor of 4 at planetary bow shocks. The thermal plasma was heated relatively little, to 100,000 K compared to 2,000,000 K in planetary magnetosheaths. We believe that 80% the flow energy goes into the pickup ions and only 20% to the thermal ions (Richardson et al. 2008b), as a few authors predicted before V2 crossed the TS (Zank et al. 1996; Gloeckler et al. 2005).

Thus in the past few years we have revolutionized our understanding of the heliosphere's interaction with our interstellar environment. We know the size of the heliosphere is about 90 AU, that the TS is not round but blunt, and that the heliosphere is asymmetric. As the Voyagers continue moving outward they will eventually enter the LISM and measure it directly. The location of the HP will not be known until we cross it, but models suggest the width of the heliosheath is 30–40 AU so that the Voyager could cross it in 2014–2018. One indication the HP is approaching is that the heliosheath flow will rotate parallel to this HP boundary. Another indication may be an increase in the magnetic field as it piles up against this boundary (Cranfill 1971). These spacecraft have sufficient power to continue several years past 2020, so there is a good chance they will make mankind's first measurements of the plasma outside our solar system.

Acknowledgements This work was supported under NASA contract 959203 from JPL to MIT (Voyager), by NNX08AE49G (Wind), and NASA OPR grant NNX08AC04G.

References

Aellig, M.R., Lazarus, A.J., Steinberg, J.T.: The solar wind helium abundance: variation with wind speed and the solar cycle. Geophys. Res. Lett. **28**, 2767–2770 (2001)

Biermann, L.: Kometenschweife und sol are Korpuskularstrahlung. Z. Astrophys. **29**, 274–286 (1951)

Burlaga, L.F.: Interplanetary Magnetohydrodynamics. Oxford University Press, New York (1995)

Burlaga, L.F., Ness, N.F., Acuna, M.H., Lepping, R.P., Connerney, J.E.P., Stone, E.C., McDonald, F.B.: Crossing the termination shock into the heliosheath. Magn. Fields Sci. **309**, 2027–2029 (2005)

Burlaga, L.F., Ness, N.F., Acuna, M.H., Lepping, R.P., Connerney, J.E.P., Richardson J.D.: Observations of magnetic fields at the termination shock by Voyager 2. Nature. **454**, 75–77 (2008)

Carrington, R.C.: Description of a singular appearance seen in the Sun on September 1, 1859. Mon. Not. R. Astron. Soc. **20**, 13–15 (1860)

Chapman, S., Ferraro, V.C.A.: A new theory of magnetic storms. Nature. **126**, 129–130 (1930)

Cranfill, C.W.: Flow problems in astrophysical systems. Thesis (Ph.D.), Department of Physics, University of California, San Diego, California (1971)

Decker, R.B., Krimigis, S.M., Roelof, E.C., Hill, M.E., Armstrong, T.P., Gloeckler, G., Hamilton, D.C., Lanzerotti, L.J.: Voyager 1 in the foreshock, termination shock, and heliosheath. Science. **309**, 2020–2024 (2005). doi: 10.1126/science.1117569

Decker, R.B., Krimigis, S.M., Roelof, E.C., Hill, M.E.: Low-energy ions near the termination shock. In: Physics of the inner heliosheath: voyager observations, theory, and future prospects, AIP conference proceedings 258, 73–78 (2006)

Decker, R.B., et al.: Shock that terminates the solar wind is mediated by non-thermal ions. Nature. **454**, 67–70 (2008)

Feldman, W.C., Asbridge, J.R., Bame, S.J., Gosling, J.T.: Long-term variations of selected solar wind properties: IMP 6, 7, and 8 results. J. Geophys. Res. **83**, 2177 (1978)

Gazis, P.R.: Limits on deceleration and asymmetry of solar wind speed. Geophys. Res. Lett. **22**, 2441 (1995)

Gloeckler, G., Fisk, L.A., Lanzerotti, L.J.: Acceleration of solar wind and pickup ions by shocks. In: Solar wind 11/SOHO 16 programme and abstract book (pdf file), European Space Agency, the Netherlands, 52 (2005)

Gringauz, K.I.: Some results of experiments in interplanetary space by means of charged particle traps on Soviet space probes. Space Res. **2**, 539–553 (1961)

Hoffmeister, C.: Physikalisch Untersuchungen auf Kometen. I. Die Beziehungen des primaren Schweifstrahl zum Radiusvektor. Z. Astrophys. **22**, 265–285 (1943)

Howe, R., et al.: Dynamic variations at the base of the solar convection zone. Science. **287** 2456–2460 (2000)

Isenberg, P.A., Smith, C.W., Matthaeus, W.H., Richardson, J.D.: Turbulent heating of the distant solar wind by interstellar pickup protons with a variable solar wind speed. In: Fleck B., Zurbuchen, T.H. (eds.) Proceedings of solar wind 11: connecting Sun and Heliosphere, ESA SP-592, pp. 347–350. European Space Agency, The Netherlands (2005)

Jokipii, J.R., Giacalone, J., Kota, J.: Transverse streaming anisotropies of charged particles accelerated at the solar wind termination shock. Astrophys. J. **611**, L141–L144 (2004)

Karmesin, S.R., Liewer, P.C., Brackbill, J.U.: Motion of the termination shock in response to an 11 year variation in the solar wind. Geophys. Res. Lett. **22**, 1153–1156 (1995)

Lazarus, A.J. McNutt, R.L. Jr.: Plasma observations in the distant heliosphere: a view from Voyager. In: Grzedzielski, S., Page, D.E. (eds.) Physics of the outer heliosphere, pp. 229–234. Pergamon Press, New York (1990)

Lee, M.A.: Effects of cosmic rays and interstellar gas on the dynamics of the solar wind. In: Jokipii, J.R., Sonnett, C.P., Giampapa M.S. (eds.) Cosmic winds and the heliosphere, p. 857. University of Arizona Press, Tucson (1995)

Linde, T.J., Gombosi, T.I., Roe, P.L., Powell, K.G., DeZeeuw, D.L.: Heliosphere in the magnetized local interstellar medium: results of a three-dimensional MHD simulation. J. Geophys. Res. **103**, 1889–1904 (1998)

McComas, D.J., Elliott, H.A., Schwadron, N.A., Gosling, J.T., Skoug, R.M., Goldstein, B.E.: The three-dimensional solar wind around solar maximum. Geophys. Res. Lett. **30**, 1517–1520 (2003). doi:10.1029/2003GL017136

Neugebauer, M., Snyder, C.W.: Solar plasma experiment. Science. **138**, 1095–1097 (1962)

Ogilvie, K.W. Hirshberg, J.: The solar cycle variation of the solar wind helium abundance. J. Geophys. Res. **79**, 4959 (1974)

Ogilvie, K.W., Steinberg, J.T., Fitzenreiter, R.J., Owen, C.J., Lazarus, A.J., Farrell, W.M., Torbert, R.B.: Observations of the lunar plasma wake from the WIND spacecraft on December 27, 1994. Geophys. Res. Lett. **23**, 1255–1258 (1996)

Opher, M., Stone, E.C., Liewer P.C.: The effects of a local interstellar magnetic field on Voyager 1 and 2 observations. Astrophys. J. **640**, L71–L74 (2006)

Opher, M., Stone, E.C., Gombosi, T.I.: The orientation of the local interstellar magnetic field. Science. **316**, 875–878 (2007). doi: 10.1126/science.1139480

Parker, E.N.: Dynamics of the interplanetary gas and magnetic fields. Astrophys. J. **128**, 664–676 (1958)

Pogorelov, N.V., Zank, G.P., Ogino, T.: Three-dimensional features of the outer heliosphere due to coupling between the interstellar and interplanetary magnetic fields. I. Magnetohydrodynamic model: interstellar perspective. Astrophys. J. **614**, 1007–1021 (2004)

Pogorelov, N.V., Zank, G.P.: The direction of the neutral hydrogen velocity in the inner heliosphere as a possible interstellar magnetic field compass field compass. Astrophys. J. **636**, L161–L164 (2006)

Ratkiewicz, R., Barnes, A., Molvik, G.A., Spreiter, J.R., Stahara, S.S., Vinokur, M., Venkateswaran, S.: Effect of varying strength and orientation of local interstellar magnetic field on configuration of exterior heliosphere: 3D MHD simulations. Astron. Astrophys. **335**, 363–369 (1998)

Richardson, J.D., Paularena, K.I., Belcher, J.W., Lazarus, A.J.: Solar wind oscillations with a 1.3 year period. Geophys. Res. Lett. **21**, 1559–1560 (1994)

Richardson, J.D., Paularena, K.I., Lazarus, A.J., Belcher, J.W.: Evidence for a solar wind slowdown in the outer heliosphere? Geophys. Res. Lett. **22**, 1469–1472 (1995)

Richardson, J.D., Liu, Y., Wang, C., McComas, D.J.: Determining the LIC H density from the solar wind slowdown. Astron. Astrophys. **491**, 1–5 (2008a)

Richardson, J.D., Kasper, J.C., Wang, C., Belcher, J.W., Lazarus, A.J.: (2008b) Termination shock decelerates upstream solar wind but heliosheath plasma is cool. Nature. **454**, 63–66 (2008b)

Silverman, S.M.: Secular variation of the aurora for the past 500 years. Rev. Geophys. **30**, 333–351 (1992)

Smith, C.W., Isenberg, P.A., Matthaeus, W.H., Richardson, J.D.: Turbulent Heating of the solar wind by newborn interstellar pickup protons. Astrophys J. **638**, 508–517 (2006)

Stone, E.C., Cummings, A.C., McDonald, F.B., Heikkila, B., Lal, N., Webber, W.R.: Voyager 1 explores the termination shock region and the heliosheath beyond. Science. **309**, 2017–2020 (2005)

Stone, E.C., et al.: Voyager 2 finds an asymmetric termination shock and explores the heliosheath beyond. Nature. **454**, 71–74 (2008)

Wang, C., Belcher, J.W.: The heliospheric boundary response to large scale solar wind fluctuations: a gasdynamic model with pickup ions. J. Geophys. Res. **104**, 549–556 (1998)

Zank, G.P., Müller, H.-R.: The dynamical heliosphere. J. Geophys. Res. **108**, (2003). doi: 10.1029/2002JA009689

Zank, G.P., Pauls, H.L., Cairns, I.H., Webb, G.M.: Interstellar pick-up ions and perpendicular shocks: implications for the termination shock and interplanetary shocks. J. Geophys. Res. **101**, 457–577 (1996)

Reconnection Process in the Sun and Heliosphere

A.C. Das

Abstract Solar flare, coronal mass ejection and many other interesting plasma and magnetic field structures in the heliosphere are believed to be generated by a powerful plasma process widely known as reconnection of magnetic field lines. The basic understanding of this reconnection process is described by considering a simple model suggested by Dungey that contains a kind of consistent flow of plasma around an x-type neutral point. There are, however, several comprehensive MHD models, some of which are also discussed in this article. Spontaneous reconnection based on tearing mode instability is described very briefly for completeness. The crucial role played by magnetic reconnection in violent energy conversion occurring in solar flare, coronal mass ejection and other related phenomena like hard x-ray and soft x-ray emissions is highlighted. Several convincing observational evidences that support reconnection model are presented. Numerical simulation is seen to be an essential part of this study that enhances our understanding of the evolution of plasmoid and multiple shocks due to reconnection of field lines. Some challenges are mentioned towards the end of the presentation.

1 Introduction

We are all aware of the magnetosphere of our planet, Earth. It is a cavity formed as a result of the interaction between the solar wind and the magnetic field of Earth. Similarly the heliosphere is like a huge bubble in space generated by the interaction of the solar wind and the interstellar medium. It may be called the magnetosphere of the Sun. The solar wind travels at supersonic speed within the solar system and it becomes subsonic at a point called the termination shock. The point where the interstellar medium and the solar wind pressure balance is called the heliopause. A bow

A.C. Das
Physical Research Laboratory, Ahmedabad 380 009, India
e-mail: anathcdas@gmail.com, acd@prl.res.in

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_7,

shock is formed where the interstellar medium moving in the opposite direction becomes subsonic as it collides with the heliopause.

The solar wind, solar flares and coronal mass ejections send material, mostly ionized particles, and magnetic fields out into the heliosphere. This flow of mass and magnetic fields creates and fills the heliopshere, which is a volume inside the local interstellar medium and contains the solar system. Another important structure is the heliospheric current sheet, which looks like a ripple in the heliosphere, created by the Sun's rotating magnetic field. It extends throughout the heliosphere and is considered to be the largest structure in the solar system. Many interesting phenomena are seen to be associated with this current sheet.

Solar flares, coronal mass ejections and the closed magnetic loop structures in the heliospheric current sheet are believed to be generated by a very powerful plasma process known as reconnection of magnetic field lines. Therefore, we would like to concentrate first on the basic understanding of this reconnection process. This idea actually started with Giovanelli's suggestion of the importance of null points based on the observation of solar flares. Later this has been developed by Dungey (1961) in the physics of magnetosphere with great success. The role of magnetic reconnection in solar flares has been described quite well in a review paper by Lakhina (2000). Following Dungey's concept, we will describe later on the general development of the process of reconnection. It may be useful to introduce first some basic understanding of plasma motions and magnetic field structures and their evolution in the solar wind plasma. The general motion of plasma and the behaviour of the magnetic field are governed by the following equations: Equation of the motion of plasma

$$\rho \frac{d\upsilon}{dt} = -\nabla p + \frac{1}{c}\ J \wedge B, \tag{1}$$

Maxwell's equation

$$\text{c Curl B} = 4\,\pi\,\text{J}, \tag{2}$$

$$\frac{\partial B}{\partial t} = -c\, Curl\ \ E, \tag{3}$$

and Ohm's law

$$J = \sigma \left(E + \frac{1}{c}\ \upsilon \wedge B \right) \tag{4}$$

Using (2), (3) and (4), it is readily obtained the following equation:

$$\frac{\partial B}{\partial t} = \xi \nabla^2 B + \nabla \wedge (\upsilon \wedge B)\,, \tag{5}$$

where $\xi = \frac{c^2}{4\pi\sigma}$, σ represents conductivity and ξ is the magnetic diffusivity. In the above equations ρ, υ, p, J, B, E, c denote the charge density, plasma velocity, scalar pressure, current density, magnetic (field) induction, electric field, and the velocity of light respectively.

The equation of motion (1) represents how the plasma is accelerated because of the gradient of pressure and the magnetic force while the induction equation (5) describes the changes in the magnetic field due to plasma motion and magnetic diffusion. Nonlinear terms indicate how the plasma and the magnetic field are coupled. In absence of plasma velocity, the ordinary electric current is governed by the Ohm's law $J = \sigma E$, where the magnetic field is secondary and can be calculated from Ampere's law c Curl B $= 4\pi J$. In the magnetohydrodynamic (MHD) theory, velocity v and the magnetic field B are primary and J and E can be calculated from

$$Curl\ B = \frac{4\pi}{c}\ J$$

and

$$E = \frac{J}{\sigma} - \frac{\upsilon \wedge B}{c}. \tag{6}$$

Now for large σ, J/σ can be neglected and the electric field is then driven by the velocity and magnetic field. Ratio of the 2nd term to the first term on the right hand side of (5) defines the magnetic Reynolds number given by

$$R_m = \frac{\upsilon B}{L} / \left(\xi \frac{B}{L^2}\right) = \frac{\upsilon L}{\xi}. \tag{7}$$

R_m is extremely large ($\sim 10^{10}$) in the solar atmosphere and therefore the magnetic field lines are frozen in. However in a thin sheet of current or in a filament in the solar atmosphere, where L is very small (in the order of a km or so), the R_m could be very small (~ 1) so the magnetic field is not frozen in and can slip through the plasma while the magnetic energy is converted into heat in the time scale $T \sim L^2/\xi$.

1.1 Basic Reconnection Process

The basic process of reconnection depends on (1) the topology of the magnetic field, and (2) the motion of plasma near the neutral point. Dungey (1958) developed this model of reconnection to describe the interaction of the solar wind with the magnetosphere of Earth. A simple model for the magnetic field and the plasma motion around the neutral point is shown in Fig. 1.

Basically the magnetic field lines are anti-parallel to each other such that there is just one neutral point N with limiting field lines called separatrix. Two limiting lines are going into N while two are coming out. This is the magnetic topology. Now let us consider the plasma flow around such a neutral point. This holds an important key for the process of reconnection. It is very difficult to describe the plasma motion in the presence of a pressure tensor and different conductivities. However, in the absence of pressure and with an infinite conductivity, the Ohm's law that can provide velocity is given by $E + (\upsilon \wedge \mathbf{B})/c = 0$. The above equation also defines an electric

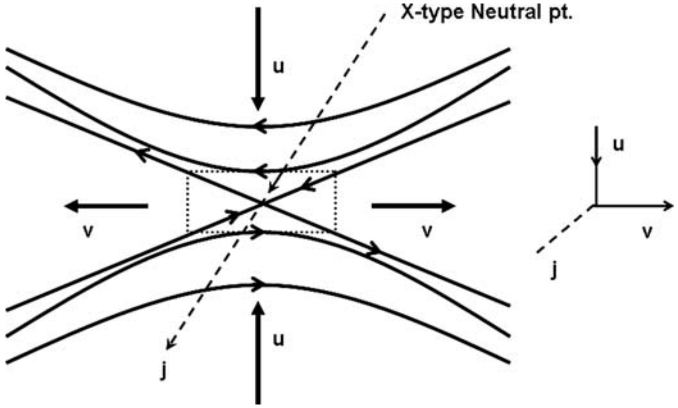


Fig. 1 Magnetic field topology and flow pattern around the neutral point (see Dungey 1961)

field. Suppose that the field lines are moving from both sides towards the neutral point as shown in Fig. 1. According to frozen-in-field approximation, they remain as unbroken field lines and still they can go to N or can go out from N. The direction of electric field associated with the motion of field lines is the same as that in above, below or at the neutral point. This kind of motion tends to increase current density at N. However, this is not reconnection.

The reconnection is the consequence of the breakdown of frozen-in-field approximation, which may be caused because of a high current density. It occurs when there is resistivity or anomalous resistivity. In this case the scenario becomes different. If one allows that these field lines do not remain as the same field line as before, then at some instant of time a pair of inflowing field lines actually becomes the limiting field lines and immediately after that they form an outflowing pair of field lines. This permits the limiting field lines to get cut at the neutral point and then reconnect to form a different set of field lines. This is possible because the frozen-in-field approximation is violated. This process is described as the reconnection of the field lines as they pass through N. This is the pictorial description of the reconnection of field lines.

We can probably proceed further with the MHD theory to describe the above process in the solar atmosphere. We have already seen that in many regions the Reynolds number R_m is usually very large and therefore, the diffusion of the magnetic field is negligible and as a result the plasma motion can be described by the equation, $E + (\upsilon \wedge B)/c = 0$, where E, B, υ are already defined earlier. The nature of the plasma flow can be virtually obtained from this equation. Hence the transport of magnetic flux near the separatrix is always associated with the plasma in such a region. However, in a thin current sheet or in a filament, Reynolds number becomes small and the frozen-in-field approximation is not valid. In such a region, the above equation is not valid. Dungey pointed out that in such a region, a large current is generated without being opposed by an electromagnetic force. The equation governing the plasma motion and the current are Maxwell's equation, the equation of

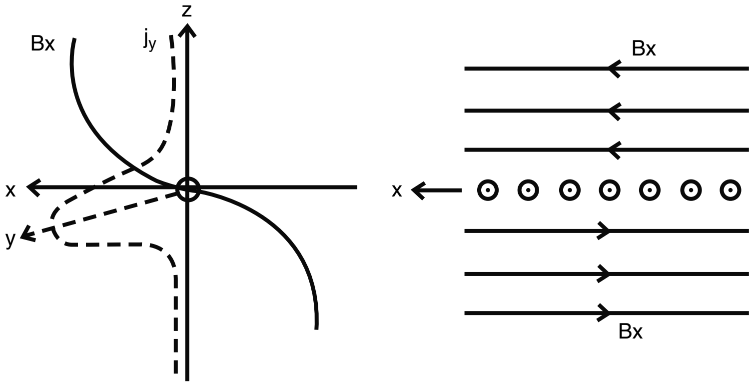


Fig. 2 Magnetic field strength and current density across the current sheet obtained from the diffusion equation. The current flows along y-direction. (Reproduced from Das 2004, copyright 2004)

motion and the Ohm's law given by

$$E + \frac{1}{c}\ \upsilon \wedge B = J/\sigma. \tag{8}$$

In thin region, near the neutral point, the diffusion term is substantial and therefore it may be useful to examine the evolution of the current and the magnetic field at the separatrix without the effect of a plasma flow. The magnetic induction (5) becomes a simple diffusion equation represented by

$$\frac{\partial B}{\partial t} = \xi\, \nabla^2 B. \tag{9}$$

In one dimension, $\frac{\partial B_x}{\partial t} = \xi\ \frac{\partial^2 B_x}{\partial z^2}$,where B_x is the magnetic field along x-direction and z is the vertical direction as shown in Fig. 2.

The solution can be obtained in terms of the error integral showing a large current flowing along the y-direction and the magnetic field lines are in the opposite direction around $z = 0$. In fact B_x varies with z and changes sign at $z = 0$. The magnetic flux diffuses from above as well as from below and gets dumped at the separatrix feeding the current. This reduces the gradient in the magnetic field and thus reducing the diffusion rate and slowing down the whole process. So the process becomes unproductive in a steady state. In order to maintain this process in a steady state, it becomes necessary to introduce a plasma flow from both sides for transporting the magnetic flux towards the neutral point at a rate at which they annihilate. The large magnetic flux through the diffusion region can maintain large current in the neutral sheet. However, this is also unphysical unless there is an outflow for these fluxes. The pictorial model suggested by Dungey (1961) contains a kind of a consistent flow of plasma around an x-type neutral point as shown in Fig. 1.

The existence of the diffusion region due to the finite resistivity is an essential requirement of the model to support the large current density. The resistivity has the classical value when it is entirely due to particle scattering by Coulomb collisions and it is shown how the Reynolds number can become appropriate in a thin sheet or a filament in the solar atmosphere for violating the frozen-in-field condition. In the magnetosphere, Coulomb collisions are extremely small and even in the neutral sheet region R_m cannot be made small. In such a plasma, an anomalous resistivity, which could be a few orders of magnitude larger, can occur due to scattering by many microscopic processes including plasma instabilities, Landau damping and so on.

This can also be understood in terms of electric field scenario suggested by Dungey (1958). Near the neutral point, some electric field is needed to drive the current, which is given by $E = \eta J$, where η is the resistivity, and J is in the y-direction and therefore the electric field also is in the same direction. Far away from the neutral point, the frozen-in-field approximation is valid and therefore there exists an electric field both above and below the neutral point bringing a flux of magnetic energy, equal to $c\ (E \wedge B)/4\pi$, towards the reversal region from both sides. However, this process of transport falls on the region where the magnetic field is zero and therefore the diffusion discussed earlier must play the role in bringing the magnetic energy into the neutral plane. This is the simplest picture of reconnection under the assumption that there exists resistivity in the region and the sheet is very thin. There exist, however, several comprehensive models, some of which will be discussed in the next lecture.

2 Reconnection Models

2.1 Sweet–Parker Model

It is assumed that the magnetic field lines are anti-parallel with a neutral sheet rather than a neutral line in between the field lines. The configuration of the magnetic field lines and the direction of flow are shown in Fig. 3.

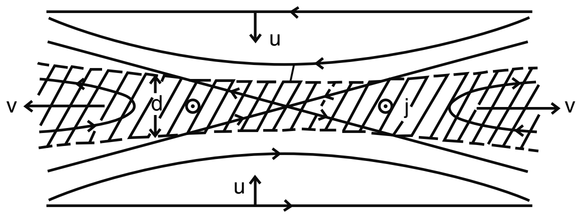


Fig. 3 Sweet–Parker model of reconnection (reproduced from Das 2004, copyright 2004)

Plasma flows with a normal velocity u from both sides into a diffusion region of the current sheet. The length of the diffusion is l while the width is denoted by d. It is further assumed that the plasma is flowing out with a velocity υ. A steady-state reconnection is maintained by the balance between the plasma flowing into the diffusion region and the equal amount of plasma out flowing in the current sheet. From conservation of mass, we get

$$ul = \upsilon d, \tag{10}$$

(a consequence of $\nabla \cdot \upsilon = 0$).

From the momentum balance equation, it is seen that $p = p_o + B^2/8\pi$, where p is the plasma pressure on the central plane where the magnetic field is assumed to be negligible or almost zero, and p_o is the pressure outside the diffusion region where the magnetic field is B. The pressure difference $(p - p_o)$ drives the plasma flow out of the diffusion region with a velocity v determined by

$$p - p_o = B^2/8\pi = \frac{1}{2}\rho\upsilon^2, \tag{11}$$

or,

$$\upsilon = \left(\frac{B^2}{4\pi\rho}\right)^{\frac{1}{2}} = \upsilon_A, \tag{12}$$

where, υ_A is the Alfven velocity.

Now the magnetic field is carried by the fluid towards the diffusion region which is controlled by the magnetic induction equation (5). The balance between the convecting flux and the flux annihilation due to the diffusion within it determines the width of the diffusion region. In a steady state, the velocity u is then given by

$$u = \xi/d \tag{13}$$

where $\xi = \eta c^2/4\pi$ and η is the resistivity. Then from (10), (12) and (13), we get $u = \upsilon\frac{d}{l} = \upsilon_A\,\xi/lu$, or, $u^2 = \upsilon_A\frac{\eta c^2}{4\pi l} = \upsilon_A^2\frac{\eta c^2}{4\pi l\upsilon_A} = \upsilon_A^2\frac{1}{R_m}$, where the magnetic Reynolds number $\mathrm{R_m}$ is given by

$$R_m = \frac{\upsilon_A l}{\eta c^2/4\pi}. \tag{14}$$

Thus the reconnection rate is given by

$$u = \frac{1}{\sqrt{R_m}}\upsilon_A. \tag{15}$$

$\mathrm{R_m}$ is a function of the Alfven velocity which is in this case is the outflow velocity from the diffusion region as the result of reconnection and l is the diffusion length which in this particular case is the overall system length. If there were no reconnection, the ordinary diffusion rate would have been given by $\upsilon_d = \xi/l = \upsilon_A/R_m$, where $\xi = \eta c^2/4\pi$. Thus it is seen that the reconnection rate is larger than the ordinary diffusion by a factor of $\sqrt{R_m}$ and since R_m is very large in interplanetary space, the reconnection is enhanced by a large factor. However, it is seen that the rate of reconnection or the inflow velocity is still very small compared to the Alfven velocity. Thus the Sweet-Parker model (Sweet 1957; Parker 1957) remains inefficient in producing a fast reconnection and thus the conversion of magnetic energy to plasma energy takes place at a slow rate.

2.2 Petschek Model

Sweet–Parker model cannot produce a large rate of reconnection because the length of diffusion region l, which is also the scale length of the system can be very large compared to the width of the region. This can be verified easily from the relation $u = (d/l)\upsilon_A$, which is much less than one and therefore the low rate of reconnection. Petschek realized this and suggested that if the size of diffusion can be limited to a smaller dimension, then the rate can be enhanced substantially because $u > \upsilon_A$ if $d > l$. Petschek (1964) pointed out that in the MHD flow in the outer region, it is possible that two standing MHD wave fronts can be maintained and these fronts are slow shocks. The diffusion region can be matched to a region of standing waves which deflect and accelerate the incoming plasma into two jets sustained between the shocks. The diffusion region is still important in the sense that the actual reconnection takes place there. A schematic view of the magnetic field and plasma flow configuration is shown in Fig. 4. A small diffusion region at the center around the neutral line still exists between the two standing waves.

Suppose α is the half angle of the exit flow or the angle of the slow shock such that it remains stationary in the flow. Now the plasma flow normal to the plane of the shock must be equal to the Alfven shock speed normal to the shock front in the frame of the out-flowing plasma shown in Fig. 4. The laws of flux conservation give

$$uB = \upsilon B_o, \tag{15}$$

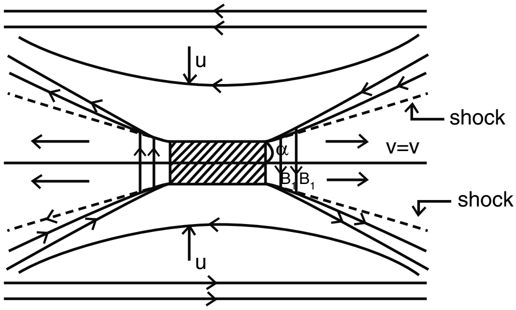


Fig. 4 Petschek model of reconnection (reproduced from Das 2004, copyright 2004)

where the magnetic field B is associated with the inflow, while B_o denotes the field in the outflow. The conservation of tangential stress gives

$$\rho u \upsilon = \frac{BB_o}{4\pi}. \tag{16}$$

Combining the above two equations, we get $\upsilon = \upsilon_A$, which is parallel to the x-axis shown in Fig. 4, similar to Sweet–Parker model.

The flow speed normal to the shock is denoted by $\upsilon_n$and is given by

$$\upsilon_n = \upsilon_A \sin\alpha, \tag{17}$$

and the normal component of the magnetic field

$$B_n = B_o Cos\alpha. \tag{18}$$

In order to keep the shock stationary with respect to the outflow of the plasma, the velocity perpendicular to the shock front must be balanced by the normal component of the Alfven wave velocity, i.e.

$$\upsilon_n = \frac{B_n}{(4\pi\rho)^{1/2}}.$$

Now using (18) and (19) we get

$$\upsilon_A \sin\alpha = \frac{B_o Cos\alpha}{(4\pi\rho)^{1/2}} = \frac{uB}{\upsilon_A}\frac{Cos\alpha}{(4\pi\rho)^{1/2}},$$

or,

$$\tan\alpha = \frac{uB}{\upsilon_A^2}\frac{1}{(4\pi\rho)^{1/2}} = \frac{u}{\upsilon_A} \tag{19}$$

For $u << \upsilon_A$, the angle α is quite small. As u increases, α will increase to accommodate the incoming flow. If $u = \upsilon A$, α becomes $\pi/4$ and the reconnection rate is enhanced considerably. It is possible because the outflow is not confined to the narrow width of the diffusion region like in the Sweet–Parker model. The outflow occurs in an expanding wedge whose angle would change according to the inflow rate. There is, however, an upper limit to the reconnection rate in the Petschek model. Equation (19) has been derived by assuming that u and B are constant throughout the inflow region starting from $x = \pm\infty$ and extending up to the diffusion region. Now that the electric field is also uniform in the steady state, the velocity and the magnetic field at infinity can be related by the relation $uB = u_\infty B_\infty$, where u and B are the velocity and the magnetic field at the diffusion region and u_∞ and B_∞ represent the velocity and the magnetic field at ∞. When α is small for $u << \upsilon_A$, the flow velocity u remain uniform throughout the inflow region with $u = u_\infty$. As α increases, u has to increase and then B decreases in the diffusion region and becomes less than B_∞. This is achieved by rotating the magnetic field vector towards the normal. Since the plasma flow is weakly perturbed in the region outside the diffusion region, it is assumed that the currents in these regions can be neglected (Petschek 1964). Magnetic field can be assumed to be potential and Vasyliunas (1975) has obtained the upper limit of the reconnection which is given as

$$u = (\pi/4)\upsilon_{A\infty}/\ln R_m,$$

where R_m is defined with respect to the speed $\upsilon_{A\infty}$. Thus it is possible to reach a fairly high reconnection rate with an appropriate boundary condition.

2.3 Spontaneous Reconnection or Patchy Reconnection

This is believed to occur when an original stable system moves to a new one with an onset of instability because of gradual changes of the stable system by an external force. The tearing mode instability is one of the most favourite candidates that can be triggered in a current sheet described earlier in a stable equilibrium (Fig. 5). If the electric field is generated consistently by a small disturbance, and not imposed by an external force, then it is called a spontaneous reconnection. The generation of these instabilities requires some kind of dissipation near the neutral point. For dissipation due to collisions, it is called a resistive instability and for dissipation due to the anomalous resistivity caused by wave turbulence or some other sources in a collisionless plasma then it is called a collisionless tearing mode.

Consider a current sheet near the neutral line in the magnetic field configuration shown in Fig. 5 wherein there exists magnetic fields with opposite directions. If the current is disturbed by a small perturbation, then some type of bunching in the current sheet occurs because of the self-generated magnetic field. The instability occurs because of the violation of frozen-in-field condition. It occurs in the region where the current is enhanced as a result of diffusion. The equation that controls the perturbation in fields and currents is given by

$$E^{(1)} + \frac{1}{c}\upsilon^{(1)} \wedge B = \frac{J^{(1)}}{\sigma},$$

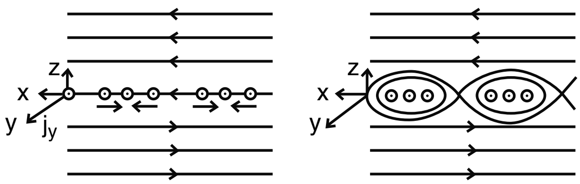


Fig. 5 Magnetic field configuration and tearing model instability

where $E^{(1)}, v^{(1)}$ and $J^{(1)}$ are the first order perturbations. At the center, where $B = 0$, the electric field and currents are governed by the conductivity in the current sheet. Then the magnetic perturbation is given by the induction equation

$$\frac{\partial b^{(1)}}{\partial t} = \frac{c^2}{4\pi\sigma} \nabla^2 b^{(1)}$$

The induction equation outside the resistive region is given by

$$\frac{\partial b^{(1)}}{\partial t} = \nabla \wedge \left(\upsilon^{(1)} \wedge B \right).$$

Usually the parameters that determine this tearing mode instability are obtained by solving these equations outside and inside the region and matching the solution at the boundary. However, one can estimate the growth of this tearing instability by examining the induction equation near the neutral line. The growth rate is given by

$$\gamma_T = \frac{c^2}{\pi \sigma d^2},$$

where length in the x direction is much longer than d, i.e. $k_x < \frac{1}{d}$ can be neglected. It depends upon the width of the current sheet. The narrower the width, the larger is the growth rate. Furthermore, the growth rate increases if the conductivity decreases. Both these results have very important impact on the growth of a solar flare or a coronal mass ejection. As a consequence, this instability is seen to produce magnetic islands or plasmoids in the current-sheet that can be blown away from the region of the neutral point.

2.4 Collisionless Tearing Mode

The tearing mode instability in a current sheet in a collision dominated plasma has been discussed already in the previous section. In a collisionless plasma this can be described adequately in the framework of Vlasov equation where the wave-particle interaction can provide the necessary dissipation in a generalized sense due to electron and/or ion Landau damping. Therefore, there will be a diffusive layer as in the case of the fluid theory and the plasma and field configurations are the same as before containing a neutral layer with a current sheet. Following the general description (see e.g., Laval et al. 1966; Galeev and Zelenyi 1976) the general expression for the growth rate can be obtained. It is seen to produce substantial growth for a very thin current sheet. One of the significant results is the generation of a magnetic island or a plasmoid. However, a finite component of the magnetic field normal to the current sheet, which magnetizes the electron in a neutral layer, stabilizes the electron tearing mode and as a result, no reconnection can take place.

Ion-tearing mode can still exist (Schindler 1974; Galeev and Zelenyi 1976). This, however, is limited to only a small range of magnetic field. In the case of a reconnection in the tail of the magnetosphere, Buchner and Zelenyi (1987) have suggested that ion-tearing can be sustained if the electron orbits become chaotic (Chen and Palmadesso 1986). However, more recent numerical simulation indicates that even the chaotic motion of electrons cannot prevent stabilization of the tearing mode. Therefore the possibility of the influence of external sources in the generation of the instability has been explored by several researchers. One of such sources considered by Sundaram et al. (1979) and Das (1992) was the effect of the background of lower hybrid/ion-cyclotron waves. It was shown that in such cases the tearing mode instability can be sustained with higher growth rate. In another approach, Lakhina (1992) considered a small perturbation at the boundary of the magnetospheric tail, as an external source to develop a theory of driven reconnection in the Earth's magneto tail. Recent work on numerical simulations has enhanced our understanding of reconnection enormously.

3 Role of Magnetic Reconnection in Solar Flares

Solar flares are believed to be generated by magnetic reconnection. Flares are generally classified into two types: compact and two ribbon flares. The two ribbon flares appear as long-lived, slowly developing large loops. During the preflare phase, a large flux tube with an overlying arcade of magnetic field lines rises slowly upward into the corona because of several possible reasons. For example, it may be due to a spontaneous eruptive MHD instability or due to a magnetic non-equilibrium state when the separation of footpoints or the twist in the field or the pressure becomes too large. As the field lines form inverted Y-shaped structure and relax, the reconnection of the field lines takes place below the filament leading to shocks generated by the process described by Petschek and to impulsive bursty regions. This is also known as emerging flux or interacting flux models. A general scenario on the basis of simulations and observations is presented in Fig. 6.

A slightly different mechanism was thought to be required to produce short-lived impulsive flares known as compact flares. However, observations made with Yohkoh Hard-X-ray Telescope (HXT) and Soft X-ray Telescope (SXT) show a compact flare with a geometry similar to that of a two ribbon flares. The reconnection region is identified as the site of particle acceleration suggesting that the basic physics of the reconnection process may be common to both the types of flares. A schematic view of this process is shown in Fig. 7.

Now we would like to elaborate the present status of our understanding from the point of view of observations.

1. Cusp-shaped loop structure in long duration events (LDE) is one of the most important discoveries of Yohkoh (SXT). An example of this kind of flare is shown at the left of the Fig. 6. This was reported to be seen within a few hours after a large-scale coronal eruption, which created a helmet streamer-like configuration,

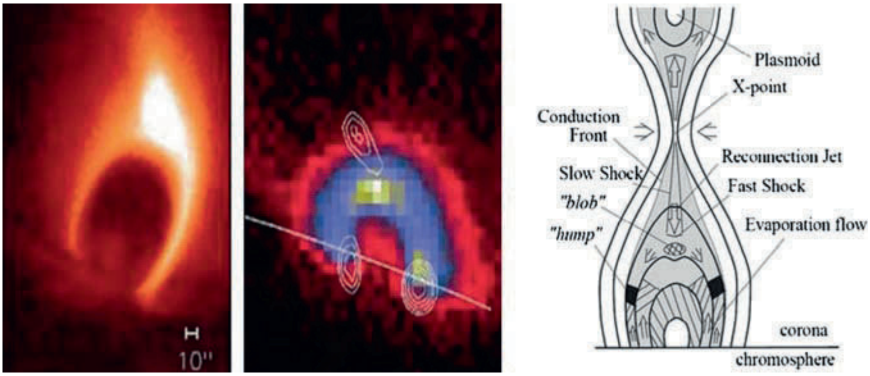


Fig. 6 Observational evidence and illustration of simulation results for the magnetic reconnection in solar flares – a general scenario. (*left*) Illustration of LDE (long duration event) flares; (*middle*) Observation of loop-top hard X-ray source, (reprinted with permission from Elsevier, K. Shibata 1996); (*right*) A schematic illustration of the reconnection model of a solar flare on the basis of simulation results.(reproduced by permission of AAS, Yokoyama and Shibata 2001)

which in turn suggests the formation of a current sheet as a result of some MHD instability. The apparent height of the loop and the distance between the footpoints of the loop are also seen to increase gradually. This actually inferred the successive magnetic field reconnection in the current sheet above the loop, which is supported by the theoretical studies. Most recent observations of flares and other related phenomena with SXT onboard Yohkoh have also shown the soft X-ray (SXR) emissions association with the reconnection. In fact these are considered to be the evidence of reconnection. With a closer examination of the observations, it is reported that a small magnetic island (or plasmoid) is ejected during the rise phase of a flare. It is likely that the ejection of the plasmoid triggered the flare. Similarly the observations by HXT, also on Yohkoh, have brought out the major discovery of a hard X-ray (HXR) loop top source above a soft X-ray bright loop during the impulsive phase. This is shown in the middle of the Fig. 6. Furthermore, long duration event (LDE), large-scale arcade formation and impulsive flares show many common features such as plasmoid/filaments ejection in the Yohkoh data. All these phenomena have led to a model for flares based on reconnection as shown in Fig. 7.

2. Cusp-shaped loops or arcades have also been found in a scale much larger than that for LDEs although the evolution of the features are similar. A large helmet streamer, which may occur due to a coronal mass ejection, has been conjectured as another view of a large-scale arcade. Yohkoh/SXT has revealed that these large-scale arcades are very similar to LDE flares from various points of view apart from the size and magnetic field strength. The formation of a magnetically closed helmet structure observed as a bright feature with the SXT on the Yohkoh spacecraft is interpreted as a consequence of magnetic reconnection that proceeded along a vertical magnetic neutral sheet formed by a mass ejection.

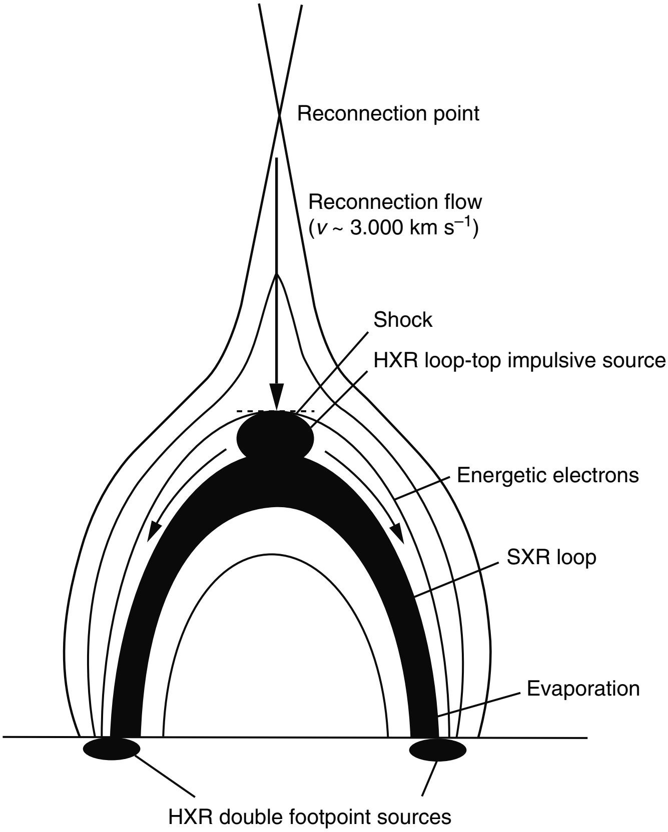


Fig. 7 A schematic view of the site of particle acceleration as a result of reconnection (Masuda et al. 1994; reproduced by permission from Macmillan Publisher Ltd., Nature 371, 1994)

3. The detection of a reconnection inflow as well as an outflow has been an important aspect to our understanding of reconnection in solar flares. In fact, this is considered to be a very important signature of reconnection near the neutral point or line.
4. *Impulsive flare*: We have already mentioned that this is the second type of flares, which does not show a cusp-shaped structure. These are bright in hard X-rays and show an impulsive phase whose duration is short. Masuda et al. (1994) have observed a loop-top hard X-ray source appearing well above a soft X-ray bright loop. This is a distinct indication that an impulsive energy release did not occur within the soft X-ray loop but above the loop. One possible physical mechanism to produce such a loop-top hard X-ray source is a magnetic reconnection occurring near the X-type neutral point much above the loop as shown in Fig. 6. It is shown by Soward and Priest (1982) that the outflow from the reconnection process produced by the Petschek mechanism is supermagnetosonic with respect to fast-mode magnetosonic wave speed and so anything that obstructs the flow will

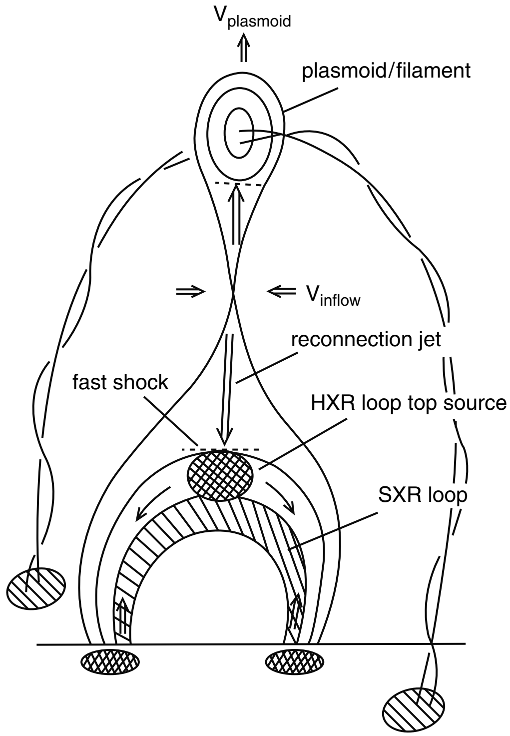


Fig. 8 A unified model on the basis of plasmoid-driven reconnection (reprinted with permission from Elsevier, Shibata 1996)

create a fast shock, which produces superhot plasma and/or high energy electrons emitting hard X-rays. A schematic view is shown in Fig. 7. The discovery of the loop-top HXR source has made it possible to think of unifying two classes of flare, LDE flare and impulsive flare by a single mechanism of magnetic reconnection. If these impulsive flares are also generated by the reconnection process, then these must be associated with a plasmoid or a filament ejection. Interestingly many such flares are found to be associated with a plasmoid ejection (Masuda et al. 1994). Using all these findings, Shibata (1996) has proposed a model based on a plasmoid ejection, which is shown in Fig. 8.

In this model the impulsive phase corresponds to the initial phase of the plasmoid injection. The ejection of the plasmoid produces a strong inflow into the x-point which drives the fast reconnection. The inflow velocity is estimated by using the conservation of mass and assuming that the density does not change appreciably. The Alfven Mach number of the inflow becomes comparable to the inflow speed expected from the Petschek model. Furthermore, according to the reconnection theory, two high speed jets from the neutral point move in opposite direction at a velocity $v_O \sim v_A$. However, in the fast reconnection process the jet can exceed

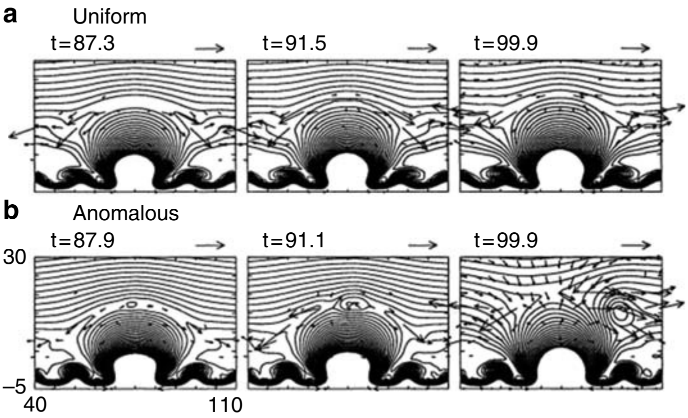


Fig. 9 Recent MHD simulation of reconnection between emerging flux and coronal field (reproduced by permission of AAS, Yokoyama and Shibata 1994)

fast-mode wave speed, and when it collides with the top of the SXR loop it generates fast shocks producing high energy electrons at the loop top which are observed as HXR. This seems to support the observation on temperature.

3.1 Reconnection Between Emerging Flux and Coronal Field

When an emerging flux interacts with the quiet coronal field, two horizontal jets or loops are seen in both sides of the emerging flux regions. Recent MHD simulation of the above physical situation has shown the formation of magnetic islands that are ejected out of the current sheet. This is shown in Fig. 9. The simulation seems to explain the above features very well by the magnetic reconnection model suggested by Yokoyama and Shibata (1994).

A localized resistivity seems to be essential for the fast reconnection. If the current sheet is long enough, the tearing mode instability with anomalous resistivity may produce such a reconnection and island formation. The plasmoid produced in this situation appears to be similar to the plasmoid ejected in larger flares shown in Fig. 6.

3.2 Recent Simulation and Present Status of Theoretical Understanding

More realistic simulation of a flare has been done by Yokoyama and Shibata (2001) by including thermal processes that are important for quantitative comparison with observations. Both the temperature and density evolution leading to the reconnection and island formation are shown in Fig. 10. The regions of X-ray emissions are

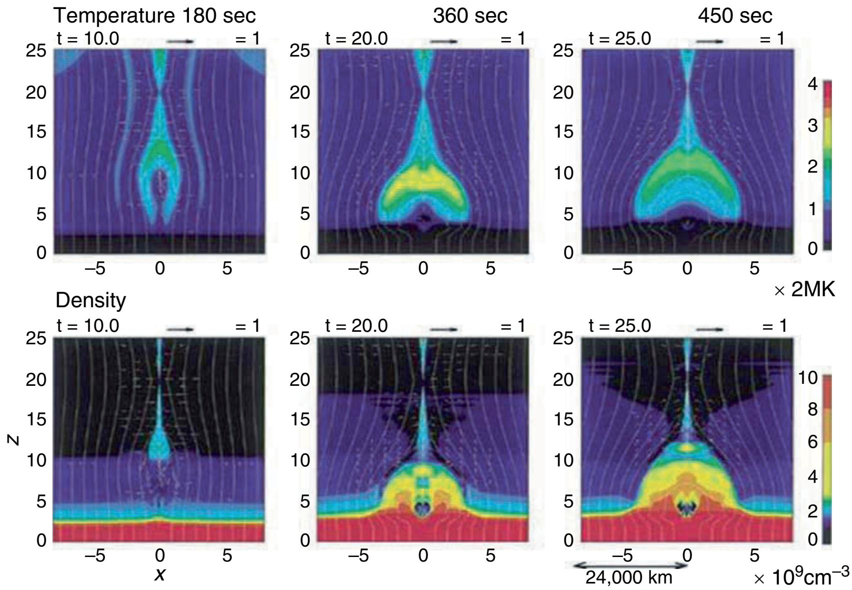


Fig. 10 A more realistic simulation model leading to reconnection and island formation (reproduced by permission of AAS, Yokoyama and Shibata 2001)

also seen in Fig. 10. The localized resistivity appears to be necessary for the fast reconnection (Ugai 1999a,b; Yokoyama and Shibata 2001).

3.3 Scale-Matching Between Macro and the Micro Features or Scales

This is an important aspect of reconnection which remains unanswered. For example, there is an enormous gap of scale sizes between the region of anomalous resistivity and a solar flare. The scale size of the anomalous resistivity d can be approximated by $d = \rho_i \sim 10\,\mathrm{m}$, where d is the thickness of the current sheet and ρ_i represents the ion Larmour radius. The scale size of a flare $\approx 10^4$ km. Therefore there is a gap of seven orders of magnitude between the small scale size and the large scale size. The MHD turbulence or the stochastic magnetic reconnection may have an answer to this problem.

3.4 Turbulent Structure in the Magnetic Reconnection Jet

Tanuma and Shibata (2005) have shown a very interesting turbulent structure containing internal shocks in the magnetic reconnection jet in solar flares. They have

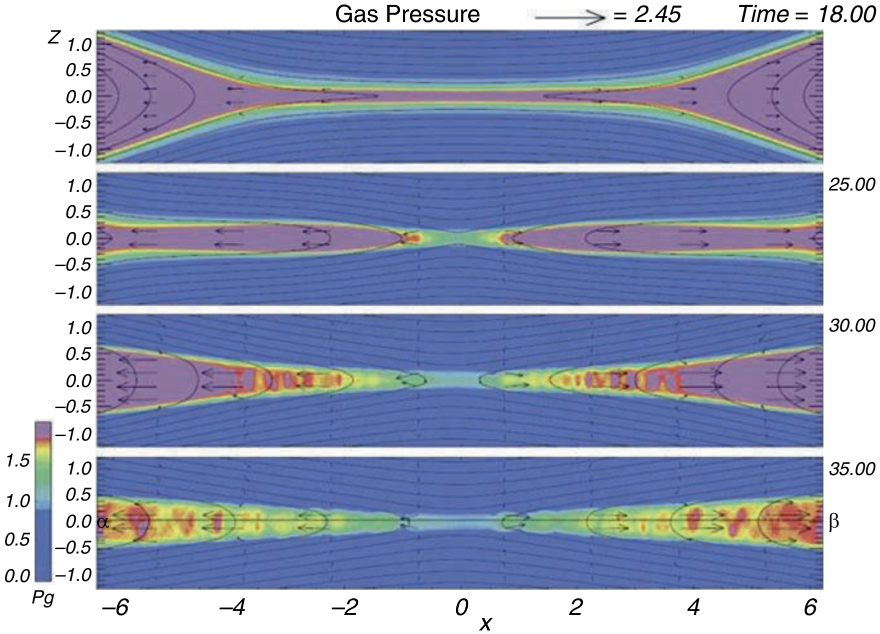


Fig. 11 Turbulent structure in the outflow. The Petschek-like reconnection is seen in the top panel while the development of multiple island-like structures are shown in other panels (reproduced by permission of AAS, Tanuma and Shibata 2005)

examined how magnetic reconnection creates multiple shocks by performing a 2-D resistive MHD simulation. This is shown in Fig. 11. Certain amount of energy is seen to be continuously transferred into the system. A secondary tearing mode instability plays a crucial role.

3.5 *Reconnection in Heliospheric Current Sheet*

There exist a large number of satellite observations giving evidence for the reconnection in the planetary space and around the heliospheric current sheet. During the solar maximum mission, a few probable coronal disconnection events appearing as the pinching of a helmet stream followed by the release and outward acceleration of U- and V-shaped structures are observed. The sequence of events is such that it indicates reconnection across the heliospheric current sheet between previously open field region and the creation of detached magnetic structure. This process then sends back the open field lines towards the Sun as closed field arches. There also exist observations of an internal magnetic reconnection within a flux rope. As the flux rope field lines at the interface of the two-speed solar wind are sheared, oppositely directed field lines are generated which press together and get reconnected

(Schmidt and Cargil 2001). Magnetic reconnection caused by resistive tearing mode in a non-periodic multiple current sheets has been investigated by using two dimensional simulations. The result shows that it produces unsteady complex reconnection which is supported by two observations of very complex structure with directional discontinuities in magnetic field current system. These discontinuities observed at the sector boundary crossing in the heliosphere may be associated with magnetic islands and plasmoids due to reconnection (Wong et al. 1997).

4 Summary

- Magnetic reconnection is the underlying driver of giant explosive releases of magnetic energy in the Sun's atmosphere that are observed as solar flare or CMEs.
- Several compelling observational evidence for reconnection which support reconnection model of solar flares are presented.
- Numerical simulation suggests that the localized resistivity is necessary for magnetic reconnection.
- There is still an enormous gap between the microscale of anomalous resistivity and the size of solar flares.
- The MHD turbulence model of reconnection shows interesting features in various cases and may play an interesting role in solving the scale-matching problem.

Acknowledgements The author would like to thank Mr. Hitesh Macwan for his help in preparing this manuscript.

References

Buchner, J., Zelenyi, L. M.: Chaotization of the electron motion as the cause of internal magnetotail instability and substorm onset. J. Geophys. Res. **92**, 13456 (1987)

Chen, J., Palmadesso, P. J.: Chaos and nonlinear dynamics of single particle orbit in a magnetotail like magnetic field. J. Geophys. Res. **91**, 1499 (1986)

Das, A.C.: Lower hybrid turbulence and tearing mode instability in magnetospheric plasma. J. Geophys. Res. **97**, 12275 (1992)

Das, A.C.: Space Plasma Physics: An Introduction. Narosa Publishing House, New Delhi (2004)

Dungey, J.W.: Cosmic Electrodynamics. Cambridge University Press, London (1958)

Dungey, J.W.: Interplanetary magnetic field and the auroral zone. Phys. Rev. Lett. **6**, 47 (1961)

Galeev, A.A., Zelenyi, L.M.: Tearing instability in plasma configurations. Sov. Phys. JETP **43**, 1113 (1976)

Lakhina, G.S.: A Kinetic theory of driven reconnection in the Earth's magnetotail. J. Geophys. Res. **97**, 2961 (1992)

Lakhina, G.S.: Magnetic reconnection. Bull. Astr. Soc. India **28**, 593 (2000)

Laval, G., Pellat, R., Vuillemin, M.: In: Plasma Physics and Controlled Fusion Research, vol. 2, p. 259. International Energy Agency, Vienna (1966)

Masuda, S.: Ph.D thesis, University of Tokyo (1994)

Masuda, S., Kosugi, T., Hara, H., Tsnueta, S., Ogawara, Y.: A loop top hard X-ray source in a compact solar flare as evidence for magnetic reconnection. Nature **371**, 495 (1994)
Parker, E.N.: Sweet's mechanism for merging magnetic field in conducting fluids. J. Geophys. Res. **62**, 509 (1957)
Petschek, H.E.: Magnetic field annihilation. In: Hess, W.N. (ed.) The Physics of Solar Wind, p. 425. NASA, SP-50, Washington D.C. (1964)
Schindler, K.: A theory of substorm mechanism. J. Geophys. Res. **79**, 2803 (1974)
Schmidt, J.M., Cargil, P.J.: Magnetic cloud evolution in a two speed solar wind. J. Geophys. Res. **106**, 8283 (2001)
Shibata, K.: New observational facts about solar flares from Yohkoh studies – evidence of magnetic reconnection and a unifield model of flares. Adv. Space Res. **17**(4/5), 9–18 (1996)
Soward, A.M., Priest, E.R.: Fast magnetic field line reconnection in a compressible fluid, I coplanar field lines. J. Plasma Phys. **28**, 335 (1982)
Sundaram, A.K., Das, A.C., Sen, A.: A nonlinear mechanism for magnetic reconnection and substorm activity. Geophys. Res. Lett. **7**, 921 (1979)
Sweet, P.A., The neutral point theory of solar flares. In: Lehnert, B. (ed.) Electromagnetic Phenomena, Cosmical Physics, p. 123. Cambridge University Press, New York (1957)
Tanuma, S., Shibata, K.: Internal shocks in the magnetic reconnection jet in solar flares: multiple fast shocks created by secondary tearing instability. Astrophys. J. Lett. **628**, L77–L80 (2005)
Ugai, M.: Basic Physical mechanism of reconnection development and magnetic loop dynamics. J. Geophys. Res. **104**(A4), 6929 (1999a)
Ugai, M.: Computer studies on the spontaneous fast reconnection model as a nonlinear instability. Phys. Plasma **6**, 1522 (1999b)
Vasyliunas, V.M. Rev. Geophys. **13**, 303 (1975)
Wong, S., Lin. Y.F., Zing, H.N.: Magnetic Reconnection in Multiple Heliospheric Current Sheets. Solar Phys. **173**, 409 (1997)
Yokoyama, T., Shibata, K.: What is the condition for fast magnetic reconnection? Astrophys. J. Lett. **436**, L197–L200 (1994)
Yokoyama, T., Shibata, K.: Magnetohydrodynamic simulation of a solar flare with chromospheric evaporation effect based on the magnetic reconnection model. Astrophys. J. **549**, 1160 (2001)

MHD Fluctuations in the Heliosphere

B. Bavassano and R. Bruno

Abstract The solar wind is an excellent laboratory to study the MHD turbulence behaviour in a collisionless plasma. This is a fundamental topic for both plasma physics and astrophysics. The impressive amount of observations at different solar distances and latitudes collected in the last decades has allowed us of reaching a good understanding on many aspects of the complex phenomenon of the solar wind variability at MHD scales. Here we discuss the character of the observed fluctuations and focus on their radial evolution as the plasma flow expands into interplanetary space. A comparison is performed between fluctuations seen in low- and high-latitude solar wind. Implications about processes of local generation of turbulence are discussed.

1 Introduction

The heliosphere is the region of space influenced by the Sun through its expanding corona, the solar wind. This is a collisionless plasma flow that offers one of the best opportunities to study plasma phenomena by in-situ measurements. Here we will focus on a topic of primary importance for both plasma physics and astrophysics, the magnetohydrodynamic (MHD) turbulence. The use of the term MHD only refers to the fact that we are looking at phenomena falling in the MHD regime (i.e., frequency well below the proton gyrofrequency). The name does not imply that these phenomena may necessarily be described by the MHD theory.

The three-dimensional structure of the solar wind is strongly dependent upon the solar cycle. As shown in Fig. 1, at low solar activity the solar wind is characterized by a bimodal structure (McComas et al. 1998), with a steady, fast wind at high

B. Bavassano (✉)
Istituto di Fisica dello Spazio Interplanetario (Istituto Nazionale di Astrofisica)
Via del Fosso del Cavaliere 100, 00133 – Roma, Italy
e-mail: Bruno.Bavassano@ifsi-roma.inaf.it

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_8,

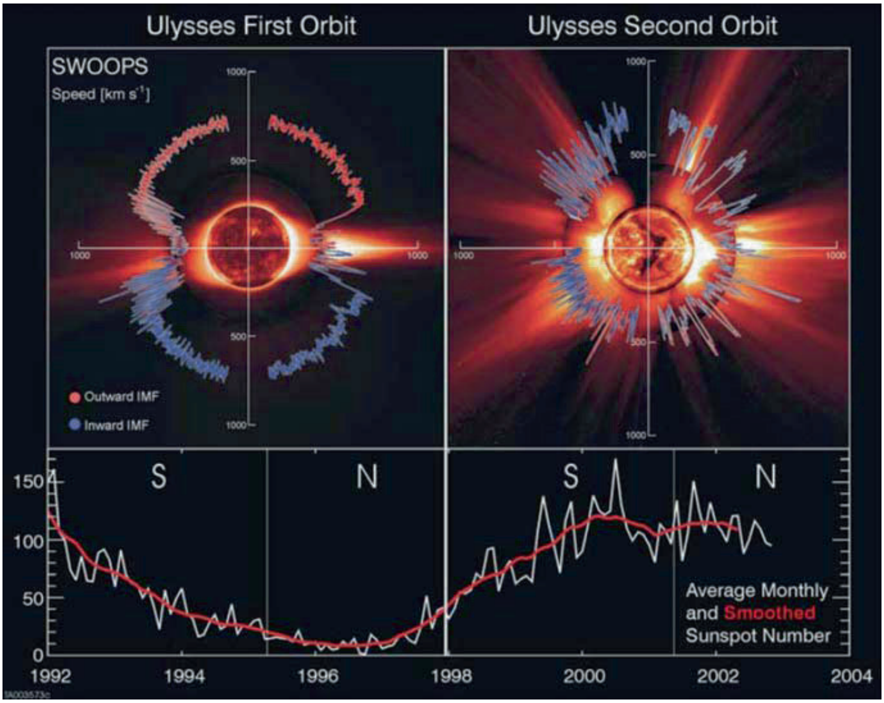


Fig. 1 Polar plots of the solar wind speed as a function of latitude for the Ulysses' first two orbits. Sunspot numbers (*bottom panel*) show that the first orbit occurred through the solar cycle declining phase and minimum while the second orbit spanned solar maximum [adapted from McComas et al. (2003), copyright 2003 American Geophysical Union, modified by permission of American Geophysical Union]

latitudes and a slower and more variable wind at low latitudes. Conversely, at high solar activity highly variable flows are observed at all latitudes and the wind structure appears to be a complicated mixture of flows (McComas et al. 2003). However, this state with variable flows at all heliographic latitudes is a short-lived feature of the heliosphere. After the reversal of the solar magnetic field, a recovery of the winds bimodal structure begins quite soon (McComas et al. 2002). The fast high-latitude wind (polar wind) and the low-latitude wind come from different sources. The fast wind is from coronal holes, while the slow wind flows from open magnetic flux-tubes in the streamer belt region.

The solar wind is a plasma with extremely low viscosity and resistivity. Both kinetic and magnetic Reynolds numbers, which essentially measure the relative weight of nonlinear and dissipative terms in the MHD equations, are high. Thus, relevant turbulent effects have to be expected. As well known, the solar wind turbulence strongly affects several aspects of the heliospheric behaviour, such as plasma heating, solar wind generation, particles acceleration, and cosmic rays propagation.

Recently it has been shown (D'Amicis et al. 2007) that the turbulence is also able to influence the geomagnetic activity.

In the seventies and eighties impressive advances have been made in the knowledge of turbulent phenomena in the solar wind (Tu and Marsch 1995). In those years, however, with spacecraft observations confined within a small latitudinal belt around the solar equator, only a limited fraction of the heliosphere was accessible. In the nineties, with the launch of Ulysses, the first spacecraft with a highly inclined orbital plane with respect to the ecliptic, the investigations have been extended to high-latitude regions of the heliosphere. This has allowed of studying how the MHD turbulence evolves in polar solar wind, a plasma flow in which the effects of large-scale inhomogeneities are considerably less important than in low-latitude wind. With this new laboratory relevant advances in interplanetary turbulence have been made (Bruno and Carbone 2005; Horbury et al. 2001).

2 Alfvénic Fluctuations

Since the first observations [e.g., Belcher and Davis (1971); Coleman (1968)] it was recognized that the solar wind MHD fluctuations have a strong Alfvénic character [Alfvénic modes have a much longer lifetime than other MHD modes Barnes et al. (1979)]. An example is shown in Fig. 2, with highly correlated variations in velocity and magnetic field (and almost constant density and magnetic field magnitude).

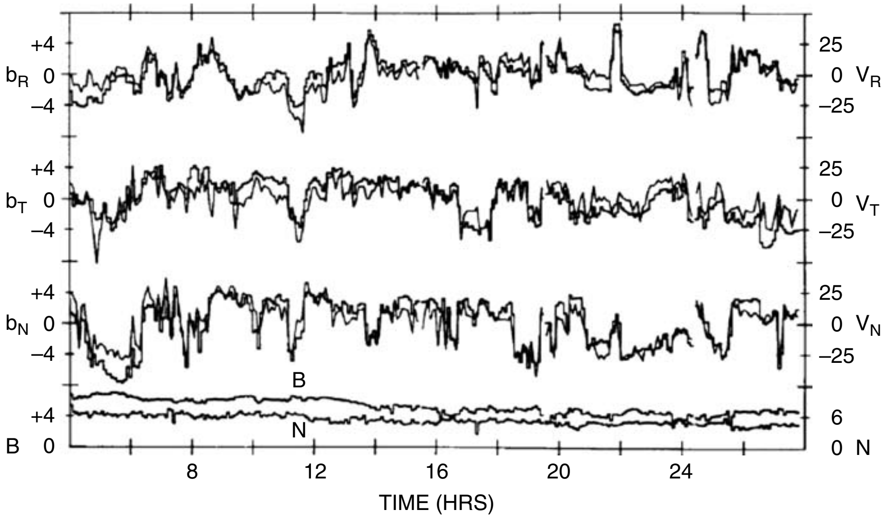


Fig. 2 One of the first observations of Alfvénic fluctuations [adapted from Belcher and Davis (1971), copyright 1971 American Geophysical Union, modified by permission of American Geophysical Union]

It is useful to briefly recall some of the parameters that are generally used to describe Alfvénic fluctuations. Basic quantities are the Elsässer's variables $\mathbf{z}_\pm = \mathbf{v} \pm \mathbf{b}$ [e.g., see Tu and Marsch (1995)], that are ideally suited to extract an Alfvénic signal from velocity and magnetic field measurements. Here $\mathbf{v}$ and $\mathbf{b}$ are the velocity and magnetic field vectors, respectively, with the magnetic field $\mathbf{b}$ scaled to Alfvén units (i.e., divided by $\sqrt{4\pi\rho}$, with ρ the mass density). Taking into account how the sign of the Alfvénic correlation depends on the propagation direction with respect to the background magnetic field, it has become common to use the above definition for a background magnetic field pointing to the Sun, while the equation $\mathbf{z}_\pm = \mathbf{v} \mp \mathbf{b}$ is taken for the opposite polarity. With this choice we have that, whatever the magnetic polarity is, $\mathbf{z}_+$ ($\mathbf{z}_-$) fluctuations always correspond to modes with an outward (inward) direction of propagation, with respect to Sun, in the plasma frame. The relative weight of the energies (per unit mass) e_+ and e_- associated to $\mathbf{z}_+$ and $\mathbf{z}_-$ fluctuations, respectively, is measured by the Elsässer ratio $r_E = e_-/e_+$. An analogous measure for the energies e_V and e_B of the $\mathbf{v}$ and $\mathbf{b}$ fluctuations is given by the Alfvén ratio $r_A = e_V/e_B$. Other related parameters are the normalized cross-helicity $\sigma_C = (e_+ - e_-)/(e_+ + e_-) = (1 - r_E)/(1 + r_E)$ and the normalized residual energy $\sigma_R = (e_V - e_B)/(e_V + e_B) = (r_A - 1)/(r_A + 1)$.

Both outward and inward propagating fluctuations are observed in the solar wind. Outward fluctuations mainly have a solar origin (or, more precisely, inside the Alfvén critical point). Conversely, inward fluctuations can only be generated outside such critical distance. As well known, the presence of both kinds of fluctuation leads to the development of nonlinear interactions.

3 Magnetic Variations

As first discussed by Bavassano et al. (1998), when solar wind fluctuations are analyzed in terms of σ_C and σ_R and occurrence frequency distributions are displayed in a $\sigma_C - \sigma_R$ plane, it clearly appears that the peak at high σ_C due to the Alfvénic population is accompanied by an extended tail towards the region close to $\sigma_C = 0$ and $\sigma_R = -1$, with a secondary peak at the end caused by fluctuations essentially of magnetic type. Figure 3 indicates that this is a quite general feature, observed for all kinds of solar wind regime (Bavassano et al. 1998). The radial expansion of the solar wind plays a role in this two-population scheme (Bavassano and Bruno 2006; Bruno et al. 2007), with an increasing weight of the magnetic fluctuations. Moreover, it has been shown (Bavassano and Bruno 2006) that the magnetic population, even when of secondary importance in terms of occurrence frequency, may correspond to a primary peak in the distribution of total (magnetic plus kinetic) energy. Thus, from the point of view of the energy content of solar wind fluctuations the magnetic population appears able to play a primary role. Finally, for the magnetic population fluctuations are more planar, in a plane perpendicular to the background magnetic field, than those for the Alfvénic population. This is reminiscent of quasi-two-dimensional fluctuations (Matthaeus et al. 1990). Yet, this kind of geometry

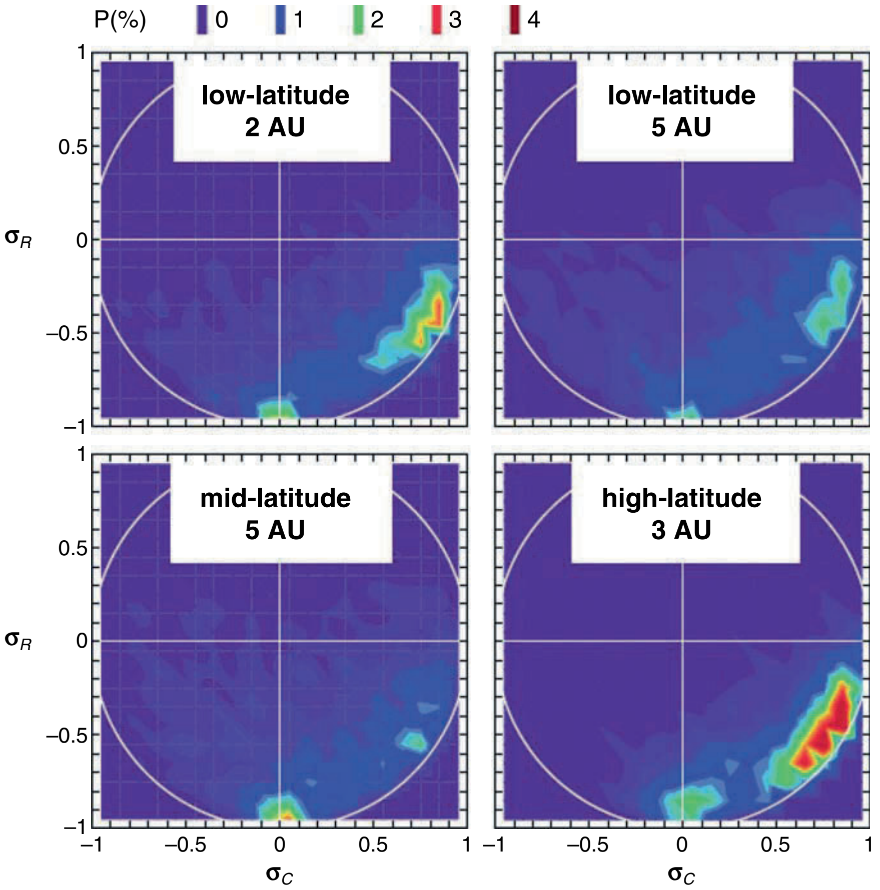


Fig. 3 Occurrence frequency distributions in the $\sigma_C - \sigma_R$ plane [modified from Bavassano et al. (1998), copyright 1998 American Geophysical Union, by permission of American Geophysical Union]

may well come from coronal magnetic flux tubes that, turbulently mixed, are swept out by the solar wind (Bavassano and Bruno 2006).

3-D plots of the fluctuation energy values are shown in Fig. 4. In addition to the highly populated clouds indicated by green ovals, that correspond to outward propagating Alfvénic fluctuations, a broad population extending towards higher magnetic energies (e_B) and lower cross-helicity (σ_C) is clearly apparent in the left plot. The same does not hold for e_V (right plot). All this confirms that solar wind fluctuations are a mixture of an Alfvénic population and a magnetic population. For this last one the values of e_B are significantly above the levels typically observed for the Alfvénic population.

In a complementary approach, it is interesting to investigate the role that the two populations have in building the distributions of magnetic and kinetic energies.

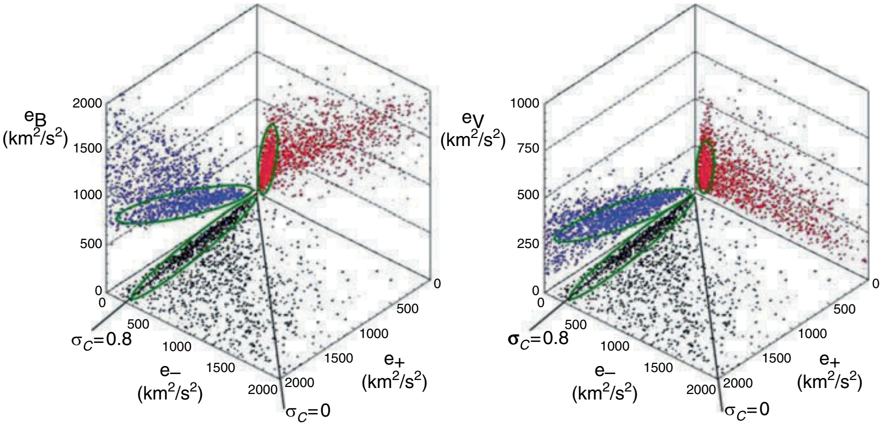


Fig. 4 3-D plots of fluctuation energy values in the polar wind. An $e_+ - e_- - e_B$ frame is used at left and an $e_+ - e_- - e_V$ frame at right (note that e_B and e_V axes have different scales). Dots of different colours indicate projections on the coordinate planes. Lines for $\sigma_C = 0.8$ and $\sigma_C = 0$ are shown in the plane $e_+ - e_-$

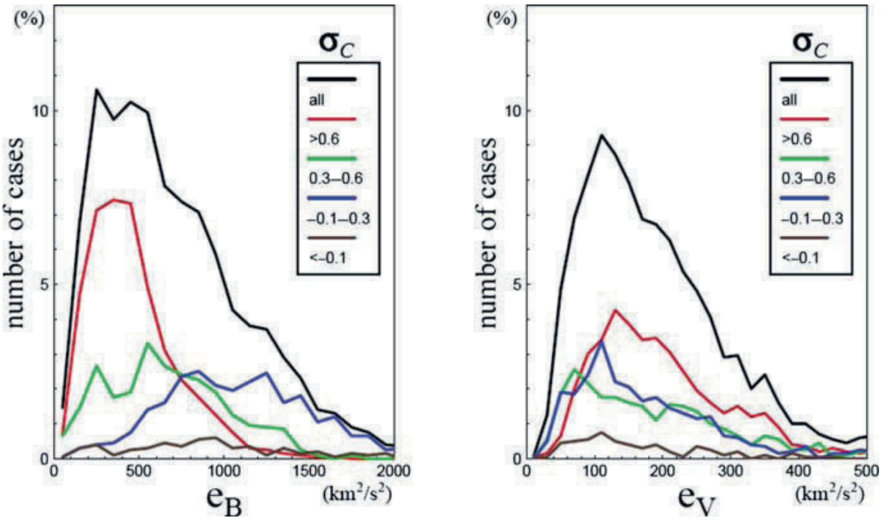


Fig. 5 Histograms of e_B (*left*) and e_V (*right*) for different ranges of σ_C

For the same data of Fig. 4 such distributions have been derived separately for four ranges of σ_C and compared to that obtained from the whole sample. The results (Fig. 5) show that the e_B and e_V distributions respond in a different way to a change in the σ_C range. While for the kinetic energy (right) the various curves have similar shapes, for the magnetic energy (left) significant differences are observed. In particular, Alfvénic fluctuations (σ_C above 0.6, red) are mainly responsible for the peak of the overall distribution (black), while non-Alfvénic fluctuations (σ_C from -0.1 to 0.3, blue) are dominant in the extended (high-e_B) tail. In other words, the peak and the tail of the e_B distribution have a different Alfvénic content or, in general, a different nature.

4 Turbulence in Low-Latitude Wind

Since the first studies on Alfvénic fluctuations in ecliptic wind [e.g., see Belcher and Davis (1971)] it was realized that fast streams (or, more precisely, their trailing edges) were the best places to observe them. In those years (mid-seventies) a hot topic was that of the turbulence generation and evolution. Observations at different heliocentric distances are crucial to this kind of study. The launch of Helios 1 and 2 (in December 1974 and January 1976, respectively), systematically covering the inner heliosphere from 0.3 to 1 AU, combined with a very stable pattern of the interplanetary medium (in a period of low solar activity) allowed to get the first evidence about a radial evolution of the solar wind turbulence (Bavassano et al. 1982; Denskat and Neubauer 1982). In the next years, with a boom of studies on this subject, impressive advances were made.

Figure 6 shows how e_+ and e_- power spectra (solid and dotted line, respectively) vary when solar distance increases from 0.3 to 0.9 AU (Marsch and Tu 1990). The e_+ spectrum declines faster than that of e_-, with the result that the two spectra approach each other. At the same time the spectral slopes evolve in such a way that an extended inertial regime expands to low frequencies.

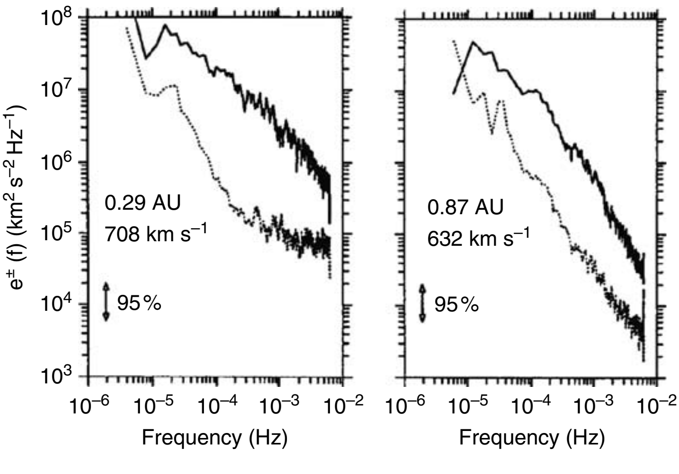


Fig. 6 Power spectra of e_+ and e_- (*solid and dotted line*, respectively) in the trailing edge of fast streams at 0.29 and 0.87 AU [adapted from Marsch and Tu (1990), copyright 1990 American Geophysical Union, modified by permission of American Geophysical Union]

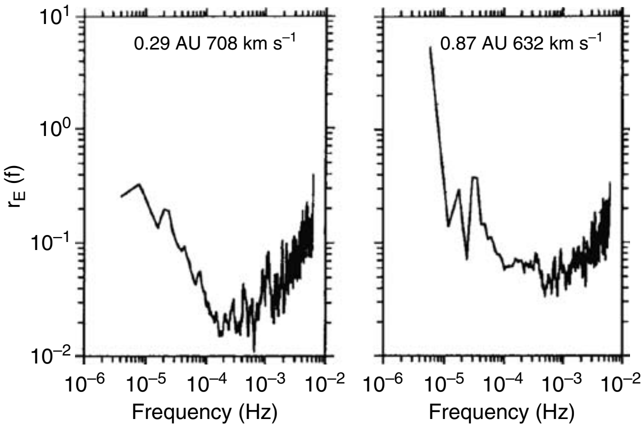


Fig. 7 Power spectra of the Elsässer ratio $r_E(= e_-/e_+)$ for the same data of Fig. 6 [adapted from Marsch and Tu (1990), copyright 1990 American Geophysical Union, modified by permission of American Geophysical Union]

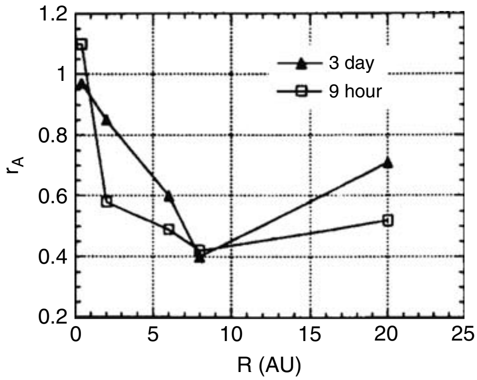


Fig. 8 Radial variation of the Alfvén ratio $r_A(= e_V/e_B)$ as seen by Helios and Voyagers between 0.3 and 20 AU [adapted from Roberts et al. (1990), copyright 1990 American Geophysical Union]

The corresponding variation in the relative weight of outward and inward fluctuation energies is shown in Fig. 7 (Marsch and Tu 1990). The very small values of r_E observed at 0.3 AU in the core of the Alfvénic regime (frequencies around $10^{-4} - 10^{-3}$ Hz) have disappeared at 0.9 AU. However, the predominance of $\mathbf{z}_+$ fluctuations remains a strong feature of the Alfvénic turbulence observed by Helios in fast streams inside 1 AU. Ulysses data from the ecliptic phase of the mission have shown that this situation persists at least up to 5 AU (Bavassano et al. 2001).

A relevant feature of the solar wind turbulence is that the magnetic energy tends to become dominant as the solar distance increases. This decreasing trend for the Alfvén ratio $r_A(= e_V/e_B)$ clearly appears in Helios data (Bruno et al. 1985; Marsch and Tu 1990). The r_A decrease, however, is not without a limit. This is seen in Fig. 8, that combines Helios and Voyagers observations to give an overall view of the radial variation of r_A between 0.3 and 20 AU (Roberts et al. 1990). The 9-h curve (squares), the one of the two reported curves that best describes the typical MHD scales, shows that r_A, after a fast and pronounced decrease, remains nearly unchanged.

5 Turbulence in Polar Wind

As already mentioned, Ulysses observations have shown that, at low solar activity, the solar wind at high latitudes is a fast and relatively steady flow [e.g., McComas et al. (1998)]. A remarkable feature of the polar wind is the ubiquitous presence of an intense flow of Alfvénic fluctuations [e.g., Goldstein et al. (1995); Horbury et al. (1995); Smith et al. (1995); Bavassano et al. (1998)]. Similar to previous ecliptic observations in fast streams, a largely dominant fraction of these fluctuations is outward propagating, with respect to the Sun, in the solar wind frame. The polar wind, a flow in which large-scale inhomogeneities are almost absent, offers the best opportunity to study the turbulence evolution under nearly undisturbed conditions (with respect to the low-latitude wind).

Figure 9 shows polar wind spectra of $\mathbf{z}_+$ and $\mathbf{z}_-$ at distances of about 2 and 4 AU from the Sun (Goldstein et al. 1995). The spectral evolution appears qualitatively similar to that observed in low-latitude fast streams, with the development of a turbulent cascade with increasing distance that moves to lower frequencies the breakpoint between the f^{-1} and $f^{-5/3}$ regimes. However, as shown by Fig. 10, for polar wind the breakpoint is at smaller scale than at similar distances in low-latitude wind (Horbury et al. 1996). Thus, spectral evolution in polar wind is slower than at low latitudes.

A general agreement exists about the fact that the slower evolution for polar turbulence has to be ascribed to the lack of a large-scale stream structure. The role

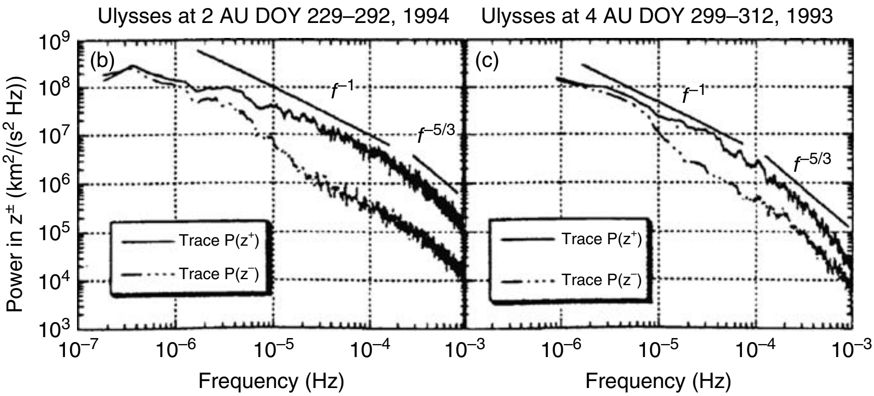


Fig. 9 Power spectra of $\mathbf{z}_+$ and $\mathbf{z}_-$ (*upper and lower curve*, respectively) in the polar wind at (*left*) ~2 AU and (*right*) ~4 AU [adapted from Goldstein et al. (1995), copyright 1995 American Geophysical Union, modified by permission of American Geophysical Union]

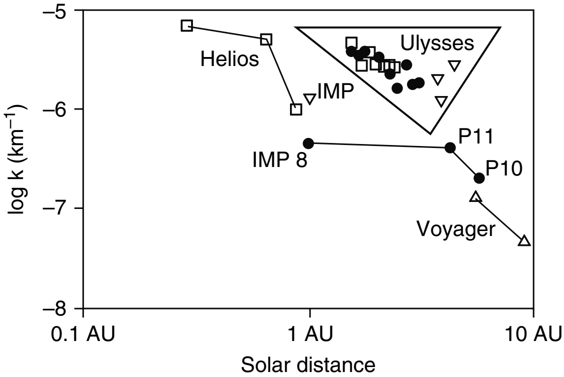
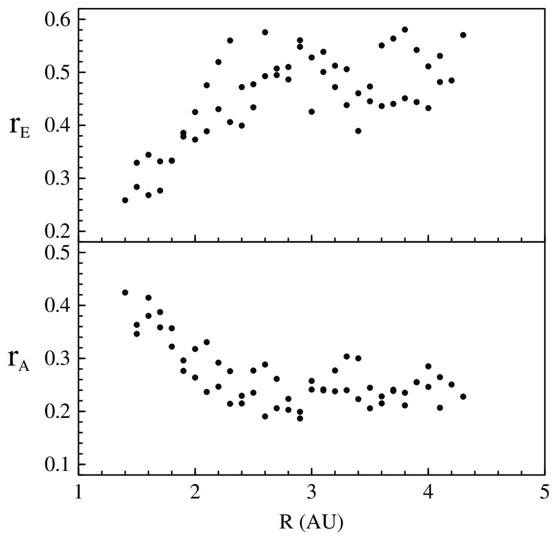


Fig. 10 Variation of the spectral breakpoint with radial distance. Data inside the upper triangle are for high-latitude observations by Ulysses, while the other data are from spacecraft near the ecliptic [adapted from Horbury et al. (1996), copyright 1996 American Geophysical Union, modified by permission of American Geophysical Union]

Fig. 11 Radial variation of (*top*) the Elsässer ratio $r_E(= e_-/e_+)$ and (*bottom*) the Alfvén ratio r_A $(= e_V/e_B)$ in polar wind (from Solar Wind Ten Proceedings, by courtesy of the American Institute of Physics)



of such a structure in accelerating the turbulence evolution has been stressed by (Bavassano et al. 1998) using Ulysses data at mid-latitudes with strong gradients in the wind velocity. It should be also mentioned that the turbulence appears younger in polar flow, since fluctuations at a given distance have had less time to evolve due to the higher wind speed [e.g., see Matthaeus et al. (1999)].

The radial variation of the Elsässer ratio r_E and the Alfvén ratio r_A in polar wind is shown in Fig. 11. Both trends, increasing for r_E and decreasing for r_A, do not go beyond some limits, in agreement with low-latitude observations.

Figure 12 combines Ulysses observations in the polar wind with those by Helios 1 and 2 within the trailing edge of low-latitude fast streams at about 0.4 and 0.8 AU

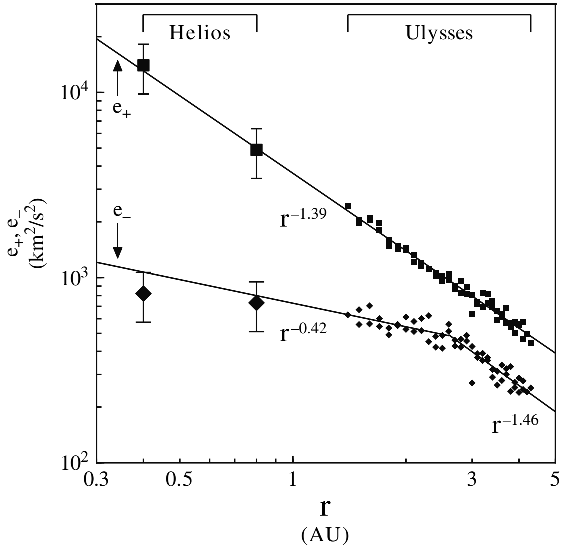


Fig. 12 A composite plot combining Ulysses observations in polar wind with those by Helios 1 and 2 inside 1 AU in the ecliptic plane [adapted from Bavassano et al. (2000), copyright 2000 American Geophysical Union, modified by permission of American Geophysical Union]. The values of e_+ (e_-) are shown as squares (diamonds), small for Ulysses and large for Helios. Best fit lines and radial power laws for Ulysses data are given

(Bavassano et al. 2000). As regards the polar wind, the e_+ values exhibit the same radial gradient over all the investigated range of distances, while for e_- a change of slope around 2.5 AU is clearly apparent. A remarkable point is that the Ulysses best fit lines are in good agreement with the Helios data.

6 Solar Wind Turbulence Models

A well established property of the Alfvénic turbulence is that, in the presence of an energy imbalance between $\mathbf{z}_+$ and $\mathbf{z}_-$ modes, the non-linear interactions act in such a way to lead the turbulence towards a state in which the only left modes are those initially prevailing (Dobrowolny et al. 1980). In this so-called aligned state the minority component has completely disappeared and a total alignment between velocity and magnetic fluctuations has been established. In the case of the solar wind, even though an energy imbalance in favour of outgoing $\mathbf{z}_+$ modes exists, the turbulence does not end with an aligned state. The only way to avoid the disappearance of the minority ingoing $\mathbf{z}_-$ modes is that one or more processes act to refill the solar wind with that kind of mode.

Several models have been proposed for turbulence generation in interplanetary space. Both non-linear processes at velocity gradients and plasma instabilities have been invoked to drive the turbulence evolution. The interaction between Alfvénic fluctuations and convective solar wind structures has also been considered. Velocity gradients are certainly relevant in low- and mid-latitude wind, while in polar wind other processes could be of primary importance, with a robust candidate represented by the parametric decay. Some of the most popular models for turbulence generation and evolution in the different solar wind regimes will be now discussed.

6.1 Models for Low-Latitude Turbulence

Velocity gradients can generate turbulence and drive the radial evolution of $\mathbf{z}_+$ and $\mathbf{z}_-$ spectra. In Fig. 13 the results of a simulation (Roberts et al. 1991) are shown [for other references on this subject see Tu and Marsch (1995)]. The velocity shear

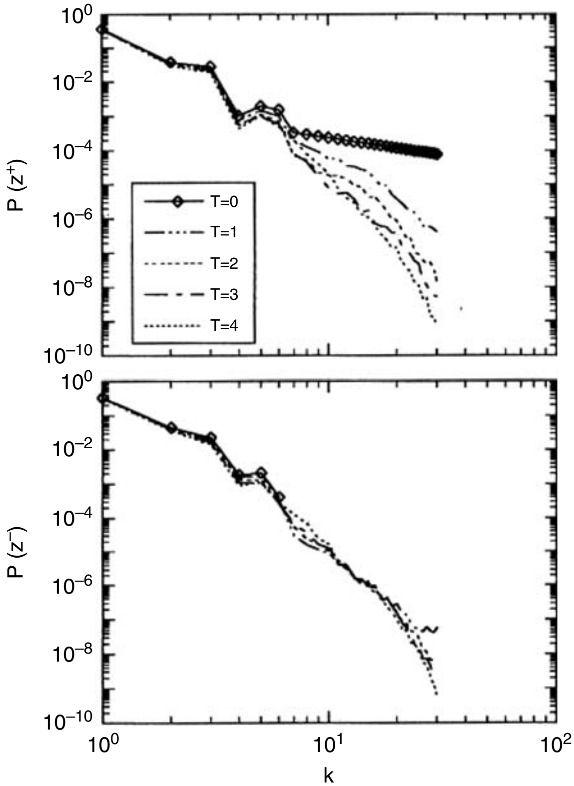


Fig. 13 Evolution of $\mathbf{z}_+$ (*top*) and $\mathbf{z}_-$ (*bottom*) power spectra as driven by a velocity shear [adapted from Roberts et al. (1991), copyright 1991 American Geophysical Union, modified by permission of American Geophysical Union]. Time evolution in the simulation corresponds to radial evolution in the solar wind, with T = 3 roughly equivalent to 1 AU

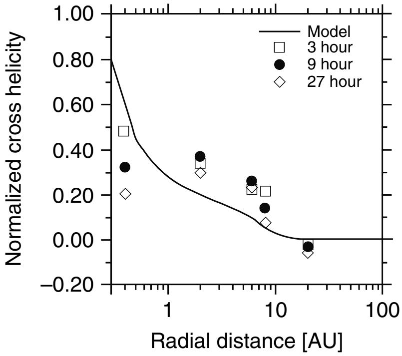


Fig. 14 The radial variation of the normalized cross-helicity as modeled by Matthaeus et al. (2004) is compared to observations by Helios (inside 1 AU) and Voyagers [adapted from Matthaeus et al. (2004), copyright 2004 American Geophysical Union, modified by permission of American Geophysical Union]

leads to an approximately equal injection of power in $\mathbf{z}_+$ and $\mathbf{z}_-$ at large scale. Starting with essentially no power for $\mathbf{z}_-$ at high wavenumbers, this injection leads to a $\mathbf{z}_-$ spectrum that quickly grows until is balanced by dissipation. Thus, the $\mathbf{z}_-$ spectrum is already well developed near the Sun. In contrast, the injection of $\mathbf{z}_+$ is inadequate to balance the dissipation of initially strong $\mathbf{z}_+$ fluctuations. Thus, the $\mathbf{z}_+$ spectrum decreases rapidly. Subsequently, when the $\mathbf{z}_+$ level is close to that of $\mathbf{z}_-$, the two spectra gradually evolve toward each other. This is strongly reminiscent of the spectra variation observed by Helios in the inner heliosphere. The simulations also indicate that magnetic field reversals, common in low-latitude wind, speed up the evolution.

More recently the cross-helicity radial decrease has been modeled as caused by velocity gradients and pickup-ion effects (Matthaeus et al. 2004). Velocity gradients drive large-scale nonlinear Kelvin-Helmholtz instabilities that inject kinetic energy only (hence without contributions to cross-helicity, that would request also a magnetic term). As regards pickup ions, in the outer heliosphere they encounter an approximately transverse magnetic field and therefore couple equally to both $\mathbf{z}_+$ and $\mathbf{z}_-$. Thus, driving supplies energy but not cross-helicity. In Fig. 14 the model predictions are compared to the observational data, as obtained by Helios and Voyagers inside and outside 1 AU, respectively. The model accounts reasonably well for the observed radial trend of σ_C, in other words driving by velocity shear and pickup-ions appears able to overcome the inherent tendency for MHD turbulence to produce aligned Alfvénic states.

An alternative way of describing the radial evolution of solar wind Alfvénic fluctuations is that proposed by Tu and Marsch (1993), based on a two-component model [see also Schmidt (1995); Schmidt and Marsch (1995)]. In this approach the fluctuations are described as a mixture of (1) Alfvén waves, created near the coronal base and propagating outward along the magnetic field lines, and (2) static magnetic structures convected by the expanding solar wind. Acomparison between predicted

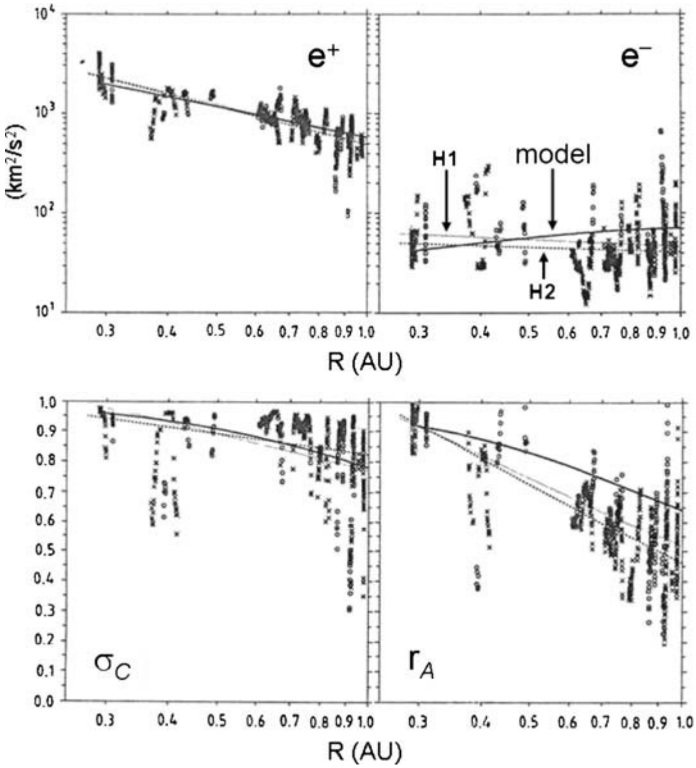


Fig. 15 Predictions (*heavy lines*) of the two-component (waves plus magnetic structures) model (Tu and Marsch 1993) are compared to the radial trends observed by Helios 1 and 2 (*light lines*) inside 1 AU [adapted from Tu and Marsch (1993), copyright 1993 American Geophysical Union, modified by permission of American Geophysical Union]

and observed radial trends for e_+, e_-, σ_C, and r_A is done in Fig. 15. In spite of the data dispersion a reasonable agreement appears to exist.

6.2 Models for Polar Turbulence

Though weaker than in low-latitude wind, polar wind velocity gradients could still give a non-negligible contribution in generating turbulence. However, it is very likely that other processes take a leading role, with the parametric decay as the most robust candidate.

As regards velocity gradients, the occurrence of cross-helicity decreases in connection with polar wind microstreams was pointed out by Neugebauer et al. (1995). Recently, the above model based on velocity gradients and pickup-ion effects

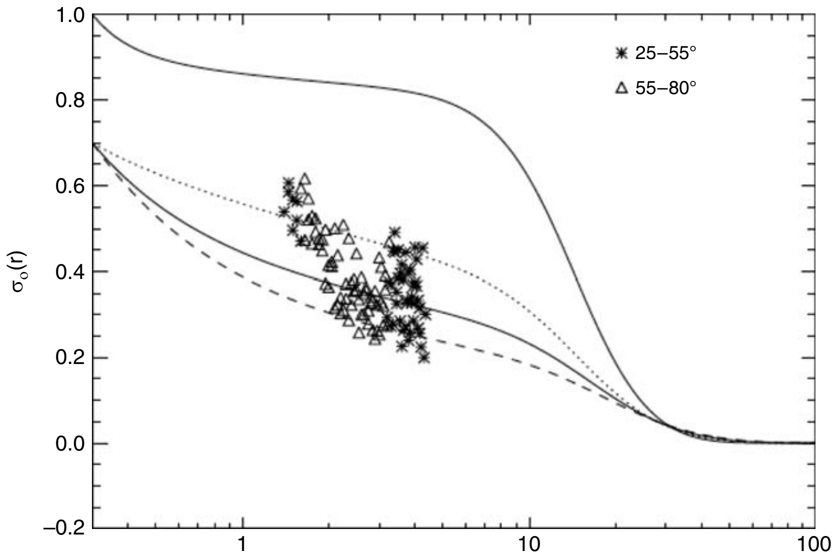


Fig. 16 The normalized cross-helicity at high latitudes as modeled by Breech et al. (2005) is compared to Ulysses observations. The different curves are for different conditions at the inner boundary [adapted from Breech et al. (2005), copyright 2005 American Geophysical Union, modified by permission of American Geophysical Union]

(Matthaeus et al. 2004) has been adapted to polar wind conditions (Breech et al. 2005). Using lower shear, higher wind speed, and lower density, a reasonable tuning of the model parameters is able of leading to solutions behaving in a way comparable to that of the Ulysses observations (Fig. 16).

As for the parametric decay, it has been proposed [e.g., see review Bavassano et al. (2004)] that a relevant role in driving the cross-helicity evolution in polar wind can be played by this process (an Alfvénic wave decay that essentially transfers energy to a backward Alfvénic mode and to a compressive mode). Simulations for the case of large-amplitude non-monochromatic Alfvénic fluctuations (Malara et al. 2000) have shown that the final state strongly depends on the value of β (thermal to magnetic pressure ratio). For $\beta < 1$ the normalized cross-helicity σ_C decreases, from an initial value of 1, to values close to 0. Thus, the instability appears able to completely destroy the initial Alfvénic correlation. In contrast, for $\beta = 1$ (a value closer to real solar wind conditions) σ_C remains different from 0 in the final state. In this case the parametric instability is not able to go beyond some limit in the disruption of the initial correlation between velocity and magnetic field fluctuations. This kind of solution, shown in Fig. 17, is reminiscent of the radial trend seen by Ulysses.

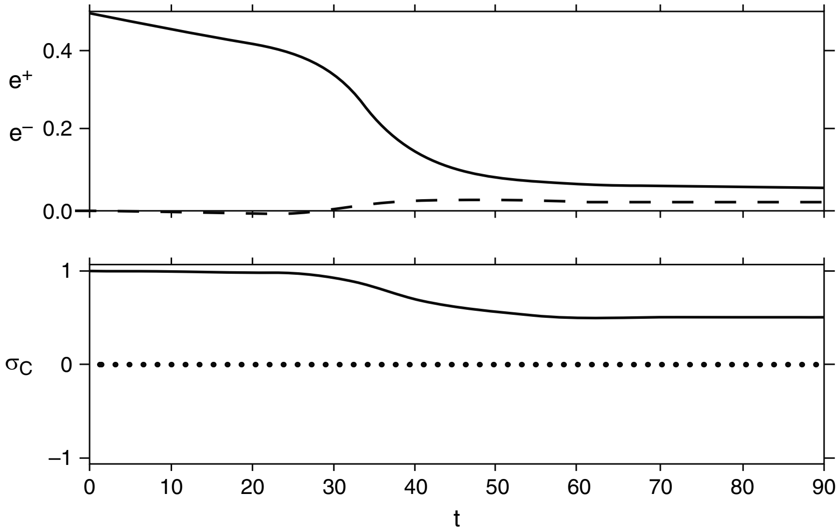


Fig. 17 Time evolution of e_+ and e_- (*top*, *solid and dashed lines*, respectively) and σ_C (*bottom*) for a parametric decay with $\beta = 1$, as modeled by Malara et al. (2000) [adapted from Malara et al. (2000), copyright 2000 American Geophysical Union, modified by permission of American Geophysical Union]

7 Final Remarks

The heliosphere is the only region of space in which in-situ measurements of particles and fields have been performed. This offers a unique opportunity for studying fundamental processes common to astrophysics and plasma physics (for instance, magnetic reconnection, particle acceleration, turbulence generation, plasma heating). Here we have discussed MHD fluctuations observed in the solar wind, a plasma flow resulting from the expansion of the solar corona. The solar wind is an excellent laboratory to investigate processes in a collisionless plasma.

In the above sections it has been shown that solar wind fluctuations at MHD scales have a strong Alfvénic component, with a predominance of fluctuations traveling (in the plasma frame) away from the Sun. However, non-negligible contributions to the wind variability also come from variations of magnetic type. All occurs in close-to-balance conditions between thermal and magnetic pressures [e.g., see Bavassano et al. (2004); Vellante and Lazarus (1987)].

Interplanetary, or local, generation of fluctuations is seen to overcome the inherent tendency for MHD turbulence to produce aligned Alfvénic states. The main candidates for solar wind turbulence driving are velocity gradients and parametric decay.

In an alternative approach, the observations may be explained in terms of a mixture of (1) Alfvén waves, created near the coronal base and propagating outward along the magnetic field lines, and (2) magnetic structures convected by the expanding wind.

In this regard, it should be stressed that the view of a bimodal character of the solar wind variability at MHD scales, with an Alfvénic turbulence propagating from quiet regions of the Sun and interspersed with highly filamentary structures convected from regions in the inner corona, is recently gaining evidence (Bavassano and Bruno 2006; Bruno et al. 2007; Chapman and Hnat 2007; Ogilvie et al. 2007). Efforts in modeling solar wind fluctuations should strongly address this point.

Finally, it is certainly important to extend in-situ measurements to close-to-Sun regions. This is the aim of future missions as Solar Orbiter and Solar Probe. While the latter will have an extremely low perihelion ($\sim$4 solar radii), but with the drawback of a very quick passage, the former will be collecting data for long periods of time at $\sim$0.2 AU in a nearly-heliosynchronous trajectory. This is a key advantage of Solar Orbiter. From this co-rotational vantage point the temporal and spatial variations of the solar wind can be disentangled unambiguously, enabling us to better understand the links between solar and heliospheric processes. More specifically, this will allow to clearly identify the different components of the solar wind variability.

Acknowledgements We are indebted to N. Gopalswamy, R. Ramesh, and K. E. Rangarajan for the invitation at the Asia-Pacific Regional School on Heliophysical Processes (Kodaikanal Observatory, Indian Institute of Astrophysics).

References

Barnes, A.: In: Parker, E.N., Kennel, C.F., Lanzerotti, L.J. (eds.) Solar System Plasma Physics, vol. 1, p. 249. North-Holland, Amsterdam (1979)
Bavassano, B., Bruno, R.: Ann. Geophys. **24**, 3179 (2006)
Bavassano, B., Dobrowolny, M., Mariani, F., Ness, N.F.: J. Geophys. Res. **87**, 3617 (1982)
Bavassano, B., Pietropaolo, E., Bruno, R.: J. Geophys. Res. **103**, 6521 (1998)
Bavassano, B., Pietropaolo, E., Bruno, R.: J. Geophys. Res. **105**, 15,959 (2000)
Bavassano, B., Pietropaolo, E., Bruno, R.: J. Geophys. Res. **106**, 10,659 (2001)
Bavassano, B., Bruno, R., Carbone, V.: In: Poletto, G., Suess, S.T. (eds.) The Sun and the heliosphere as an integrated system, vol. 317, p. 253. Astrophysics and space science library, Kluwer, Dordrecht (2004)
Belcher, J.W., Davis, L., Jr.: J. Geophys. Res. **76**, 3534 (1971)
Breech, B., Matthaeus, W.H., Minnie, J., Oughton, S., Parhi, S., Bieber, J.W., Bavassano, B.: Geophys. Res. Lett. **32**, L06103 (2005). doi:10.1029/2004GL022321
Bruno, R., Carbone, V.: The solar wind as a turbulence laboratory. Living Rev. Solar Phys. **2** (2005). http://www.livingreviews.org/lrsp-2005-4. Cited 15 March 2008
Bruno, R., Bavassano, B., Villante, U.: J. Geophys. Res. **90**, 4373 (1985)
Bruno, R., DAmicis, R., Bavassano, B., Carbone, V., Sorriso-Valvo, L.: Ann. Geophys. **25**, 1913 (2007)
Chapman, S.C., Hnat, B.: Geophys. Res. Lett. **34**, L17103 (2007). doi:10.1029/2007 GL030518
Coleman, P.J., Jr.: Astrophys. J. **153**, 371 (1968)
D'Amicis, R., Bruno, R., Bavassano, B.: Geophys. Res. Lett. **34**, L05108 (2007). doi:10.1029/2006GL028896

Denskat, K.U., Neubauer, F.M.: J. Geophys. Res. **87**, 2215 (1982)
Dobrowolny, M., Mangeney, A., Veltri, P.: Phys. Rev. Lett. **45**, 144 (1980)
Goldstein, B.E., Smith, E.J., Balogh, A., Horbury, T.S., Goldstein, M.L., Roberts, D.A.: Geophys. Res. Lett. **22**, 3393 (1995)
Horbury, T.S., Tsurutani, B.T.: In: Balogh, A., Marsden, R.G., Smith, E.J. (eds.) The heliosphere near solar minimum: The Ulysses perspective, p. 167. Springer-Verlag, Berlin (2001)
Horbury, T.S., Balogh, A., Forsyth, R.J., Smith, E.J.: Geophys. Res. Lett. **22**, 3401 (1995)
Horbury, T.S., Balogh, A., Forsyth, R.J., Smith, E.J.: Astron. Astrophys. **316**, 333 (1996)
Malara, F., Primavera, L., Veltri, P.: Phys. Plasmas **7**, 2866 (2000)
Marsch, E., Tu, C.-Y.: J. Geophys. Res. **95**, 8211 (1990)
Matthaeus, W.H., Goldstein, M.L., Roberts, D.A.: J. Geophys. Res. **95**, 20,673 (1990)
Matthaeus, W.H., Zank, G.P., Smith, C.W., Oughton, S.: Phys. Rev. Lett. **82**, 3444 (1999)
Matthaeus, W.H., Minnie, J., Breech, B., Parhi, S., Bieber, J.W., Oughton, S.: Geophys. Res. Lett. **31**, L12803 (2004). doi:10.1029/2004GL019645
McComas, D.J., Bame, S.J., Barraclough, B.L., Feldman, W.C., Funsten, H.O., Gosling, J.T., Riley, P., Skoug, R., Balogh, A., Forsyth, R., Goldstein, B.E., Neugebauer, M.: Geophys. Res. Lett. **25**, 1 (1998)
McComas, D.J., Elliot, H.A., Gosling, J.T., Reisenfeld, D.B., Skoug, R.M., Goldstein, B.E., Neugebauer, M., Balogh, A., Geophys. Res. Lett. **29**(9) (2002). doi:10.1029/2001GL014164
McComas, D.J., Elliott, H.A., Schwadron, N.A., Gosling, J.T., Skoug, R.M., Goldstein, B.E.: Geophys. Res. Lett. **30**(10), 1517 (2003). doi:10.1029/2003GL017136
Neugebauer, M., Goldstein, B.E., McComas, D.J., Suess, S.T., Balogh, A.: J. Geophys. Res. **100**, 23,389 (1995)
Ogilvie, K.W., Coplan, M.A., Roberts, D.A., Ipavich, F.: J. Geophys. Res. **112**, A08104 (2007). doi:10.1029/2007JA012248
Roberts, D.A., Goldstein, M.L., Klein, L.W.: J. Geophys. Res. **95**, 4203 (1990)
Roberts, D.A., Ghosh, S., Goldstein, M.L., Matthaeus, W.H.: Phys. Rev. Lett. **67**, 3741 (1991)
Schmidt, J.M.: Ann. Geophys. **13**, 475 (1995)
Schmidt, J.M. Marsch, E.: Ann. Geophys. **13**, 459 (1995)
Smith, E.J., Balogh, A., Neugebauer, M., McComas, D.J.: Geophys. Res. Lett. **22**, 3381 (1995)
Tu, C.-Y., Marsch, E.: J. Geophys. Res. **98**, 1257 (1993)
Tu, C.-Y., Marsch, E.: Space Sci. Rev. **73**, 1 (1995)
Vellante, M., Lazarus, A.J.: J. Geophys. Res. **92**, 9893 (1987)

Radio Emission Processes: Part I

K.R. Subramanian

Abstract This chapter provides an overview of the radio emission mechanisms in the Solar atmosphere. Incoherent (free - free emission and gyro - emission) and coherent (plasma emission and electron cyclotron maser emission) emission mechanisms are discussed.

1 Introduction

Radio emission from the Sun is produced incoherently or coherently. Incoherent emission consists of free–free emission (bremsstrahlung) and gyro – emission. Plasma emission and electron cyclotron maser emission are the main radio emission processes in the case of coherent radio emission. Incoherent emission results from continuum processes, such as thermal particle distribution, that produce through Columb collisions or thermal/mildly relativistic electron distribution that produce through gyroresonance/gyrosynchrotron emission. Coherent emission is produced by kinetic instabilities from unstable particle distribution. No spectral lines in emission or absorption due to atomic/molecular transitions have been discovered in the radio frequency band on the Sun.

2 Bremsstrahlung

The term bremsstrahlung means braking radiation in German and is the electromagnetic radiation produced by the acceleration of a charged particle like an electron in the Coulomb field of ambient ions.

In a fully ionized plasma, electrons and ions move freely interacting with each other through their electrostatic charges. The most important interaction is the

K.R. Subramanian
Indian Institute of Astrophysics, Koramangala, Bangalore – 34, India
e-mail: subra@iiap.res.in

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_9,

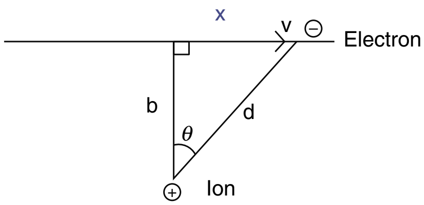


Fig. 1 An electron passing by a heavy ion. Due to weak scattering, low energy radio photons are produced. The distance of closest approach is called the impact parameter b. The distance between the ion and the electron is d and θ is the angle between b and d

scattering of a free electron in the Coulomb field of an ion. Since the electron and the ion are free particles before and after the interaction, the radiation emitted is called as free–free radiation. A photon is emitted with an energy corresponding to the difference of the outgoing to the incoming kinetic energy of the electron according to the principle of conservation of energy. Figure 1 shows an example of a binary collision between an electron of *charge* $+\, e$ and velocity v with an ion of charge Z_i. The effect of this collision is to deflect the incoming electron by an amount which depends on the impact parameter b. Due to the small angle collisions, the trajectory of the incoming electron is determined by a multitude of small deviations. The bremsstrahlung emission from a single electron is calculated using the equations given in Rybicki and Lightman (1979). It is assumed that the electron energy in the initial stage is greater than energy lost by the electron when it interact with the ion. For radio waves, distant encounters of electrons with ions called small angle deflections are more important than the relatively close encounters and large deflections.

When an electron of charge e passes by an atom having Z ions, due to the electrostatic force, the electron is accelerated both parallel and perpendicular to the direction of electron's velocity. Emission due to acceleration in the perpendicular direction is in the radio band. The force in the perpendicular direction is given by

$$F_{\perp} = m_e a_{\perp} = \frac{Ze^2 \cos(\theta)}{d^2} = \frac{Ze^2 \cos^3(\theta)}{b^2}, \tag{1}$$

where d is the distance between the electron and the ion, b is the impact parameter and m_e is the mass of the electron. According to Larmor's formula, the total power radiated by the electron is $\propto a^2$ and is given by

$$P = \frac{2}{3}\frac{e^2 a^2}{c^3}, \tag{2}$$

where a is the acceleration, e is the electric charge, and c is the speed of light. The total energy radiated by the electron in time dt is

$$W = \int_{\infty}^{\infty} P dt, \tag{3}$$

This energy is emitted in a single pulse of duration $\tau = \frac{b}{v}$, so that the power spectrum is flat over all frequencies upto a maximum frequency $\nu_{\max} = \frac{1}{2\pi\tau} = \frac{v}{2\pi b}$ and falls of rapidly at high frequencies. Assuming the power spectrum is flat up to frequency $\nu = \frac{v}{2\pi b}$, we get the average energy per unit frequency emitted by a single electron as $W_\nu = \frac{W}{\nu_{max}}$. Therefore W_ν is given by

$$W_\nu = \frac{\pi^2}{2} \frac{Z^2 e^6}{c^3 m_e^2} \left[\frac{1}{b^2 v^2} \right] \tag{4}$$

In an ensemble of particles, the number of electrons passing any ion per unit time with impact parameter b to $b+db$ and speed ranging from v to $v + dv$ is $N_e(2\pi b db) v f(v) dv$ where $f(v)$ is the normalised speed distribution of electrons. The number $N(v, b)$ of such encounters per unit volume/unit time is given by $2\pi b db(v f(v) dv) N_e N_i$. where N_e and N_i are the electron and ion number densities. The isotropic spectral power emitted at frequency ν is given by $4\pi\epsilon_\nu$ where ϵ_ν is the emission coefficient.

$$4\pi\epsilon_\nu = \int_{b=0}^{\infty} \int_{v=0}^{\infty} W_\nu(v, b) N(v, B) dv db, \tag{5}$$

With the derived value of W_ν the above equation becomes

$$\frac{\pi^3 Z^2 e^6 N_e N_i}{c^3 m_e^2} \int_{v=0}^{\infty} \frac{f(v)}{v} dv \int_{b=0}^{\infty} \frac{db}{b}, \tag{6}$$

In the above equation the term $\displaystyle\int_{b=0}^{\infty} \frac{db}{b}$ diverges. Therefore let us consider the limits of b from $b_{\min}$ to $b_{\max}$. It is convenient to write $\ln\left(\frac{b_{\min}}{b_{\max}}\right)$ as $G_{ff}(v, \nu)$ which is called as Gaunt factor (Karzas and Latter 1961). In the case of a non relativistic Maxwellian velocity distribution. The emission coefficient is

$$\epsilon_\nu = \frac{\pi^2 Z^2 e^6 N_e Ni}{4 c^3 m_e^2} \left(\frac{2 m_e}{\pi k T}\right)^{3/2} ln\left(\frac{b_{max}}{b_{min}}\right), \tag{7}$$

Assuming local thermo dynamic equilibrium (LTE) at temperature T the absorption coefficient k_ν can be related to the emission coefficient and the brightness by the Kirchoff's law by

$$k_\nu = \frac{e_\nu}{B_\nu(T)} = \frac{\epsilon_\nu c^2}{2kT\nu^2}, \tag{8}$$

The emission coefficient is therefore

$$k_\nu = \frac{1}{\nu^2 T^{3/2}} \left[\frac{Z^2 e^6 N_e N_i}{c} \right] \left[\frac{1}{\sqrt{2\pi m_e^3 k^3}} \right] \frac{\pi^2}{4} ln \left(\frac{b_{max}}{b_{min}} \right), \tag{9}$$

The properties of the radiation field is modified when it propagates through a medium to the observer and is described by the radiative transfer equation (Kraus 1966):

$$\frac{dI_\nu}{ds} = \eta_\nu - k_\nu I_\nu, \tag{10}$$

The term η_ν is a source term while the term $k_\nu I_\nu$ is an absorption term. The later is used to define the optical depth along the line of sight, i.e. $\tau_\nu = \int_0^s k_\nu ds$
The solution of (10) is of the form

$$I_\nu = I_{\nu(0)} e^{-\tau_c} + \frac{\eta_\nu}{k_\nu}(1 - e^{-\tau_c}), \tag{11}$$

In the case of a plasma cloud sitting between the source and the observer, the observed brightness is the combination of the source and contribution from the plasma cloud. Expressing the intensity in terms of temperature, we get

$$T_b = T_s e^{-\tau_c} + T_c(1 - e^{-\tau_c}) \tag{12}$$

The optical depth in the case of free–free emission is

$$\tau_\nu = \int -k_\nu ds \propto \int \frac{N_e N_i}{\nu^2 T^3/2} ds \approx \int \frac{N_e^2}{\nu^2 T^3/2} ds, \tag{13}$$

From the observed brightness temperature the flux density can be calculated using the equation

$$S = \frac{2k_B}{\lambda^2} \int T_b d\Omega, \tag{14}$$

where k_B is the Boltzmann's constant, λ is the wavelength and T_b is the brightness temperature and $d\Omega$ is the size of the source. Since the absorption coefficient scales with square of the density, the free–free emission is dominant in regions of high density.

3 Gyro-Emission

When a non-relativistic electron gyrates about the magnetic field, cyclotron radiation is emitted at the frequency of gyration called the gyrofrequency and is given by $f_B = eB/2\pi m_e c = 2.8B$ MHz where B is in Gauss. The electron gyrofrequency is independent of velocity and therefore independent of the kinetic energy of the electron ($E = \frac{1}{2}m_e v^2$ when v≪c). Since the gyrofrequency is single valued it really represents monochromatic or line radiation rather than continuum radiation. For low velocity electrons ($v < 0.03c$), only a single gyrofrequency emission occur. Gyroresonance emission involves a resonance between the electromagnetic waves and electron spiralling along the magnetic field lines at the electron gyrofrequency. This resonance produces strong coupling between the electrons and radiation at low harmonics of the gyrofrequency ($f = sf_B$ ($s = 1, 2, 3, i \ldots$)). At mildly relativistic speeds (electron energies 100–300 keV), 10–100 harmonics of the cyclotron frequency occur, and overlap to produce a continuous spectrum. This form of emission is called gyrosynchrotron emission. At highly relativistic speeds, the emission occurs at many harmonics so close together that emission looks like a continuum emission. This type of emission is called synchrotron emission.

4 Synchrotron Emission

The general Larmor's equation is a nonrelativistic one and and are correct only in the inertial frames where the electron velocity v≪c. Using the Lorentz transform, the results can be transformed into other frames. We now calculate the total power radiated by a ultra relativistic electron in a magnetic field parallel to the x-axis. The primed coordinates are used to describe an inertial frame in which the electron is nearly at rest. The Larmor equation in this case is given by

$$P' = \frac{2e^2 \left(a'_\perp\right)^2}{3c^3}, \tag{15}$$

where $a'_\perp$ is the magnetic acceleration in the observer's frame. We have

$$a_\perp = \frac{a'_\perp}{\gamma^2}, \tag{16}$$

The Larmor equation now can be written as

$$P' = \frac{2e^2}{3c^3}\left(a'_\perp\right)^2 = \frac{2e^2 a_\perp^2 \gamma^4}{3c^3}, \tag{17}$$

Since $a_\perp = \omega v_\perp$ and $\omega = \frac{eB}{\gamma mc}$ in the relativistic case, the time averaged power radiated by an electron is given by

$$P = \frac{2e^2}{3c^3}\gamma^2 \frac{e^2 B^2 v^2 sin^2\theta}{m^2c^2}, \tag{18}$$

where θ is the angle between the **v** and **B** and is called the pitch angle.

The above power can be expressed in terms of σ_T, Thomson cross section of an electron ($\sigma_T = \frac{8\pi}{3}\left(\frac{e^2}{m_e c^2}\right)^2$) and the magnetic energy density ($U_B = \frac{B^2}{8\pi}$)

The power radiated by a single electron is therefore

$$P = \left[\frac{8\pi}{3}\left(\frac{e^2}{m_e c^2}\right)^2\right] 2\left(\frac{B^2}{8\pi}\right) c\gamma^2 \frac{v^2}{c^2} sin^2\alpha, \tag{19}$$

The average power $< P >$ per electron in an ensemble of electron with the same γ, but random pitch angle is

$$< P >= 2\sigma_T \beta^2 \gamma^2 c U_B < sin^2\alpha >= 2\sigma_T \beta^2 \gamma^2 c U_B \frac{2}{3}, \tag{20}$$

The radiated power depends on the physical constants, square of the electron energy, the magnetic energy density and the pitch angle. Synchrotron emission is strongly beamed along the direction of motion which turns out to be perpendicular to the acceleration. The emission is concentrated into an angle along the direction of motion of order $1/\gamma$. Since $1/\gamma$ is $\ll 1$, the radiation is confined into a narrow beam of width $2/\gamma$ between the nulls as shown in the Fig. 2. But for any reasonable distribution of particles that varies smoothly with the pitch angle, the elliptical components will cancel out as emission cones will be partially linearity polarized.

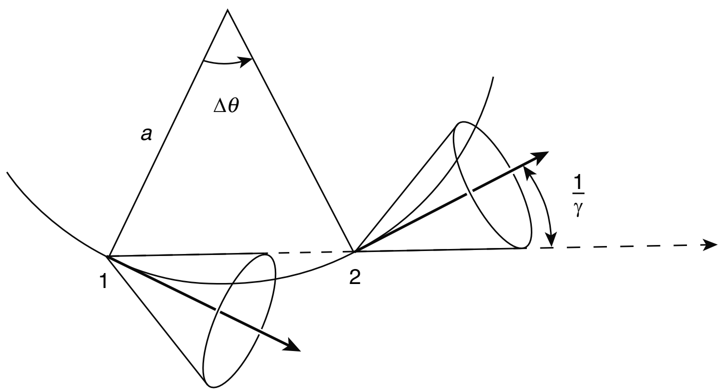


Fig. 2 Emission comes at various points of an accelerated particle's trajectory

4.1 Synchrotron Spectrum

The beaming of the radiation has a very important effect on the observed spectrum emitted by the electrons. As the electron cycles around the helical path along the magnetic field lines, any emission directed towards a distant observer is seen only when the beam is aligned with the observer's line of sight. In this case the observer sees a flash of radiation for a period which is much shorter than the gyration period. The over all spectrum of emission is the Fourier transform of the time series of pulses. Thus as a function of time, the power $P(t)$ emitted by an electron is a succession of pulses of width τ separated by $2\pi\gamma/\omega_{ce}$ (ω_{ce} is called gyro frequency): i.e. occurring with the frequency $\omega_{ce}/2\pi\gamma$.

The volume emissivity (power per unit frequency per unit volume per unit solid angle) of relativistic electrons is given by

$$\epsilon_\nu = \int_E P(\nu, E)N(E)dE, \tag{21}$$

where $P(\nu, E)$ is the total power from that one electron with energy E radiates and $N(E)\,dE$ is the number of electrons per unit volume and per unit solid angle moving in the direction of the observer and whose energy lies in the range of E and $E + dE$. The particle energy distribution can be either Maxwellian or power law. In the first case the dominant mechanism is the gyroresonance emission at discrete and low harmonic number of ω_{ce} for which emissivity can be written as a function of the density n_e, the magnetic field strength B and its direction relative to the observer. In the case of nonthermal power law, the energy distribution of electrons is of the form $AE^{-\delta}$. The gyrosynchrotron emissivity then has a power law spectrum which can be used to estimate the injected particle distribution as well as the strength of the magnetic field B. Calculations of the emissivity for an isotropic velocity distribution of electrons in a homogeneous magnetic field have been performed by Ramaty (1969).

Let us consider the case of synchrotron emission from N electrons with the same velocity and pitch angle. The power radiated is

$$P = \frac{dE}{dt} = \frac{4}{3}\sigma_T \beta^2 \gamma^2 c U_B \tag{22}$$

at the single frequency $\nu = \gamma^2 \nu_G$. The emission coefficient from an ensemble of electrons is

$$\epsilon_\nu d\epsilon_\nu = \frac{dE}{dt} N(E)dE, \tag{23}$$

$$\text{since } \; E = \gamma m_e c^2 = \left(\frac{\nu}{\nu_G}\right)^2 m_e c^2, \tag{24}$$

The emission coefficient is

$$\epsilon_\nu = \frac{4}{3}\sigma_T \beta^2 \gamma^2 c U_B K E^{-\delta} \frac{m_e c^2 \nu^{-1/2}}{2{\nu_G}^{1/2}}, \tag{25}$$

Synchrotron spectrum of a power law itself is a power law.

5 Summary

We have discussed a few radio emission processes in this chapter. The free–free bremsstrahlung radiation mechanism is responsible for the quiet Sun radio emission. The intensity of emission depends on the atmospheric layer because of the dependence on the free–free opacity, the plasma density and the local temperature. Using typical temperature of 10^6 K and density of $10^9\,\mathrm{cm}^{-3}$ we get the optical depth $\tau \ll 1$ in the whole microwave range. At decameter wavelengths, the corona is optically thick and ray – tracing calculations are generally used to determine the temperature and density of the corona (Subramanian 2004). In the absence of flares, the the radio emission of solar active regions is due to the free–free emission. Gyroresonance emission at the low harmonics ($s = 1, 2, 3, \ldots$) is responsible for bright coronal emission from the Sun in the microwave band. Gyrosynchrotron emission is commonly observed as a broadband microwave spectrum in a typical frequency range of 2–20 GHz, and the spectrum of gyrosynchrotron emission peaks around 5–10 GHz. Most of the radio emission from solar and stellar flares are due to gyrosynchrotron emission. Synchrotron emission is important in extreme energy environments like extragalatic radio sources.

References

Karzas, W.J., Latter, R.: Astrophys. J. Suppl. S. **6**, 167 (1961)
Kraus, J.D.: Radio Astronomy. Cygnus-Quasar Books, Ohio (1996)
Ramaty, R.: Astrophys. J. **158**, 753 (1969)
Rybicki, G.B., Lightman, A.P.: Radiative Processes in Astrophysics. Wiley, New York (1979)
Subramanian, K.R.: Astron. Astrophys. **426**, 329 (2004)

Radio Emission Processes: Part II

1 Plasma Emission

Plasma emission process is the most important coherent process to produce solar radio bursts in metric wavelengths. The important feature of the coherent emission process is that the brightness temperature of the emitted radiation is much higher ($T_B \gg 10^{10}$ K) than can be explained by the incoherent emission. The decametric radio emission from Jupiter and auroral kilometric emission (AKR) from Earth are also due to the coherent plasma emission process.

In a plasma a wide range of wave oscillations and modes can exist. Plasma emission is an indirect emission process when a beam of electrons propagate from the flare site outwards towards the corona. As the beam of electrons propagate, they excite plasma oscillations in different layers of the Sun with increasing height. The generation of plasma waves by this process is called Langmuir turbulence. Because the plasma waves are longitudinal electrostatic waves, they can not directly produce electromagnetic radiation, and the electromagnetic waves are generated through some secondary process. The basic process is that the free energy available in the Langmuir turbulence is converted into electromagnetic radiation by scattering of the plasma waves on the ions or by coalesclence of two waves. Radiation can occur at the fundamental and harmonic frequency. The fundamental frequency is related to the local plasma frequency of the medium and is given by

$$\nu_p = \frac{e}{2\pi}\sqrt{\frac{N_e}{\epsilon_0 m}}, \tag{1}$$

where N_e is the electron density and m is the mass of the electron and ϵ_o is the permittivity. The numerical value for the plasma frequency is given by $\nu_p = 9{,}000\sqrt{(N_e)}$ Hz, where N_e is in cm^{-3}. The first step in the production of plasma emission is the generation of Langmuir waves. Due to the wave - particle interaction, the wave can grow by getting energy from the particle, if the velocity of the particle is greater than the phase velocity of the plasma wave. For particle with a velocity greater than the phase velocity of the wave, the wave-particle interactions generate Langmuir turbulence, and reduce the velocity of the particle. This instability is called bumb-in-tail instability. Such electrons have a velocity distribution with a positive slope $\left(\frac{df(v_z)}{dv_z} > 0\right)$ near v_ϕ as shown in Fig. 1. The mutual interaction of the waves and the particles is described by a pair of quasilinear equations. The feed back on the distribution of particles referred to as a quasi-linear relaxation involves smoothing out of the bump near v_ϕ to form a plateau $\frac{df(v_z)}{dv_z} \approx 0$. It was pointed by by Sturrock (1964) that in a homogeneous beam – plasma model, the beams should propagate only a few hundred kilometers before losing all their

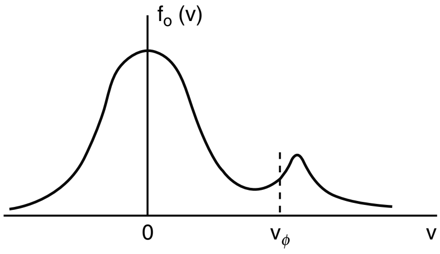


Fig. 1 Beam instability

energy to the Langmuir waves. But it is well known from observations that electron beams propagate through the corona, and some of them are known to propagate beyond the orbit of Earth still generating Langmuir waves and radiation. A number of theories have been suggested to over come this problem and to stabilize the beam. One is based on the saturation of the instability due to inhomogenities in the beam. Another way is the influence of the ambient density fluctuations on the growth rate of the beam – plasma instability.

Plasma emission is a multistage process and involves (a) formation of an unstable beam distribution by velocity distribution, (b) generation of Langmuir turbulence and (c) conversion of plasma waves into electromagnetic radiation. The dispersion relation between the angular frequency ω and the wave vector k for Langmuir waves is given by

$$\omega^2(k) = \omega_p^2 + \frac{3}{2}v_{th}^2, \tag{2}$$

where v_{th} is the thermal velocity and ω_p is 2π times the plasma frequency. For the above equation, the solution for ω is of the form $\omega = \omega_r \pm i\omega_i$. The variation of the amplitude with time is of the form $e^{-i\omega t} = e^{-i\omega_r t}e^{\Gamma t}$, where $\Gamma = \omega_i$. $\Gamma > 0$ means exponential growth of the perturbation. Electrons with the velocity v_{th} can undergo wave-particle interactions with waves $\omega(k)$ when they fulfill the Doppler resonance condition:

$$\omega - \frac{s\Omega}{\gamma} - k_{\|}v_{\|} = 0, \tag{3}$$

The wave growth $\Gamma(k, f(v))$ or absorption rate $\Gamma < 0$ can be calculated for a given particle velocity distribution function $f(v)$. For nonrelativistic electrons the growth rate is

$$\Gamma(k) = \left(\frac{\pi}{2}\right)\frac{\omega_p^2}{k^2}\left(\frac{\omega}{n_o}\right)\left(\frac{df_o\left(v_{\|}\right)}{dv_{\|}}\right)_{\frac{\omega}{k}} \approx \left(\frac{\pi}{2}\right)\frac{n_b}{n_o}\left(\frac{v_{ph}}{\Delta v}\right)^2 w, \tag{4}$$

where n_o is the ambient electron density, n_b is the beam electron density, and $v_{ph} = \frac{\omega}{k}$ is the phase speed. For every wave – wave interaction, the energy and momentum equations have to be fulfilled, i.e. matching conditions of frequencies and wave vectors:

$$\omega_1 + \omega_2 = \omega_3, \qquad k_1 + k_2 = k_3, \tag{5}$$

For example, a primary Langmuir wave ($\omega_1, \mathbf{k_1}$) can couple with an ion acoustic wave ($\omega_2, \mathbf{k_2}$) to generate a secondary Langmuir wave ($\omega_3, \mathbf{k_3}$). Since the phase speed of the ion acoustic wave is much smaller than for Langmuir waves, the primary and secondary Langmuir waves have similar frequencies and wave vectors. For conversion of plasma waves to electromagnetic waves Melrose (1987) suggested the following wave – wave interactions:

$$\begin{gathered} L + S \rightarrow L', \\ L + S \rightarrow T, \\ L + S \rightarrow L', \\ T + S \rightarrow L, \\ T + S \rightarrow T, \\ L + L' \rightarrow T, \end{gathered}$$

Here L denotes the Langmuir wave, S the ion acoustic wave and T the electromagnetic wave. In the above equations, the first process is important to generate Langmuir turbulence, while the second and third processes generate fundamental plasma emission, the fourth is important for scattering of the transverse waves and the last for the generation of second harmonic radiation. Growth of the Langmuir waves occur until it is limited or saturated owing to (a) the beam of electrons passes or it otherwise changes its characteristics to destroy the resonance, (b) the growing wave may be scattered or altered due to scattering off ions, inhomogenities or low frequency waves. The wave will interact with the streaming electrons diffusing them in the velocity space, and it also removes the positive slope. Figure 2 shows a schematic diagram of the processes involved.

The saturated energy density in the Langmuir waves (W_L) or equivalently the effective temperature (T_L) of the Langmuir waves is defined by

$$W_L = \int \frac{k_B T_L}{(2\pi)^3} d^3\mathbf{k}, \tag{6}$$

where k_B is the Boltzmann's constant. Several studies have shown that the saturated value of W_L is $\approx 10^{-5}$ times the energy density of the background plasma, i.e. $\approx 10^{-5} n k_B T$. Hence the limiting value of T_L is of the order of

$$T_L \approx 10^8 \frac{v_b^2}{c^2} \frac{v_b}{\omega_p} T, \tag{7}$$

which is $\leq 10^{15}$ K.

Solar radio emission at a given frequency ν arise only from a region where the plasma frequency ν_p is equal or lower than ν. Because n_e decreases as a

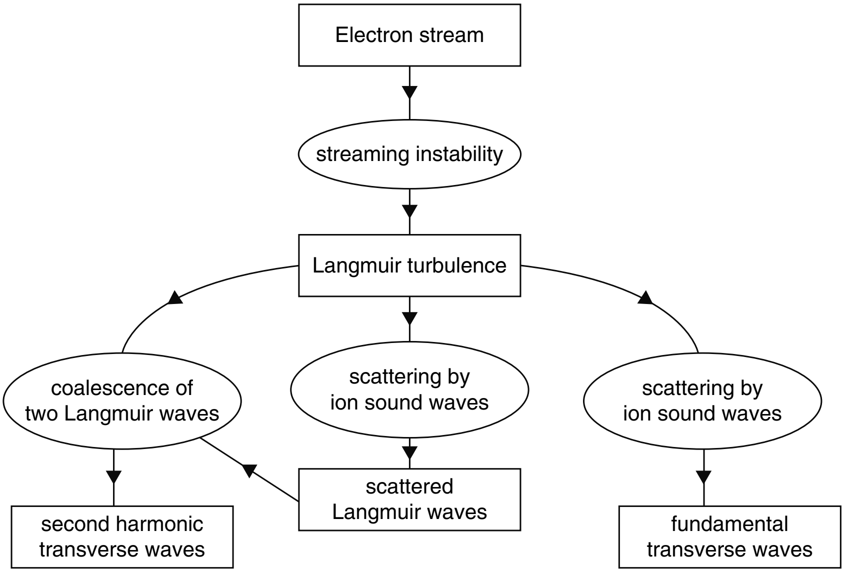


Fig. 2 Theoretical description of emission of electromagnetic radiation by an electron stream in a plasma

function of height in the solar atmosphere, ν_p also decreases with height, and lower frequencies must arise from greater heights.) the beam of electrons passes or otherwise changes its characteristics to destroy the resonance, Electron densities in the range of $n_e = 10^8 - 10^{10}$ cm^{-3} give rise to plasma radiation in the frequency range of 100–1,000 MHz. During a solar flare, many types of radio bursts are generated at frequencies near the plasma frequency or its harmonics and they come from a thin layer above the plasma level, i.e. above the height where $\nu_p = \nu$. In a dynamic spectrum where the intensity is displayed as a function of both frequency and time, the signature of a beam of particles is a sloping emission band. If we know the density of the medium at different heights, we can estimate the speed of the particles in the beam. Figure 3 shows a schematic diagram of an idealized dynamic spectrum frequently produced by large flares. The nomenclature and details of different radio bursts are given in McLean and Labrum (1985).

2 Electron Cyclotron Maser Emission

Magnetospheres of magnetized planets emit nonthermal radiation with an intensity and variability which cannot be explained as s simple cyclotron emission by trapped energetic particles. Since 1980 electron cyclotron maser emission (ECME) has become accepted as a mechanism to explain the radio emission from magnetized planets. MASER is the acrononym for microwave amplification by stimulated

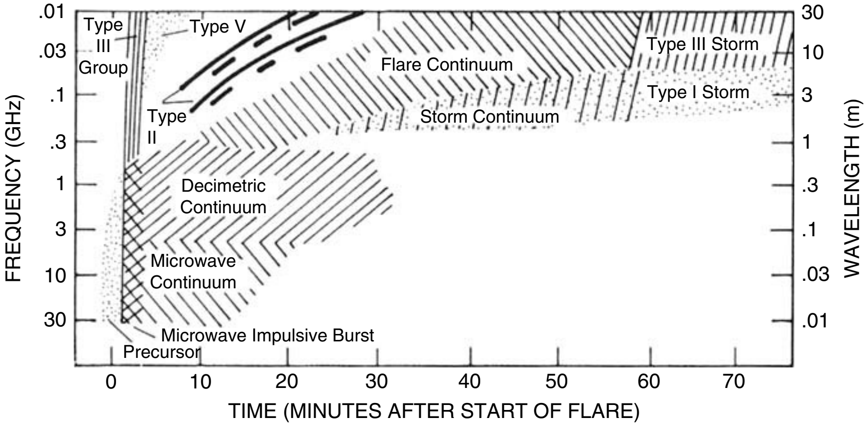


Fig. 3 Idealized solar radio dynamic spectrum after a large flare (Dulk 1985)

emission of radiation. The basic requirements for MASER to operates are: (a) population inversion in the electron distribution as compared with equilibrium, (b) a pump for the MASER and (c) relatively strong magnetic field or low density so that the electron cyclotron frequency ω_{ce} is much greater than the plasma frequency ω_{pe}. The most common form of population inversion in an astrophysical situation is the loss cone distribution produced when electrons are energized in magnetic flux tubes that have converging legs and their foot points are in a high density atmosphere. This condition is satisfied above the auroral zone of Earth, in the magnetosphere of Jupiter and Saturn and in some magnetic flux tubes in the low solar corona. The radio emission from these objects and the spike bursts from the Sun are usually interpreted in terms of ECME. The brightness temperature of these emission is $\geq 10^{15}$ K and the emissions are 100% circularly polarized except for the AKR. In a magnetic field during the acceleration of electrons, the radiated emission is divided into both parallel and perpendicular components. The velocity distribution of the particles can be parallel to the magnetic field $\left(\dfrac{df}{dv_\parallel} > 0\right)$ as in the case of a beam of particles or perpendicular as in the case of a loss cone $\left(\dfrac{df}{dv_\perp} > 0\right)$. The following discussions are based on the work of Dulk (1985).

In the case of the cyclotron MASER emission, the wave – particle resonance condition is given by

$$\omega - k_\parallel v_\parallel = s\omega_e \left(1 - \frac{v^2\|}{c^2} - \frac{v_\perp^2}{c^2}\right)^{-1/2}, \tag{8}$$

From the above equation, we can find that for the generation of radiation at frequencies near ω_{ce}, the term $k_\parallel v_\parallel$ should be small. This happens when the radiation is produced with wave vectors almost perpendicular to the magnetic field. This occurs when the particles are trapped in regions bounded by two sites of high magnetic field like in a loss cone. From the conservation of energy and magnetic moment it can be shown that as a particle moves towards region of high magnetic field, the velocity parallel to the magnetic field direction decreases and the velocity perpendicular to the magnetic field increases. The instability can develop due to this anisotropic velocity distribution and transfer energy directly into electromagnetic radiation through ECME.

Due to the magnetic reconnection in the solar atmosphere, acceleration of electrons takes place, and the electron energy can be divided into components both parallel and perpendicular to the magnetic field. Equation (8) implies that resonance is possible only if $\omega^2 < s^2\omega_{ce}^2 + c^2k_\parallel^2$. Depending on the distribution $f(v_\parallel, v_\perp)$ of electrons in the velocity space, the wave either extracts energy from the resonant electrons and grows or loses energy to them and is damped. In the former case, the intensity of the wave at any point increases exponentially with time until the MASER saturates. The dominant contribution to the growth rate Γ_s for the sth harmonic is given by

$$\Gamma_s \sim \int d^3\mathbf{v}\mathbf{A_s}(\mathbf{v}, \mathbf{k}\frac{\mathbf{df}}{\mathbf{dv_\perp}}\delta(\omega - \mathbf{k_\parallel v_\parallel} - \mathbf{s}\omega_{\mathbf{ce}}/\gamma), \tag{9}$$

where $A_s > 0$. Because of the delta function this corresponds to an integration around the resonance ellipse, and growth can occur only if $\frac{df}{dv_\perp} > 0$. This condition is analogous to the requirement of population inversion for the operation of a LASER. The source of free energy for all the MASER emission is believed to be a single-sided loss cone distribution. This type of distribution is formed when electrons are accelerated downwards along converging magnetic field lines near the surface of a planet or star.

Melrose and Dulk (1982) estimated the brightness temperature of MASER to be

$$\frac{k_B T_B}{mc^2} = \eta \frac{c}{r_0\omega}\frac{\Gamma}{\omega_e}, \tag{10}$$

where r_0 is the classical radius of the electron and η is the fraction of the energy density of the driving electrons converted to MASER radiation. The maximum growth rate for the fundamental x-mode for energetic electron number densities of $\sim 10^{13}$ m^{-3} and energy of 10 KeV near 1 GHz are typically $\sim 10^{-3}$. For values of $\eta \sim 10^{-3}$, the brightness temperature $\sim 10^{17}$ K.

3 Summary

In this chapter we have discussed the coherent radio emission processes, i.e. plasma and electron cyclotron emission mechanisms. In the case of plasma emission, the beam of particles have a non-Maxwellian velocity distribution with a positive slope in the velocity distribution. Losscones which have positive slope in the

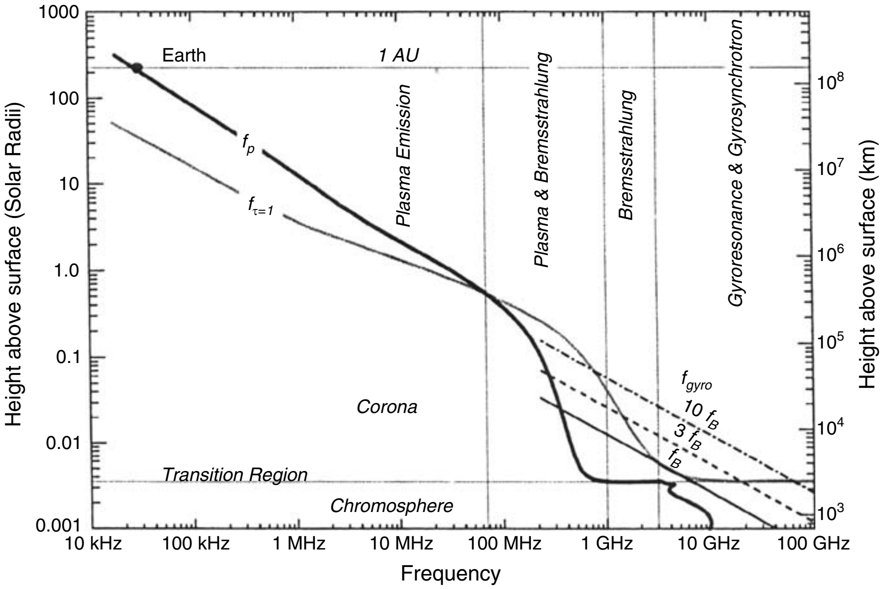


Fig. 4 Regimes of dominant emission in the solar atmosphere as a function of radio frequency (Gary 2004)

perpendicular direction to the magnetic field also produce radio emission coherently. The solar radio bursts (type I, II, II, IV and V) below 300 MHz are usually explained by the plasma emission processes. Coherent plasma emission is also responsible for decimetric radio emission in the band 300–3,000 MHz. The radio bursts provide useful diagnostics of density and magnetic field, complementary to other wavelengths. The ECME is responsible for the radio emission from magnetospheres of magnetized planets like Jupiter. The high brightness temperature ($>10^{15}$ K) of millisecond spike burst observed during solar flares is explained by electron cyclotron emission mechanism. The various emission processes discussed in this and the previous chapter and their occurence in the solar atmosphere as a function of frequency is shown in Fig. 4.

References

Dulk, G.A.: Annu. Rev. Astron. Astr. **23**, 169 (1985)

Gary, D.E.: In: Gary, D.E., Keller, C.U. (eds.) Solar and Space Weather Radiophysics, p. 74. Kluwer Academic, Dordrecht (2004)

McLean, D.J., Labrum, N.R. (eds.): Solar Radio Physics. Cambridge University Press, Cambridge (1985)

Melrose, D.B.: Solar Phys. **111**, 89 (1987)

Melrore, D.B., Dulk, G.A.: Astrophys. J. **259**, 844 (1982)

Sturrock, P.A.: Solar Flares, p. 357. Colarado University Associated Press, Boulder (1980)

Elemental and Charge State Composition in the Heliosphere

Eberhard Möbius

Abstract The heliosphere is filled with material from three samples of matter, solar material in the form of the solar wind and solar energetic particles, material from the neutral fraction of the surrounding interstellar medium, and galactic cosmic rays. In this lecture we will address the first two samples that constitute the main plasma component and evidence of their further acceleration in the heliosphere. We will review the observation techniques and some key observations, which provide information about fractionation, acceleration, and transport. For brevity we limit the discussion of the solar wind to some recurring topics that connect particularly well to the discussion of further acceleration and fractionation.

1 Introduction: Why Composition and Q-State?

Observation of elemental, isotopic and charge composition in heliospheric particle populations provides constraints on contributing sources and their composition. Such measurements, when taken over a wide energy range, reveal the characteristic signatures of heating, acceleration and transport of all ion species. For neutrals, ionization and filtration processes also leave their unmistakable mark. Often, these processes are highly selective in mass, charge and energy distribution. Thus, heliospheric composition measurements generate some of the most important and complex datasets, which provide tremendous opportunities for the analysis of physical processes and cosmologically important compositional patterns.

The heliosphere is filled with and exchanges material from three different samples of matter, solar material as solar wind and a variety of energetic particles, the neutral fraction of the surrounding interstellar medium, and galactic cosmic rays

---

E. Möbius
Space Science Center and Department of Physics, University of New Hampshire
e-mail: eberhard.moebius@unh.edu

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_10,

(GCR). On the grand scheme these samples represent different stages of the evolution of matter in the universe and our galaxy. While primordial matter only consisted of H and He, with a very small contribution of Li and B, right after the Big Bang, heavy elements were formed in consecutive generations of stars and injected into the interstellar medium by winds of aging massive stars and supernovae. It is thought that the fraction of heavy elements continuously increases with the age of the universe. With its formation 4.5 billion years ago, the solar system froze-in a cloud of interstellar matter at that time, with the Sun representing the lion share and most unbiased sample. The interstellar material surrounding the solar system is a contemporary sample with 4.5 billion years of additional evolution. Together with the reasonably well-understood Big Bang nucleosynthesis (Geiss and Gloeckler 2007) the first two samples provide three composition data points over the history of the universe. GCRs contribute valuable insight into elements and isotopes, which are relatively rare, and have been mostly formed in interaction between cosmic rays and interstellar gas, such as B and Be. Therefore, a detailed abundance pattern provides key information on the evolution of matter.

Figure 1 gives a simplified overview of sources and acceleration sites in the heliosphere. All three main sources contribute to the observed particle populations. The shading in Fig. 1 indicates how the relative importance of the solar wind (white) and the interstellar medium (grey) changes with distance from the Sun. Planets with their atmospheres, comets, asteroids, and even dust (not shown) serve as additional injectors. While these objects mostly belong to the solar system and thus inherited largely the composition of the protosolar nebula, they have undergone a variety of fractionation processes during the formation and evolution of the solar system. Only the Sun is thought to have retained mostly the original composition, since nuclear fusion is well contained in its core. A final fractionation occurs because mostly volatile elements are found in and thus released from

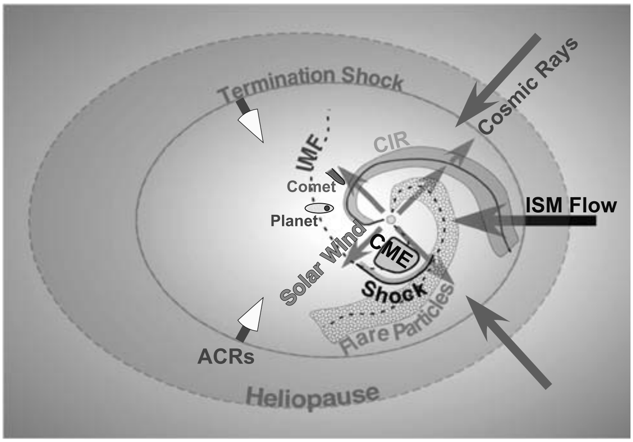


Fig. 1 Schematic view of the heliosphere with various particle populations, their sources, and related acceleration sites (from Möbius and Kallenbach 2005)

atmospheres. On the flip side, this last step in the chain, along with the environmental conditions, which mostly determine the ionic charge state upon release, allows identification of these local sources through element and charge composition. For example, the solar wind as the extension of the 1–2 MK corona contains charge states that correspond to this temperature (C^{5+}, O^{6+}, and Fe^{9-12+}), while comet and planet atmospheres release almost exclusively neutral and singly ionized components.

Also shown in Fig. 1 are the most prominent acceleration sites in the heliosphere. Through solar flares, the Sun features a second channel of particle release in addition to the solar wind, which as will be seen later exhibits starkly different characteristics. Wherever the solar wind abruptly changes its velocity, it is compressed, often until a shock forms, with the consequence of effective particle acceleration. Where fast solar wind overtakes slow wind a co-rotating interaction region (CIR) forms, a pattern that rotates exactly with the Sun. Likewise, coronal mass ejections (CMEs) with their often rather high speeds lead to a similar interaction, in most cases with a leading interplanetary (IP) shock. It is reasonable to assume and indeed observed that in these interaction regions particles out of all present sources are accelerated, posing a challenge and at the same time providing opportunities for diagnostics. In this limited overview we restrict ourselves to particle populations produced in the inner heliosphere, thus omitting anomalous cosmic rays (ACRs) and GCRs.

2 Key Composition Parameters

The nuclear charge number Z, which also equals the number of electrons, determines the element. To distinguish isotopes the mass M with mass number A must be obtained, where $M = A\,\mathrm{m_p}$. Often, it is sufficient to measure A to also identify the element, because the most important isotope of each element does not have a mass counterpart in neighboring elements.

The ionic charge state Q is determined by how many electrons an atom has lost. Q can range from 0 to Z and thus should be clearly distinguished from the latter. Generally, higher temperatures lead to the extraction of more electrons, because higher energy is needed to eject additional electrons from inner shells. For equilibrium conditions the relation between temperature and Q has been compiled for a number of elements (Mazzotta et al. 1998, and references therein). A combination of ionic charge Q and mass A along with their actual energy E usually determines the acceleration and transport of different species.

In summary, the combination of A and/or Z, Q, and E is needed to carry out a successful composition measurement. Often either A or Z is sufficient, and for many studies Q is not needed. Still, to obtain composition and energy, a combination of two or more measurement techniques is needed. In this paper we will present a few such combinations that have been implemented in space-born instruments, followed by observations of the solar wind and energetic particles.

3 Techniques for Solar Wind and Suprathermals

Detailed composition measurements in the solar wind started with the deployment of thin metal foil sails on the Moon during the Apollo missions and their return to Earth. After solar wind had been implanted over several hours or days on the Moon, the captured content was analyzed in the laboratory (Geiss et al. 1972), leading to the first average element and isotope abundances in the solar wind. However, this method does not allow tracking of temporal changes, long term monitoring, and has other limitations, which called for in-situ methods.

3.1 *Electrostatic Analyzers*

Since the beginning of the space age solar wind observations are being made using electrostatic analyzers in the form of Faraday Cups (Vasyliunas 1971) and hemi-spherical analyzers (Gosling et al. 1978) that utilize retarding potentials and electrostatic deflection, respectively, for energy selection. Figure 2 shows functional principles for the example of a schematic solar wind spectrum shown here as a function of deflection voltage (U_{Defl}) for a hemispherical analyzer. It demonstrates how such an analyzer provides species identification in combination with the defined flow speed of the solar wind, which translates into a common energy per mass E/A for all species. A hemispherical analyzer consists of two concentric hemispheres with radius R and separation ΔR, of which the inner one usually is biased with a negative deflection voltage. For ions that pass the analyzer on a central trajectory the centripetal force $A m_{\mathrm{p}} v_{\mathrm{o}}{}^2/R$ balances the electrostatic force $Q e U_{\mathrm{Defl}}/\Delta R$. For a given U_{Defl} ions of a certain energy per charge E/Q pass through the analyzer. In our example the speed of the particles is the solar wind speed, i.e. $v_{\mathrm{o}} = v_{\mathrm{sw}}$. Combining E/Q with the common E/A, the analyzer separates species with distinctly different A/Q, as shown in the lower left of Fig. 2 for H^+ and He^{2+}. The separation capability is determined by the width of the solar wind velocity distribution $\Delta v/v_{\mathrm{o}}$ and the analyzer resolution $\Delta(E/Q)/(E/Q)$. With $\Delta v = v_{\mathrm{th}}$ (solar wind thermal speed) and $\Delta(E/A)/(E/A) \approx 2\Delta v/v_{\mathrm{o}}$, the combined A/Q resolution of the analyzer becomes:

$$\frac{\Delta(A/Q)}{A/Q} = \sqrt{\frac{4v_{th}^2}{v_o^2} + \frac{\Delta(E/Q)^2}{(E/Q)^2}}. \tag{1}$$

For a typical analyzer resolution of 3–10% species identification depends heavily on the solar wind being cold. H^+ and He^{2+} typically can be resolved, but it becomes much more challenging, if not impossible, to separate O^{6+} and Fe^{10+} from He^{2+} and each other. Moreover, this method is restricted to the bulk solar wind.

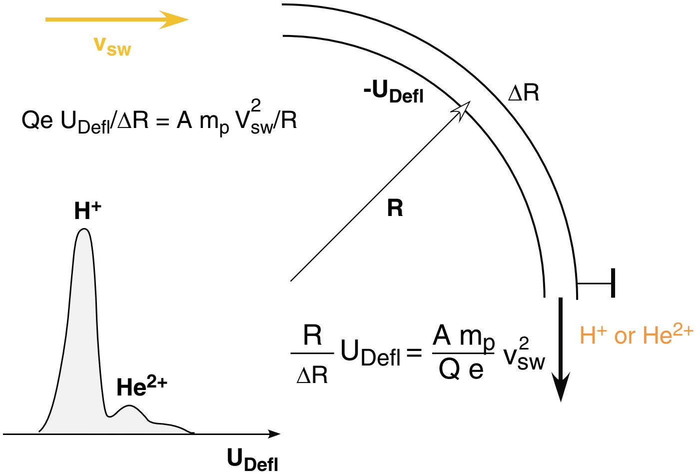


Fig. 2 Schematic view of solar wind observations with an electrostatic analyzer

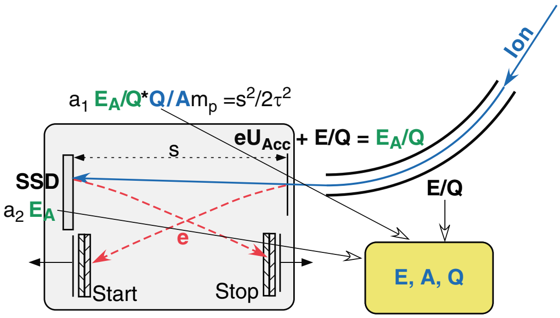


Fig. 3 Schematic cross-section and functional diagram of a time-of-flight mass spectrometer

3.2 Time-of-Flight Spectrometers

Progress towards mass resolution independent of the solar wind distribution was made with magnetic mass spectrometers and most recently with time-of-flight spectrometers. Figure 3 shows a schematic cross-section and the governing relations for a time-of-flight (TOF) mass spectrometer.

After passing an electrostatic analyzer, which selects the ions for E/Q, they pass a thin ($\approx$1–3 μg cm^{-2}) carbon foil, from where they move on to a solid-state detector (SSD) (or a micro-channel plate (MCP)) for a path length s. To achieve a reasonably good resolution after energy loss in the foil, the energy of incoming ions is boosted by post-acceleration with $U_{\mathrm{Acc}} = 15$–25 kV, resulting in

$$E_{\mathrm{A}}/Q = \mathrm{e}U_{\mathrm{Acc}} + E/Q. \tag{2}$$

The ions release secondary electrons from the C-foil and the surface of the SSD, which – by means of appropriate electrostatic potentials – are accelerated and steered to MCPs for detection as start and stop signals to record the TOF τ. The energy per mass of the incoming ion is then

$$\frac{a_1 E_A}{Am_p} = \frac{s^2}{2\tau^2}, \tag{3}$$

where $a_1 \leq 1$ is the fraction of the ion energy after passing the foil. Combining 2 and 3 yields A/Q. The SSD returns the residual ion energy $a_2 E_A$, where a_2 is the fraction of the energy that is actually detected. Both a_1 and a_2 vary with energy and species and are determined in calibration at an ion accelerator. Including the residual energy in the analysis yields E, A, and Q of the incoming ion separately. An overview of TOF spectrometers may be found in Wüest (1998) and Young (2002).

3.3 *High-Resolution Time-of-Flight Spectrometers*

Linear TOF spectrometers as described above are limited in their resolution by the energy and angle spread after energy and angular straggling due to the interaction with the C-foil. These limitations can be circumvented if the ions traverse a ballistic trajectory in a retarding electric field whose strength increases linearly with distance from the plane in which the entrance foil and the stop detector are located (Fig. 4). Because of the linearly increasing electric field $E(y) = k \cdot y$ the TOF τ is half the period of a harmonic oscillator for an ion with mass per charge A/Q.

$$\tau = \frac{\pi}{\omega} = \pi \sqrt{\frac{Am_p}{ek}}. \tag{4}$$

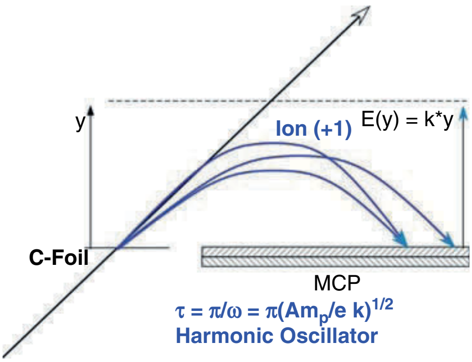


Fig. 4 Schematic cross-sectional view of a high-resolution TOF spectrometer, including ballistic trajectories of ions in the retarding electric field E(y). The time-of-flight τ equals half the period of a harmonic oscillator

Relation 4 is for singly charged ions because the majority of ions will either be neutral or singly charged after passage of the C-foil with energies that range from 1 to 30 keV e^{-1}, requiring a retarding potential of ≈20 kV for 45° trajectories. Neutral particles leave the TOF on a straight line, while singly charged ions reach the stop detector. Neither initial energy nor incoming angle matter for the resulting TOF, which solely depends on A/Q, or on A for $Q = 1$ (Wüest 1998, and references therein).

4 Key Composition Observations in the Solar Wind

The solar wind exhibits large variations in its abundances of heavy ions on time scales of hours or even down to minutes that can amount up to almost a factor of ten. Also the charge states of the ions vary along with the abundance variations. Usually, charge states are expressed in terms of their freeze-in temperature, i.e. the temperature for which the observed charge states occur in equilibrium state. Because the solar wind expands rapidly as it emerges from the corona its density falls and collisions become so rare that the original charge states are frozen in during the expansion of the solar wind (for a review see Bochsler 2007).

4.1 The FIP Effect

Systematic studies of solar wind abundances with Ulysses SWICS have shown that low freeze-in temperatures and high abundances of, for example Si and Fe compared with O, are associated with high-speed solar wind and vice versa (von Steiger et al. 2000; Fig. 5). In particular, the abundance ratios show much sharper boundaries between the solar wind types than speed. Therefore, the abundances are often taken as a better indicator for fast and slow solar wind.

When organizing the solar wind abundance ratios according to the first ionization potential (FIP), it turns out that generally the elements with values below the FIP of H (13.6 eV) are enhanced over those with FIP values at or above that of H, when compared with photospheric abundances. This is interpreted as a competition of magnetic lifting out of the photosphere and gravitational settling.

4.2 Isotope Fractionation

Measuring the isotopic ratios of refractory elements in the solar wind has not turned up a significant fractionation relative to the ratios in the closest proxies of the protosolar nebula, meteorites and Jupiter. A weak trend in the ratio with solar wind speed is found, consistent with the expected effects of Coulomb drag, but so far

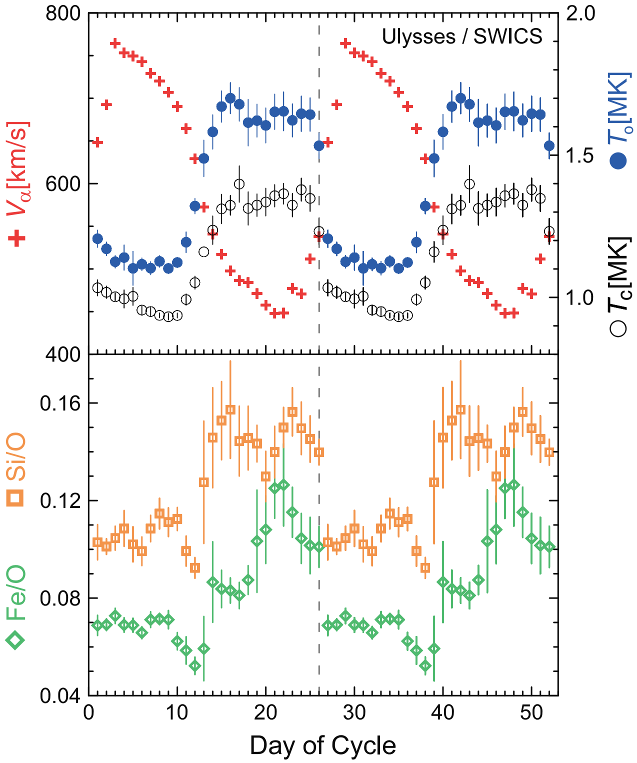


Fig. 5 Superposed epoch analysis of Ulysses SWICS solar wind composition data for fast to slow solar wind transitions (von Steiger et al. 2000). Overlaid are solar wind speed, C and O temperatures (*top*), Si/O and Fe/O (*bottom*)

these results are not statistically significant yet (Kallenbach et al. 1998, and review by Bochsler 2007). However, occasionally a substantial fractionation of ^{3}He is observed in CME ejecta (cf. review by Zurbuchen and Richardson 2006).

4.3 Implantation of Singly Charged Pickup Ions

The heliosphere is also populated with a number of neutral gas sources. Ionization by solar UV photons, charge exchange with solar wind ions and electron impact turns neutrals into ions. These newborn ions are picked up by the interplanetary magnetic and convectional electric fields, forced to gyrate about the magnetic field, and consequently swept radially outward with the solar wind. Through pitch-angle diffusion and adiabatic cooling of the velocity distribution in the radially expanding solar wind these pickup ions typically fill a sphere in velocity space with radius equal to the solar wind speed around the solar wind velocity (see review by Szegö et al. 2000).

The most extended source of such pickup ions is the neutral interstellar gas, which penetrates the entire heliosphere. He^+ ions were detected first (Möbius et al. 1985), followed by H^+ and heavy ions (see review by Gloeckler and Geiss 2001). While interstellar ions become more important at larger distances from the Sun due to their continuous implantation into the solar wind, the so-called inner source, thought to originate from interplanetary dust, is concentrated close to the Sun (for a review see Allegrini et al. 2005). In addition, to these extended sources, localized pickup ions have been found in the vicinity of comets or planetary atmospheres. Common distinctions of pickup ions from the solar wind are their broad velocity distributions and the fact that they are singly charged.

5 Techniques for Composition of Energetic Ions

To expand composition measurements into the suprathermal and medium energy range up to $\approx 1\,\mathrm{MeV\,nuc^{-1}}$ the same measurement techniques as presented in Sect. 2 for solar wind energies can be used. In fact, the task is simplified for somewhat higher energies because no post-acceleration with its high voltage challenges is needed. However, with increasing energy the TOF decreases for a given path length, which, with limited timing resolution, determines the upper energy limit for TOF sensors ($\approx 1\,\mathrm{MeV\,nuc^{-1}}$).

5.1 $\Delta E - E$ Telescopes

Above a few $100\,\mathrm{keV\,nuc^{-1}}$, overlapping with the TOF technique, energy loss in matter, which has been limiting TOF resolution through energy straggling in the C-foil, can in turn be used as a tool. The most important interaction of ions while penetrating matter consists of Coulomb collisions with electrons. For individual collisions at non-relativistic energies the momentum change Δp_e of the ion is:

$$\Delta p_e \approx \overline{F} \cdot \overline{\Delta t} = \frac{Z^* \cdot e^2}{4\pi \cdot \varepsilon_o \cdot b^2} \cdot \frac{2b}{v} = \frac{2Z^* \cdot e^2}{4\pi \cdot \varepsilon_o \cdot b \cdot v}. \quad (5)$$

F is the Coulomb force and Z^* is the effective charge of the penetrating ion, which is equal to Z for energies $> 1\,\mathrm{MeV\,nuc^{-1}}$. For illustration purposes the total interaction has been approximated by integration only along a path of $2b$ (b being the impact parameter, i.e. the closest distance between the collision partners). The corresponding energy loss ΔE_e follows as:

$$\Delta E_e = \frac{\Delta p^2}{2m_e} = \frac{2\,Z^{*2} e^4}{(4\pi\varepsilon_o)^2 b^2 v^2 m_e}. \quad (6)$$

The total energy loss ΔE is the sum over all collisions with different impact parameters b for the electron density and thickness ΔL of the detector. Carrying only the relevant proportionalities ΔE is:

$$\Delta E \sim Z^{*2}/v^2 \sim Z^{*2} A/E \tag{7}$$

From ΔE and E the nuclear charge Z and mass number A can be determined. Figure 9 shows an arrangement of a thin detector for ΔE determination and a thick detector to stop the ion and determine E. For high energies both can be SSDs of different thickness. For energies below a few MeV per nucleon the thin detector has to be a gas detector, which features a much lower matter density than solids. In a gas counter the penetrating ions ionize a certain number of atoms, and the resulting electrons are collected as a charge signal, which is proportional to the energy loss ΔE, hence proportional counter. For a description of these techniques see the book by Ahmed (2007).

The lower portion of Fig. 6 shows a ΔE–E representation of an SEP event taken with ACE SEPICA (Möbius et al. 1998). Mostly individual tracks are clearly

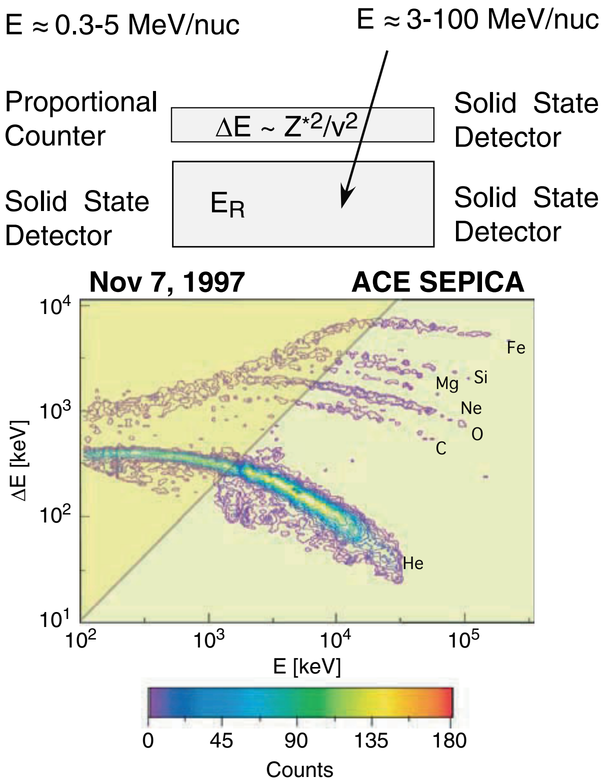


Fig. 6 *Top*: Schematic view of a ΔE–E particle telescope with combinations for different energy ranges. *Bottom*: ΔE–E diagram for an SEP event. Tracks for individual elements are clearly separated except for energies below $\approx$0.3 MeV nuc^{-1}

identifiable for each element. At the highest energies they fall off as $1/E$ (see Sect. 4.3). Below $\approx$1 MeV nuc^{-1} the tracks of heavier ions turn over and merge with other species at the low energy end (indicated by the yellow triangle), reflecting pickup of electrons in the detector material, thus reducing Z^* to $Z^* < Z$.

5.2 *Ionic Charge State Measurement at High Energies*

To also determine the ionic charge state Q requires the addition of a third measurement. At moderately high energies ($\leq$3 MeV Q^{-1}) this can be achieved with electrostatic deflection, but not with spherical analyzers. Firstly, for MeV energies this would lead to a combination of rather large radii and extremely high deflection voltages. Secondly, the necessary stepping through energy makes the effective geometric factor too low for the steep energy spectra. Therefore, mechanical collimator with subsequent deflection and position measurement is used (ACE SEPICA example in Fig. 7, Möbius et al. 1998).

Above a few MeV/Q electrostatic deflection becomes too small in any reasonably sized sensor. To extend charge state measurements the assistance of the Earth's magnetic field has been enlisted. Sako (2010) describes how penetrating cosmic ray fluxes at low Earth orbit satellites and on the ground depend on magnetic rigidity R and the location within the Earth's magnetic field:

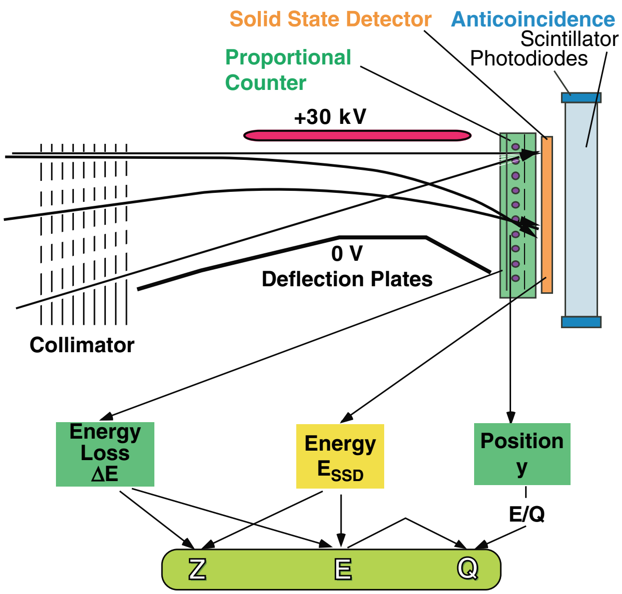


Fig. 7 Schematic cross-sectional view of the ACE SEPICA sensor (Möbius et al. 1998) with its functional principles

$$R = \frac{pc}{Qe} = \frac{c\sqrt{2Am_pE}}{Qe}. \tag{8}$$

Close to the poles the cut-off rigidity depends distinctly on the magnetic latitude. Thus Q can be determined if A and E are measured with a sensor as described above. The cut-off method is being employed successfully on SAMPEX for 0.5 to several 100 MeV nuc^{-1} (Baker et al. 1993, and references therein).

6 Key Observations with Energetic Particles

The general behavior of solar energetic particles (SEPs) is discussed by Sako (2009) and shall not be repeated here. We will rather concentrate on compositional aspects in heliospheric energetic particles. It should suffice to note that SEPs are usually subdivided into two groups. Large gradual events typically last for days, are accompanied by CMEs, and averaged over many events show abundances and ionic charge states of the corona. Mostly small, impulsive events, only last for a few hours, are associated with compact flares, and show substantial enhancements in ions heavier than O and in ^{3}He over ^{4}He, hence ^{3}He-rich events. For an overview the reader is referred to the paper by Reames (1999) and references therein.

6.1 Elemental and Isotopic Composition

6.1.1 Gradual SEPs

In several extended surveys it has been found that, averaged over a number of events, gradual SEPs show abundances and charge states that reflect coronal values. However, they also exhibit substantial event-to-event variations, which range from significant depletions to enhancements, for example, in Fe (Fe-poor and Fe-rich) (Reames 1999). Usually, the observed enhancements/depletions exhibit a trend in Q/A indicating Q/A dependent selection processes that may be attributed to acceleration, escape, and/or transport of the ions in these events.

Because on average gradual SEPs reflect the composition of the corona they also bear the imprint of the FIP effect as found in coronal and solar wind composition. While also the strength of the FIP effect varies from event to event, it is typically closer to the more pronounced value observed in the slow solar wind (Reames 1999). This may indicate the source material for the acceleration originates from the closed field region where also the slow wind stems from. However, it should be noted that the observed variations in the FIP effect make it impossible at the moment to pin down exactly the source of the gradual SEPs.

6.1.2 Impulsive SEPs

The event-to-event variations in gradual SEP composition are usually only substantial for ions with starkly different A/Q values and they vary in both directions. Conversely, abundances in impulsive SEPs can also be substantially enhanced for neighboring elements, and heavy ions are always enhanced over light ions. For example, Ne is substantially enhanced over O, up to a factor of 10 in some events. Again, the enhancement of heavy ions can be organized in terms of A/Q, assuming ionic charge states that resemble a 3–5 million K corona (Reames 1999). As can be seen in Fig. 8, heavy ion abundances increase approximately exponentially in A/Q, with ultra-heavy elements ($Z = 60-80$) exceeding a factor of 1,000 (Reames 2000; Mason et al. 2004). However, the substantial enhancement of ^{3}He in these events does not fit into this trend, because $A/Q = 1.5$ goes into the opposite direction. Compared to the natural $^3\text{He}/^4\text{He}$ ratio of 5.10^{-4} enhancements up to 10^4 have been observed with $^3\text{He}/^4\text{He}$ ratios exceeding 1.

Proposed models invoke selective heating and acceleration through resonant wave-particle interaction. Suggested are ion-acoustic turbulence, electrostatic ion cyclotron waves, electromagnetic ion cyclotron waves, and cascading Alfvénic turbulence (For an overview see the review by Reames 1999). While any of these models explain some of the observations, mostly qualitatively, none of them provides a complete quantitative explanation. The recently observed strong enhancement of ultra-heavy ions has dealt even the most successful models a significant blow. It should be emphasized though that all of them so far rely on differences in A/Q for any enhancement. Therefore, ionic charge states for all species are very important to assess any explanatory model.

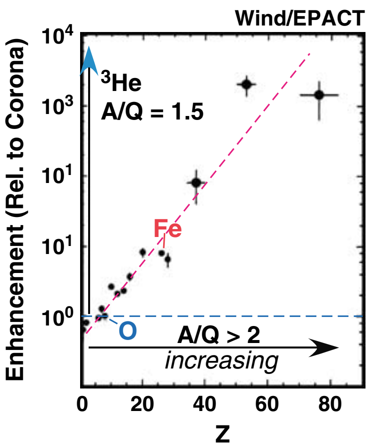


Fig. 8 Abundances in impulsive events as compared with the corona, organized in A/Q (adapted from Reames 2000)

6.2 *Ionic Charge States*

First charge state observations, averaged over all impulsive events of an entire year returned a puzzling result. The mean charge of Fe was found to be typically $Q \approx 20$, and all species up to Si appeared to be fully stripped, i.e $Q = Z$ (Klecker et al. 2007). With $A/Q = 2$ under fully stripped conditions there is no leverage for A/Q dependent fractionation. With ACE SEPICA resolution and collection power have been improved substantially so that individual impulsive event observations, even separated by energy, have become available.

6.2.1 Energy Dependence of Q

Early ionic charge state observations for both types of SEPs showed no energy dependence and supported the conclusion that charge states were frozen-in based on the temperature of the source plasma. However, observations in an extended energy range with SAMPEX showed a charge state increase above a few MeV per nucleon. The first major SEP event observed with ACE even showed an increase in the Fe charge state of $\Delta Q > 4$ at 0.2–0.5 MeV nuc^{-1}, which was also seen simultaneously with SAMPEX, followed by several similar cases. Such a strong energy dependence of the charge state requires stripping of the energetic ions in ambient material as an explanation (review by Klecker et al. 2007).

Because the most important ionization process in plasmas involves electron impact each element is effectively ionized up to the point when the necessary ionization potential exceeds the electron thermal energy. When energetic ions pass through the plasma at a speed that exceeds the thermal electron speed the ion energy starts to play a role in the formation of the charge state. At plasma temperatures of 10^6 K (1–2.10^6 K are typical for the corona) ion speeds at $\approx$0.25 MeV nuc^{-1} are comparable to the electron thermal speed, i.e. the energy where the charge state increase indeed starts some of the events. However, to leave a noticeable effect in the charge state of the energetic ions enough collisions must occur between the source and the observer. The necessary condition is either expressed as a column density $n\,l$ (with plasma density n and path length of traversal l) or a product $n\tau$ with total travel and acceleration time τ. For typical values of the ionizing cross sections of $10^{-16}\,\mathrm{cm}^{-2}$ the two conditions read $n\,l \geq 10^{19}\,\mathrm{cm}^{-2}$ and $n\,\tau \geq 10^{10}\,\mathrm{cm}^{-3}\mathrm{s}$.

These new findings suggested a strong energy dependence also for impulsive SEP events, whose source is in the solar atmosphere. This would be consistent with partial ionization at low energies where the observed strong ion fractionation supposedly starts and full ionization at high energies. Indeed such behavior was found for impulsive events in observations with ACE SEPICA and at even lower energies with SOHO STOF as shown in Fig. 9 together with modeling that assumes energetic ions achieving equilibrium charge states for each energy after emerging from plasmas with two different temperatures (Klecker et al. 2007). To achieve quantitative agreement finite residence times in the plasma and adiabatic deceleration of the ions from the source to 1 AU must be included. The low energy observations

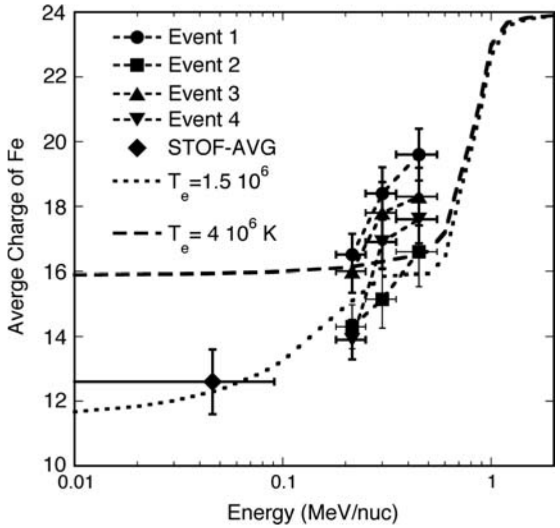


Fig. 9 Fe charge states in four impulsive SEP events and equilibrium models with stripping (Klecker et al. 2007)

constrain the source temperature to $\approx$1.2–1.8. 10^6 K. The observed energy dependence requires acceleration of the ions in the lower corona – for all impulsive and certain gradual SEP events that exhibit a similar behavior.

6.3 Sources and Fractionation in Interplanetary Space

While SEPs in impulsive events have been traced to acceleration at the flare site, for CME related events shock acceleration of ambient material is invoked. A subset of these events shows ongoing particle acceleration even at 1 AU. As it turns out CME driven shocks are not the only ones accelerating particles in interplanetary space. The succession of slow and fast solar wind leads to compression regions, which rotate with the Sun, hence co-rotating interaction regions (CIR). Compression regions and the shocks that usually form at $\geq$1 AU are known as formidable particle accelerators from localized energetic particle flux increases. For overviews see Reames (1999) and Möbius and Kallenbach (2005).

Given the location of this acceleration, the solar wind has been implicated as a source for the resulting energetic particles. However, the observation of prominent contributions of ^{3}He and heavy ions with the abundance pattern of impulsive SEP events at interplanetary shocks and of He^+ in CIRs and at shocks support the conclusion that source populations with substantially wider velocity distributions than the relatively cold solar wind are injected and accelerated much more effectively in interplanetary space. The abundance of ^{3}He and heavy ions at shocks correlates well with the abundance pattern measured before the shock arrival, while there is no correlation with solar wind abundance. He^+ that stems from interstellar pick up ions and is present as a suprathermal distribution between $v = 0$ and $2v_{sw}$ shows up in He^+/He^{2+} ratios that exceed the abundance of pickup ions relative to the solar wind by factors of 100–200. Both findings clearly support that suprathermal particle populations are strongly favored in the injection process into acceleration as argued by Möbius and Kallenbach (2005) in their review of recent observations.

7 Summary and Conclusions

Instrumentation that provides elemental, isotopic, and ionic charge state composition has become available with impressive resolution and collection power for an extended energy range, which covers three samples of cosmic matter in the heliosphere: the Sun, from solar wind and SEPs, the local interstellar cloud, from pickup ions and anomalous cosmic rays, and finally galactic cosmic rays. When it comes to the question, how well these observations reveal the composition of these sources we discover a number of processes that alter the original abundance ratios to what is observed. In turn, composition measurements have helped tremendously to understand these fractionation processes. Each step in their understanding and more complete samples will better constrain the source composition.

Starting at the original sources, abundances may be biased in the transition from neutral atoms to ions, through differences in the first ionization potential (FIP) between species. In addition to gravitation and heat, charged particles can also be affected by electric and magnetic fields, which may lead to preferential escape from the source regions. For subsequent fractionation steps the mass per charge ratio A/Q appears to be the ruling parameter, which is determined once element, isotope, and charge state are known. However, quite often the ionic charge state can change substantially in the interaction with intervening matter. As a result, the observed charge states in these populations vary with energy, thus carrying the imprint of the entire interaction history. As a consequence, charge states must be measured over the entire energy range of interest, along with elemental composition, to pin the observations unquestionably to fractionation models.

In summary, composition measurements have been and are still invaluable for the determination of source distributions and all the following physical processes that alter abundances in the course of injection, acceleration, escape, and transport.

Acknowledgements Work on this manuscript was partially supported under NASA Grants NAG 5–12929 and NNG06GD55G.

References

Ahmed, S.N.: Physics and engineering of radiation detection. Academic, Boston (2007)

Allegrini, F., Schwadron, N.A., McComas, D.J., Gloeckler, G., Geiss, J.: Stability of the inner source pickup ions over the solar cycle. J. Geophys. Res. **110**, (2005). doi:10.1029/2004JA010847

Baker, D.N., et al.: An overview of the solar, anomalous, and magnetospheric particle explorer (SAMPEX) mission. IEEE Trans. Geosci. Remote Sens. 31:531–541 (1993)

Bochsler, P.: Minor ions in the solar wind. Astron. Astrophys. Rev. **14**, 1–40 (2007)

Geiss, J., Buehler, F., Cerutti, H., Eberhardt, P.: Solar-wind composition experiment. NASA SP-289 NASA, Washington, D.C 25–122 (1972)

Geiss, J., Gloeckler, G.: Linking primordial to solar and galactic composition. Space Sci. Rev. **130**, 5–26 (2007)

Gloeckler, G., Geiss, J.: Heliospheric and interstellar phenomena deduced from pickup ion observations. Space Sci. Rev. **97**, 169–181 (2001)

Gosling, J.T., Asbridge, J.R., Bame, S.J., Feldman, W.C.: Rev. Sci. Instrum. **49**, 1260 (1978)

Kallenbach, R., et al.: Fractionation of Si, Ne, and Mg Isotopes in the solar wind as measured by SOHO/CELIAS/MTOF. Space Sci. Rev. **85**, 357–370 (1998)

Klecker, B., Möbius, E., Popecki, M.: Ionic charge states of solar energetic particles: A clue to the source. Space Sci. Rev. **130**, 273–282 (2007)

Mason, G.M., et al.: Abundances of heavy and ultraheavy ions in ^{3}He-rich solar flares. Astrophys. J. **606**, 555–564 (2004)

Mazzotta, P., Mazzitelli, G., Colafrancesco, S., Vittorio, N.: Astron. Astrophys. Suppl. **133**, 403 (1998)

Möbius, E., et al.: Direct observation of He^{+} pick-up ions of interstellar origin in the solar wind. Nature **318**, 426–429 (1985)

Möbius, E., et al.: Space Sci. Rev. **86**, 447 (1998)

Möbius, E., Kallenbach, R.: Acceleration in the heliosphere. In: Geiss, J., Hultqvist, B. (eds.) The heliosphere and beyond/10 years of ISSI, ISSI-SR003, 165, (2005)

Reames, D.V.: Particle acceleration at the Sun and in the heliosphere. Space Sci. Rev. **90**, 413–491 (1999)

Reames, D.V.: Abundances of trans-iron elements in solar energetic particle events. Astrophys. J. **540**, L111–L114 (2000)

Sako, T.: Solar energetic particles: Acceleration and observations. Astrophys. Space Sci. doi:10.1007/978-3-642-11341-3

Szegö, K., et al.: Physics of mass loaded plasmas. Space Sci. Rev. **94**, 429 (2000)

Vasyliunas, V.M.: Deep space plasma measurements. In: Lovbergs, R.H. (ed.) Methods of experimental physics, vol. 9, pp. 49–88. (1971).

von Steiger, R., et al.: Composition of quasi-stationary solar wind flows from ulysses/solar wind ion composition spectrometer. J. Geophys. Res. **105**, 27217–27238 (2000)

Wüest, M.: Time-of-flight ion composition measurement technique for space plasmas. In: Pfaff, R.F., Borovsky, J.E., Young, D.T. (eds.) Measurement techniques in space plasmas – particles, Geophys. monograph vol. 102, p. 141. AGU (1998).

Young, D.T.: Mass spectrometry for planetary science. In: Mendillo, M., Nagy, A., Waite, J.H. (eds.) Atmospheres in the solar system: comparative Aeronomy, Geophys. monograph vol. 130, p. 353. AGU (2002).

Zurbuchen, T.H., Richardson, I.G.: In-situ solar wind and magnetic field signatures of interplanetary coronal mass ejections. Space Sci. Rev. **123**, 23–41 (2006)

Sako, T.: Solar energetic particles: Acceleration and observations, this volume.

Planetary Atmospheres

H. Chandra

Abstract The planets in the solar system are described in terms of the physical parameters like radius, area, mass, gravity, distance from the Sun. Orbital and rotational periods, composition, effective temperature and magnetic parameters of different planets are also described. The nomenclatures of the division of planetary atmospheres in terms of temperature structure, composition and ionization are also described. The barometric equation, scale height and the variation of pressure and density in the atmosphere, escape velocity, escape flux of atmospheric gases, escape temperature are also discussed. The basic theory of the turbulent mixing and diffusion processes, atmospheric dynamics and atmospheric circulation are also covered. A brief description of the planetary atmospheres other than Earth is also presented.

1 The Planets

There are eight planets in the solar system (Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus and Neptune; Mercury being the nearest and Neptune farthest from the Sun) that revolve around the Sun in their specific orbits, which lie more or less in the Sun's equatorial plane. Planetary data are listed in Tables 1 and 2 (Degaonkar 1975).

The planets in the solar system can be divided into two categories.

(1) The inner or the terrestrial planets comprising of Mercury, Venus, Earth and Mars, which are small in size, warm, have higher densities of the order of $5\,\mathrm{gcm}^{-3}$ or more and sizes comparable to that of Earth. The terrestrial planets also rotate slower than outer planets.
(2) The outer planets or the Jovian planets (Jupiter, Saturn, Uranus and Neptune), which are much larger in size, colder and have low densities $\approx 1.5\,\mathrm{g\,cm}^{-3}$ (Jupiter like planets are hence called Jovian planets) and rotate much faster.

H. Chandra
Physical Research Laboratory, Ahmedabad 380 009, India
e-mail: hchandra@prl.res.in

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_11,

Table 1 Planetary data

Planet	Mean radius (km)	Mean density (gcm^{-3})	Average distance from the Sun (AU)[a]	Length of year (days)	Rotation period-(days)	Inclination (degree)
Mercury	2,439	5.42	0.39	88	58.7	<28
Venus	6,050	5.25	0.72	225	−243	<3
Earth	6,371	5.51	1.00	365	1.00	23.5
Mars	3,390	3.96	1.52	687	1.03	25
Jupiter	69,500	1.35	5.2	4,330	0.41	3.1
Saturn	58,100	0.69	9.5	10,800	0.43	26.7
Uranus	24,500	1.44	20	30,700	−0.89	98.0
Neptune	24,600	1.65	30	60,200	0.53	28.8

[a] $1\ \text{AU} = 1.496 \times 10^{13}$ cm

Table 2 Other planetary parameters

Planet	Surface area earth=1	Mass earth=1	Gravity earth=1	Escape velocity ($km\,s^{-1}$)	Atmosphere
Mercury	0.15	0.05	0.37	4.3	Trace?
Venus	0.9	0.81	0.89	10.4	CO_2 (96%), N_2 (3.5%), SO_2 (130 ppm)
Earth	1.0	1.0	1.0	11.2	N_2 (78%), O_2 (21%), Ar (0.9%)
Mars	0.3	0.11	0.39	5.1	CO_2 (95%), N_2 (2.7%), Ar (1.6%)
Jupiter	120	318	2.65	60.0	H_2 (86%), He (14%), CH_4 (0.2%)
Saturn	85	95	1.65	36.0	H_2 (97%), He (3%), CH_4 (0.2%)
Uranus	14	14	1.0	22.0	H_2 (83%), He (15%), CH_4 (2%)
Neptune	12	17	1.5	22.0	H_2 (79%), He (18%), CH_4 (3%)

Temperature, density and composition are the basic atmospheric parameters. Temperature is the most significant parameter of the atmosphere as it reflects the processes of energy absorption in the atmosphere and at the ground. A part of the solar radiation reaching the top of the atmosphere is reflected back by clouds and atmosphere. This fraction of the solar radiation reflected back to space is known as the albedo of the planet. Out of the remaining radiation most of it, mainly in the visible, penetrates down to ground and heats the surface to a certain temperature T_e known as the effective temperature of the planet. For fast rotating planets (Earth, Mars, Jupiter)

$$\sigma T_e^4 = (\text{S.C.}/4)(1 - A),$$

where S.C. is the solar constant and A the albedo of the planet.

The solar flux expected at the orbit of planet outside its atmosphere, its albedo and effective computed temperature T_{eff} are listed in Table 3. The actual temperature would depend on the presence or absence of the atmosphere, sunlit or dark condition etc. For Earth the actual temperature 288 K is warmer than the effective temperature.

In our planetary system there are bodies which have little or no atmosphere and magnetic field (Moon, Mercury), bodies which have substantial atmospheres but very little or no magnetic field (Venus and Mars) and bodies having both atmosphere

Table 3 Effective temperature of planets

Planet	Solar flux (10^{16}erg cm^{-2} s^{-1})	Albedo	T_{eff} (K)
Mercury	9.2	0.06	442
Venus	2.6	0.71	244
Earth	1.4	0.38	253
Mars	0.6	0.17	216
Jupiter	0.05	0.73	87
Saturn	0.01	0.76	63
Uranus	0.004	0.93	33
Neptune	0.001	0.84	32

Table 4 Magnetic field parameters of planets

Planet	Magnetic dipole moment (Me)	Core radius (km)	Magnetic dipole tilt (degrees)	Magnetic dipole offset (in planetary radii)
Mercury	3.1×10^{-4}	~1,800	2.3	0.2
Venus	$< 5 \times 10^{-5}$	~3,000	–	–
Earth	1	3,485	11.5	0.07
Mars	3×10^{-4}	~1,700	(15–20)	–
Jupiter	1.8×10^{4}	~52,000	11	0.1
Saturn	0.5×10^{3}	~28,000	1.5 ± 0.5	< 0.05
Uranus		–	58.6	0.3
Neptune	–	–	46.8	0.55

and intrinsic magnetic field (Earth, Jupiter). Magnetic field parameters of different planets are listed in Table 4.

There seems to be a distinct difference between the atmospheres of terrestrial planets and Jovian planets. It is agreed that the atmospheres of the terrestrial planets are of secondary origin, having lost the primordial (original) constituents long time ago. This does not appear to be the case of Jovian planets, which have reducing atmospheres with a large amount of hydrogen, suggesting that they represent primordial atmospheres whose origin is similar to that of the solar system. The Jovian atmospheres are few hundreds kilometers thick and resemble the Sun in their overall composition (H_2 84%, He 16%).

The secondary atmospheres that we see on the terrestrial planets are most likely the result of outgassing of volatile materials from the planetary interior and radioactive decay products like helium and argon. The only exception is that of oxygen on Earth that seems to be related to the presence of plant life. Mercury being closest to the Sun and so being hot (~500 K), practically all its atmosphere has evaporated away, though an upper limit of 10^{-2} mb pressure due to CO_2 has been inferred from ground-based spectroscopic observations. A high-density means Mercury has a large solid core with 50% iron and a crust. So there could be argon and helium in its small tenuous atmosphere due to degassing of its internal radioactive material plus some amount of argon, helium and neon accreted from the solar wind.

Table 5 Composition of dry air by volume at the earth's surface

N_2	O_2	Ar	CO_2	Ne	He	Kr
78.09%	20.95	0.93	0.03	0.0018	0.00053	0.0001

Such a tenuous atmosphere may give rise to the surface pressure of 10^{-10}–10^{-12} mb compared to Earth's surface pressure of 10^3 mb.

The atmospheres on Venus, Earth and Mars also appear to be different at the present time although main outgassing constituents from the interior of planets have been water vapor, CO_2 and small amount of nitrogen. The surface temperatures of Venus, Earth and Mars are 750, 300 and 250 K and their atmospheric pressures are in the ratio of 100: 1: 0.01, respectively. Venus and Mars have CO_2 as the principal constituent. Earth has N_2 and O_2 as major constituents (Table 5); water is present in all the three forms (liquid, solid and gas) and has accelerated the removal of CO_2 from the atmosphere. Venus being very hot, no water can be expected on its surface. Whatever little water was there in its atmosphere, it must have been photodissociated by sunlight into hydrogen and oxygen: hydrogen would have escaped into space. Mars on the other hand being a cold planet, its volcanic steam must have got frozen on its surface allowing CO_2 to accumulate in its atmosphere.

Water vapor is the most variable part in Earth's atmosphere and may vary from negligible amounts to $\sim$5% by volume at any particular time. The ozone is merely a trace at the surface but peaks between 20 and 30 km with a peak value of about 0.016 cm at standard temperature and pressure (STP) and the total equivalent thickness varying between 0.15 and 0.40 cm at STP from the equator to the poles.

2 Atmospheric Divisions in Terms of Temperature Structure

The atmosphere of a planet can be divided into different regions**.** The terminology generally accepted is based on the vertical distribution of temperature and composition. In terms of the temperature structure, the atmosphere is divided into following regions (see Fig. 1 for Earth).

2.1 Troposphere

The troposphere is the lowermost part of the atmosphere where the primary heat source is the planetary surface. The heat is convected by turbulent motions leading to a convective or adiabatic distribution. The adiabatic lapse rate is given by $-g/c_p$, where g is the acceleration due to gravity and c_p the specific heat at constant pressure. The lapse rate depends on planet's acceleration due to gravity and composition. For Earth the theoretical estimate is $10\,\mathrm{K\,km^{-1}}$, while the observed value is $6.5\,\mathrm{K\,km^{-1}}$ due to the presence of water vapor and large-scale circulation. The

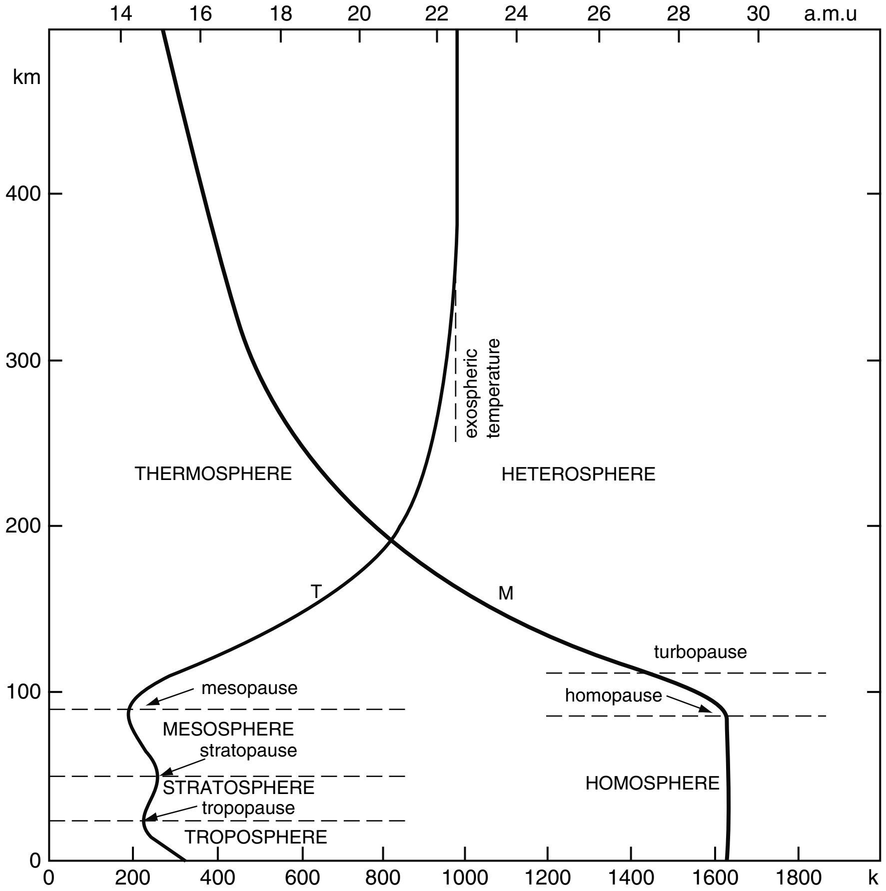


Fig. 1 Atmospheric divisions in terms of temperature (T in K) and mean molecular mass (M in atomic mass unit; after Giraud and Petit 1978)

troposphere terminates at the tropopause where the decrease in temperature ceases. Earth's tropopause varies from 18 km at the equator to 8 km at the poles.

2.2 Stratosphere

Above the tropopause the radiative processes govern the temperature distribution. In the terrestrial stratosphere, temperature increases with height mainly by the absorption of solar UV radiation by ozone reaching a maximum at the stratopause (50 ± 5 km).

2.3 Mesosphere

Above the stratopause, temperature decreases reaching a minimum at the mesopause. The mesopause for Earth is located at 85 $\pm$ 5 km. The decrease in temperature is due to the presence of CO_2 and H_2O, which provides a heat sink by radiating in the infrared. The mesosphere is dynamically unstable as convection is still prevalent.

Without a stratospheric heat source as in the case of Earth, a planetary atmosphere may not possess a stratopause, and the region above the tropopause may be called either the stratosphere or the mesosphere. Figure 2 shows the model temperature distributions for Earth, Mars and Venus while the model temperature distribution for the Jovian atmosphere is shown in Fig. 3.

2.4 Thermosphere

Above the mesopause, solar EUV radiation is absorbed and part of it is used in heating; therefore, in the thermosphere temperature increases with altitude. In the lower thermosphere convection is principal process of heat transport while in the upper thermosphere heat is transported by conduction leading to an isothermal region where temperature is constant above about 500 km (thermopause). Heating is due to the absorption of the solar EUV flux by atmospheric constituents. Still the heat transport due to conduction and convection and loss by the IR radiation by constituents govern the energy budget. Collisions between charged particles and neutrals, Joule heating, atmospheric waves (tides, internal gravity waves), hydromagnetic waves and the solar wind are other sources of heating. Atmospheric temperature shows diurnal (minimum at 06 h, maximum at 17 h local time) and solar cycle variations.

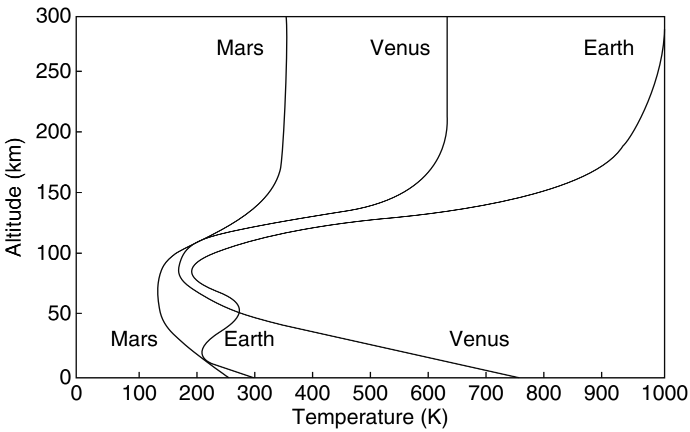


Fig. 2 Model temperature distributions for Venus, Earth and Mars (after Bauer 1973)

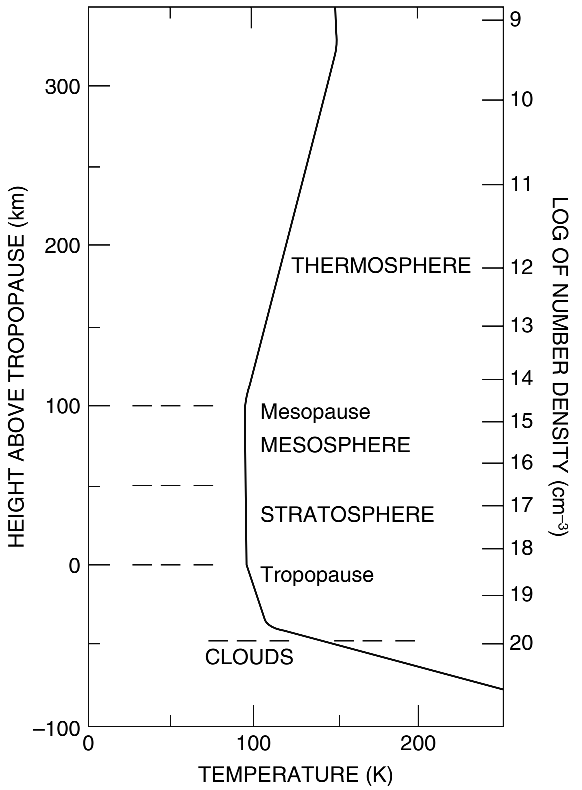


Fig. 3 Model temperature distribution as function of altitude and number density for the Jovian atmosphere (after Hunten 1969)

Figure 4 shows the vertical temperature profiles at the time of diurnal minimum and maximum for the terrestrial atmosphere.

2.5 Exosphere

In the region above 500 km (for Earth), the mean free path becomes large and collisions become negligible so that the light atmospheric constituents whose velocity exceeds the gravitational escape velocity can escape the atmosphere. This region is also called the exosphere. The exobase is also called the baropause since the atmosphere below this is also referred to as the barosphere, the region where barometric law holds. In the exosphere velocity distribution becomes non-Maxwellian due to the escape of high velocity particles. For Mars and Venus the exopause is around 200 km.

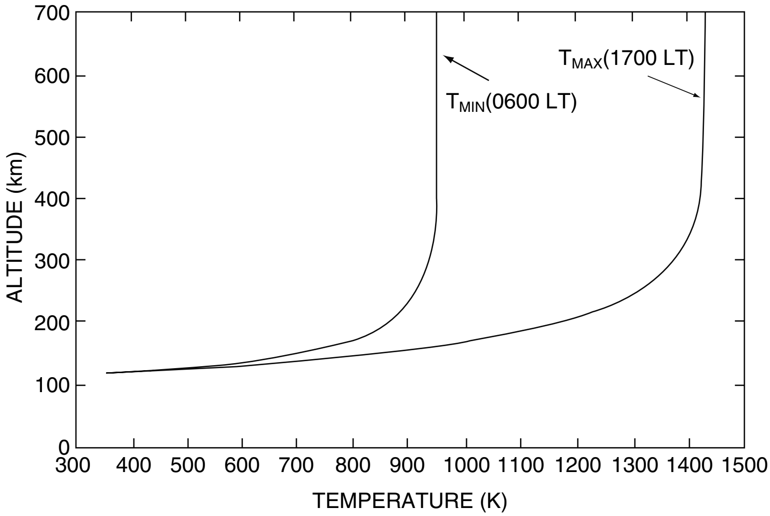


Fig. 4 Vertical temperature profiles in the thermosphere at the time of diurnal maximum and minimum (after Bauer 1973)

3 Other Divisions

3.1 Ionosphere

The ionosphere is a region of the upper atmosphere where charged particles (electrons and ions) are present. The solar extreme ultraviolet and X-rays and corpuscular radiation like cosmic rays and energetic particles (solar cosmic ray protons, the solar wind) are responsible for ionization. Meteors also produce some ionization in the atmosphere.

The ionosphere is situated above 50 km and above in the Earth's atmosphere. The upper boundary of the ionosphere cannot be specified exactly as it merges into the plasmasphere where dominant species is the hydrogen ion (proton and hence also referred to as the protonosphere). The plasmasphere terminates at the plasmapause situated between 4 and 6 Earth radii depending upon the geomagnetic quiet or disturbed conditions.

3.2 Magnetosphere

The magnetosphere is defined as the boundary within which the effect of the planet's magnetic field is felt. It is different in the dayside and nightside because of the interaction of the solar wind with the geomagnetic field. For Earth, the magnetopause boundary is around 10 Earth radii in the dayside.

The atmosphere can also be divided in terms of its composition. Below about 100 km the air is well mixed by turbulence and the composition of major constituents does not vary with altitude (homosphere). Above this (turbopause) a diffusive separation of constituents takes place (heterosphere).

4 Terrestrial Gravity

For a rotating Earth the acceleration due to gravity $g^{/}$ is given by

$$g^{/} = GM/R^2. \quad [R \text{ is distance from center of earth}] \tag{1}$$

In terms of local altitude z, Earth's radius R_o and sea level value $g_o^{/}$

$$g_o{}^{/} = GM/(R_o + z)^2 = GM/[(R_o)^2(1 + z/R_o)^2] \tag{2}$$

or

$$g^{/} = g_o{}^{/}/(1 + z/R_o)^2.$$

A reference frame spinning with Earth is not an inertial system of reference. In such a system there is in addition a centrifugal force directed perpendicular to the Earth's axis of rotation. The resulting apparent gravitational acceleration or g is no longer directed towards the center of Earth (except at the equator and poles)

$$g = g^{/} - \Omega^2(R_o + z)\cos^2\phi. \quad [\phi \text{ is geographic latitude}] \tag{3}$$

At 6.6 R_o from Earth's center, the centrifugal term equates the gravitational acceleration, at the equator. For smaller distances we can approximate

$$g = g_o(1 - 2z/R_o) \tag{4}$$

5 Hydrostatic (Barometric Equation)

The decrease of pressure or density with altitude is described by the hydrostatic distribution. The downward force due to gravity is balanced by the pressure difference (hydrostatic equation). For a cylinder of small height dh and unit cross section the pressure difference across the height dh is

$$dP = -nmg \cdot dh \tag{5}$$

Using the perfect gas law

$$P = nkT \tag{6}$$

$$(1/P)\, dP/dh \equiv -mg/kT = -1/H, \tag{7}$$

where $H = kT/mg$ is defined as a scale height. Integrating equation (7)

$$P = P_o \exp(-h/H), \tag{8}$$

where P_0 is the pressure at height $h = 0$. For Earth's atmosphere $H = 8$ km (in the homosphere) and about 70–80 km at 500 km. The scale height is also defined as the height at which density or pressure falls to 1/e.

If the whole atmosphere above the height h_o is compressed to a uniform pressure p_o, the vertical extent would be the scale height H_o.

Equation (8) can also be written in the form

$$P = P_o \exp^{-z}. \quad [z \text{ is the reduced height} = (h - h_o)/H] \tag{9}$$

Since $P/P_o = n/n_o = \rho/\rho_o$, similar relations hold for density and the number of particles per unit volume.

$H =$ constant is valid only for a limited altitude range. In general

$$P(z)/P(z_o) = n(z)\, T(z)/n(z_o)T(z_o) = \exp - \int dz/H(z). \tag{10}$$

In a mixture of gases P is the sum of partial pressures hence,

$$P_i(z)/P_i(z_o) = n_i(z)\, T(z)/n_i(z_o)\, T(z_o) = \exp - \int dz/H_i(z). \tag{11}$$

The model pressure distributions for the atmospheres of Mars, Earth and Venus are shown in Fig. 5.

6 Escape of Atmospheric Gases

Particle velocity in thermal equilibrium follows a Maxwellian distribution. If $f(v)$ is the probability of particles with velocity between v and $v + dv$, one can find the number of particles with velocity between v and $v + dv$, which in polar coordinates is

$$Nf(v)\, dv = N[\exp(-(v/v_m)^2)/(\pi v_m^2)^{3/2}]v^2 dv \sin\theta\, d\theta\, d\phi, \tag{12}$$

where V_m is the most probable velocity $= (2kT/m)^{1/2}$.

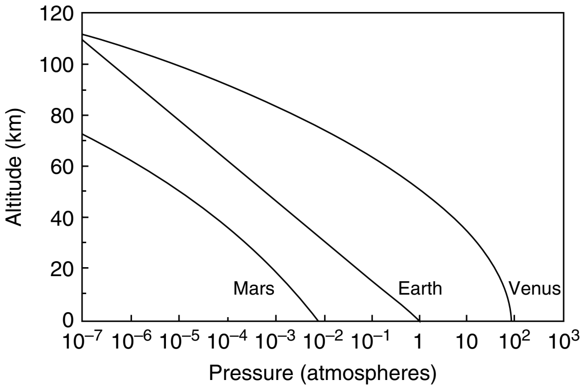


Fig. 5 Model pressure distributions for the atmospheres of Venus, Earth and Mars based on observations and theory (after Bauer 1973)

Particles can escape if the kinetic energy exceeds the potential energy of the gravitation field. If the lowest velocity for escape is v_e then

$$mv_e^2/2 = GMm/R^2 = mgR \tag{13}$$

or

$$v_e = (2gR)^{1/2} = (2MG/R)^{1/2}. \tag{14}$$

This gives $v_e = 11.2(M/R)^{1/2}\,\mathrm{km\,s^{-1}}$,

where M and R are the mass and radius of a planet in terms of the mass and radius of Earth. For Earth $v_e = 11.2\,\mathrm{km\,s^{-1}}$.

$$\textit{Also}\; v_e/v_m = (2gR)^{1/2}/(2kT/m)^{1/2} = (R/H)^{1/2}. \tag{15}$$

For Earth this comes out to be approximately 8. This means $v_e > 8v_m$.

Thus particles with velocity at the tail of the distribution only will be able to escape. As $mv^2/2 = 3kT/2$, the escape temperature T_e can be defined corresponding to the escape velocity. These are 84,000 K for O, 21,000 K for He and 5,200 K for H.

In addition particles must move upward and do not collide. If h_1 is the altitude at the starting point and h_2 the altitude of the point at an angle θ from the vertical, where we are finding the number of particles that are not stopped, then the collisions are proportional to the distance $(h_2–h_1)\ \sec\theta$, number density and the physical cross section α. The probability of escape $p(\theta)$, along the direction θ, can thus be estimated. The average probability for escape comes out to be

$$\int [\{\exp(-\beta\mu)\}/\mu^2]d\mu \cdot [\beta = N_1\sigma H : \mu = \sec\theta] \tag{16}$$

For $\beta = 0.25$ the average probability $= 50\%$. That is when half of the particles with velocity $v > v_e$ will be able to escape. This corresponds to the height $h_x \sim 600$ km for Earth.

The number density $N_x = \beta/\sigma H = 2.5 \times 10^7$ particles/cm^3.

At these altitudes due to diffusive separation helium and hydrogen are at the top. Because of the lower mass value (m) they have a higher thermal velocity ($v^2 = 3kT\,m^{-1}$) so larger fraction of these gases have $v > v_e$.

The effect of planetary rotation on the escape flux as a function of latitude for different values of the escape parameter and rotational parameter is shown in Fig. 6. Here the escape parameter X is defined

$$X = (v_\infty/u_o)^2,$$

where v_∞ is the minimum escape velocity and u_o the most probable velocity.

The rotational parameter Y is defined as

$$Y = (\Omega R/u_0),$$

where Ω is the angular rotational velocity and R the radial distance from the center of the planet.

On a rotating planet particles with velocities in the direction of rotation will obtain the escape velocity more easily than those moving in the opposite direction. The escape flux is several times greater at the equator than at high latitudes. The escape energies for atmospheric constituents from the exospheres of Mercury, Mars and Venus are shown in Fig. 7.

7 Diffusion

The effect of turbulent mixing or molecular diffusion (due to the gradients in the relative concentration as a result of slight deviations from the Maxwellian distribution) is to transport them with a vertical flux. Diffusion becomes more important at higher altitudes as collisions decrease. The vertical drift velocity due to the pressure gradient is given by

$$w = -(D/n)\,\partial n/\partial z. \quad [D \text{ is the diffusion coefficient}] \tag{17}$$

$$\text{Also } \partial p/\partial z = kT\,\partial n/\partial z.$$

Equating this to the drag force due to the collisions $nm\,\nu w$

$$kT\,\partial n/\partial z = -nm\,\nu w. \tag{18}$$

Combining the two equations we get the diffusion coefficient given by

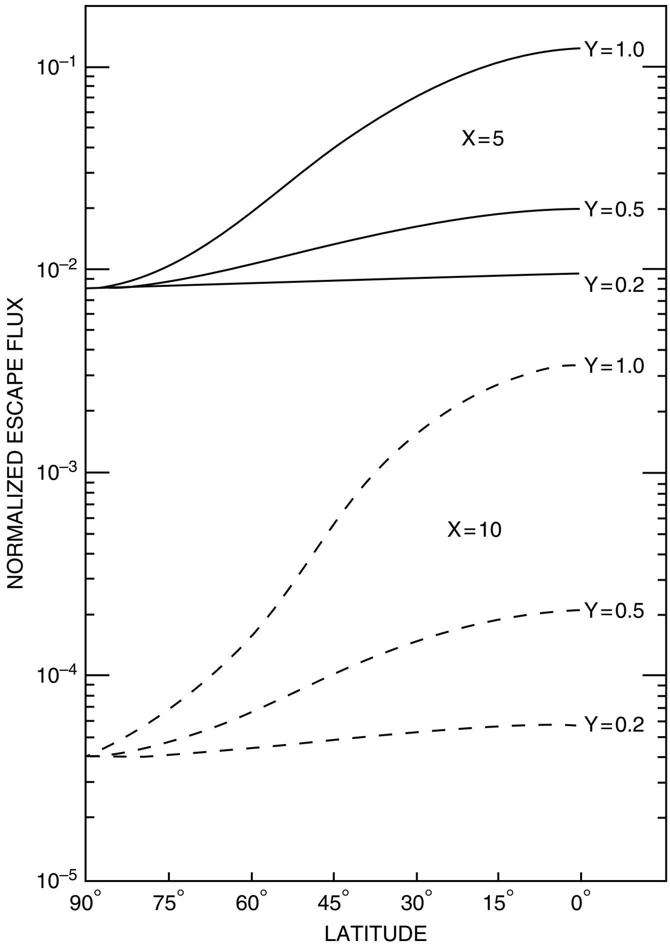


Fig. 6 Effect of planetary rotation on the escape flux as a function of latitude, for parametric values of the escape parameter $X = (v_\infty/u_o)^2$ and the rotational parameter $Y = \Omega R/u_o$. v_∞ is the minimum escape velocity $(2gR)^{1/2}$ and u_0 is the most probable velocity at the temperature T_∞ (after Hartle 1971)

$$D = kT/m\nu. \tag{19}$$

The collision frequency decreases with height so the diffusion coefficient increases. As ν is proportional to $nT^{1/2}$

$$D \propto T^{1/2}n^{-1}.$$

For a vertical motion the gravitational force must be included. Therefore

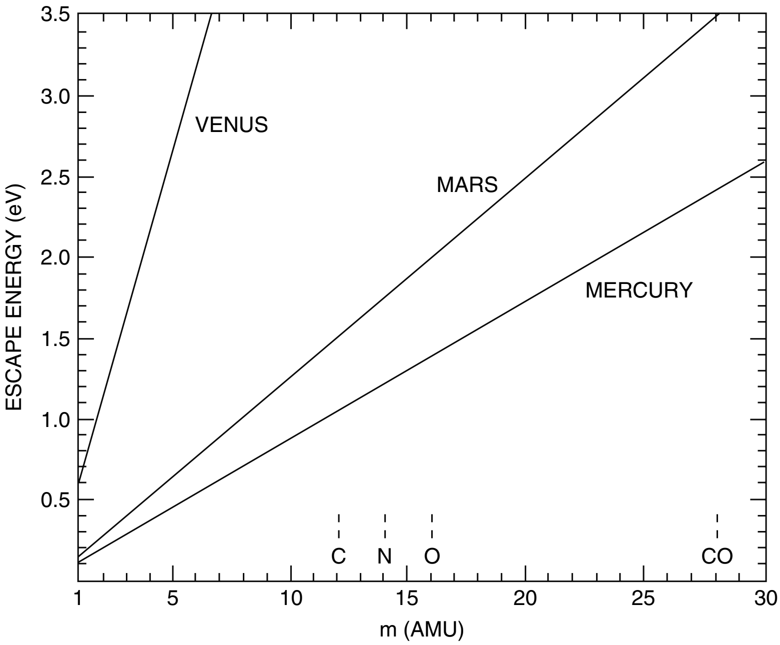


Fig. 7 Energy required for escape of atmospheric constituents from the exospheres of Mercury, Mars and Venus (after Bauer 1973)

$$nm\,\nu w = -dp/dz - nmg. \tag{20}$$

The values of p, D and the scale height of the neutral atmosphere H_n are

$$p = nkT,\ D = kT/m\nu \quad \text{and} \quad H_n = kT/mg.$$

Hence

$$nw = -D[(dn/dz) + (n/H_n)]. \tag{21}$$

Therefore

$$w = -(D/n)[(dn/dz) + (n/H_n)] = -(D/n)(dn/dz) - D/H_n. \tag{22}$$

Including the effect of turbulent mixing the vertical flux can be described by

$$\phi_i = -(D_i + K)[(dn_i/dz) + (n_i/T)(dT/dz)] - n_i[D_i/H_i + (K/H)]. \tag{23}$$

Here D_i is the diffusion coefficient for the ith species and K is the eddy diffusion coefficient. Further $\phi_i = 0$ under equilibrium condition.

When $D_i \ll K$; $K(dn_i/dz) = -n_i K/H$, [under isothermal case $(dT/dz = 0)$] or $(dn_i/dz) = -n_i/H$ in the turbosphere.

When $K \ll D_i$; $(dn_i/dz) = -n_i/H_i$. [for isothermal condition]
At the turbopause $D_i = K$.

The time constant for the molecular diffusion (the characteristic time for attaining diffusive equilibrium) is given by

$$t_{Dj} = H^2/D_j = (H^2/b_j)\,n(h). \tag{24}$$

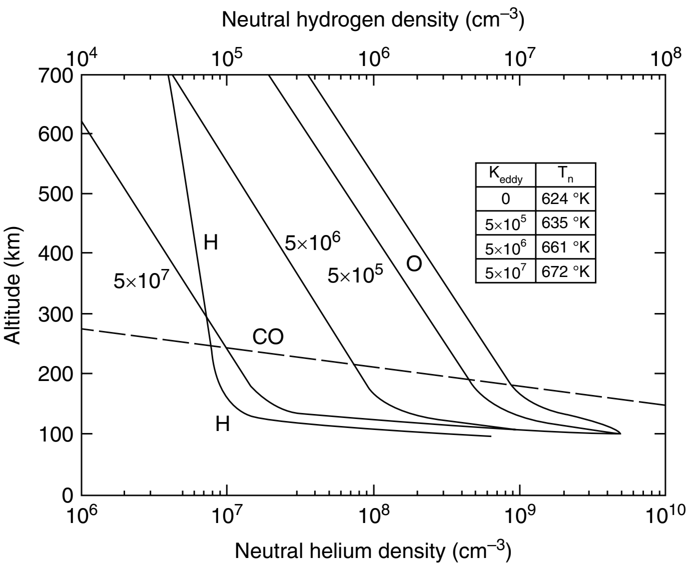
Here b_j is the collision term. Thus the diffusive equilibrium is attained in a shorter time at a higher altitude.

Similarly the time constant for turbulent mixing is given by

$$t_{KD} \approx H^2/\mathrm{K_D}. \tag{25}$$

At the turbopause $\tau_{Dj} = \tau_{KD}$.

The effects of eddy and thermal diffusion are illustrated in Fig. 8 for the atmosphere of Venus. The figure shows the distribution of hydrogen and helium in a CO_2 atmosphere at a temperature of 650 K. Helium distributions for different eddy diffusion coefficients are shown. The distribution for hydrogen is not significantly dependent on eddy diffusion or thermal diffusion because of the strong upward diffusion supporting the escape flux. The effect of thermal diffusion is also included



K_eddy	T_n
0	624 °K
5×10^5	635 °K
5×10^6	661 °K
5×10^7	672 °K

Fig. 8 Distribution of hydrogen and helium in a CO_2 atmosphere at $T = 650$ K (representative of Venus atmosphere). Helium distributions for different eddy diffusion coefficients K_D are given and also for thermal diffusion ($K_D = 0$). Below 200 km scale heights of the light constituents are approximately that of CO_2 due to maximum upward flow. At higher altitudes they approach that of a diffusive equilibrium (after Herman et al. 1971)

and can be noticed for the case with eddy diffusion equal to zero. Below 200 km, the scale heights of the light constituents are approximately that of CO_2 but approach that of diffusive equilibrium at higher altitudes.

The density distribution of a minor constituent under flow and the diffusive equilibrium for a constant value of the scale height is shown in Fig. 9. While in the main mixed atmosphere the density distribution is governed by $\exp(-z/H)$ in the diffusive equilibrium condition it is governed by $\exp(-z/H_j)$. Here H is the scale height of the mixed atmosphere and H_j for the minor constituent.

8 Atmospheric Dynamics

Dynamics of the atmosphere is due to three forces acting on a given quantity of air: (1) gravity, (2) hydrostatic pressure and (3) friction.

The effect of Earth's rotation on its axis from west to east with a period of 1 day results in the centrifugal force ($m\omega^2 r$) and the angular momentum ($m\omega r^2$) on any object on the surface or above. Those have a maximum at the equator and equal to zero at the poles (r is equal to $R\cos\theta$, and it is a function of latitude).

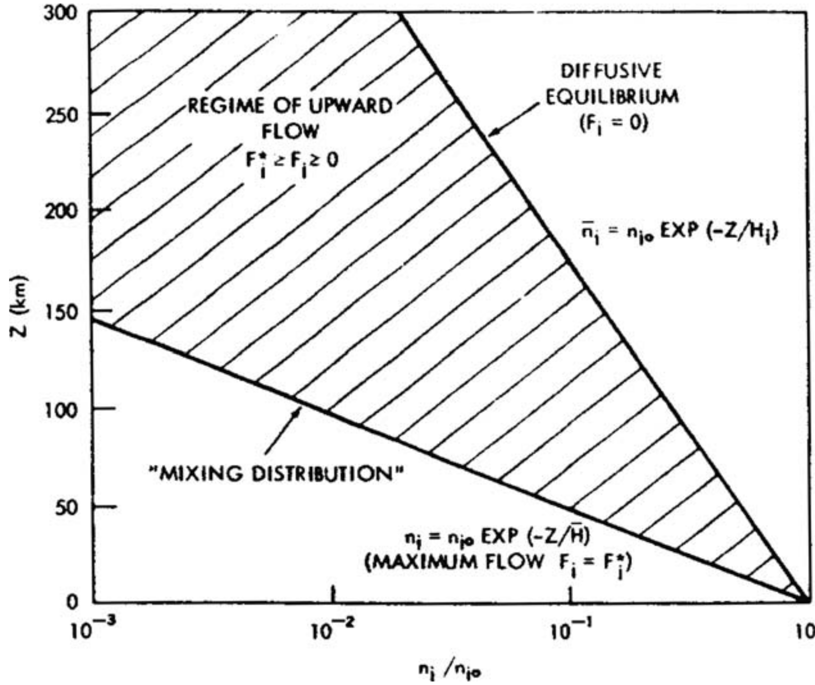


Fig. 9 Idealized density distribution of a minor constituent in an isothermal atmosphere under mixing, diffusive equilibrium and flow conditions (after Bauer 1973)

The rotation of Earth introduces a fictitious force, the Coriolis force, on a wind.

8.1 Equations of Motion

Newton's second law of motion applied to an element of fluid of density ρ moving with velocity $\mathbf{V}$ in the presence of pressure gradient ∇p and gravitational field ($\mathbf{g}$') is given by the Navier-Stokes equation (following Houghton 1977)

$$D\mathbf{V}/Dt = \mathbf{g}' - (\nabla p/\rho) + \mathbf{F}, \tag{26}$$

where $\mathbf{F}$ is the frictional force acting on the element. The equation applies to an absolute or inertial frame of reference. In the frame with respect to Earth's surface, which is rotating with angular velocity $\mathbf{\Omega}$, we need an expression appropriate to the rotating frame.

A vector $\mathbf{A}$ in a frame rotating at the angular velocity $\mathbf{\Omega}$ will have a component of motion $\mathbf{\Omega} \times \mathbf{A}$ in the inertial frame due to the relative motion of the two frames. So in the rotating frame the equation becomes

$$D\mathbf{V}/Dt + 2\mathbf{\Omega} \times \mathbf{V} + \mathbf{\Omega} \times \mathbf{\Omega} \times \mathbf{r} = \mathbf{g}' - (\nabla p/\rho) + \mathbf{F}, \tag{27}$$

or

$$D\mathbf{V}/Dt + 2\mathbf{\Omega} \times \mathbf{V} = \mathbf{g} - (\nabla p/\rho) + \mathbf{F}, \tag{28}$$

where $\mathbf{g} = \mathbf{g}' - \mathbf{\Omega} \times \mathbf{\Omega} \times \mathbf{r}$ is the acceleration due to gravity and includes the centrifugal term. The term $2\mathbf{\Omega} \times \mathbf{V}$ is the Coriolis term and it is perpendicular both to the direction of the motion and the Earth's axis of rotation.

If u, v, w are the components of the velocity in x, y, z directions and **i, j, k** are unit vectors along each axis, one can expand the terms in the equation. Considering the motions (synoptic scales or larger) with systems of scales of 1,000 km in the horizontal dimension and the vertical scale of about 1 scale height (10 km), the vertical velocities are much smaller than the horizontal velocities. Therefore to a first approximation, the equation of motion can be simplified to

$$D\mathbf{V}/Dt = f\mathbf{V} \times \mathbf{k} - (\nabla p/\rho) + \mathbf{F}, \tag{29}$$

where $f = 2\Omega \sin\phi$, [ϕ is the latitude]

and

$$D\mathbf{V}/Dt = \mathbf{i}Du/Dt + \mathbf{j}Dv/Dt$$

8.2 Geostrophic Approximation

For a large-scale motion away from the surface, the friction **F** is small. Further for a steady flow with a small curvature

$$D\mathbf{V}/Dt \approx 0.$$

The resulting motion is then geostrophic. The geostrophic velocity $\mathbf{V}_\mathrm{g}$ is given by

$$f\mathbf{V} \times \mathbf{k} = (\nabla p/\rho) \tag{30}$$

and flows parallel to isobars with a speed which is proportional to the horizontal pressure gradient. In the northern hemisphere, wind flows in a clockwise direction around centers of high pressure (anticyclones) and anticlockwise around centers of low pressure (depressions or cyclones).

The geostrophic approximation works well at heights above about 1 km (for lower heights friction becomes important) and for latitudes greater than about 10°. The Coriolis term is zero at the equator. For lower heights the velocity is no longer parallel to isobars.

8.3 Cyclostrophic Motion

At low latitudes f is small and for a motion having a large curvature (like for a tropical cyclone) another approximation is the cyclostrophic approximation in which the acceleration of the air towards the center is balanced by the pressure gradient:

$$V^2/r = (1/\rho)\,\partial p/\partial r. \quad [r \text{ is distance, center considered origin}] \tag{31}$$

If both the acceleration term and the Coriolis term are included, the resulting solution is known as gradient wind:

$$V^2/r + (1/\rho)\,\partial p/\partial r + fV = 0. \tag{32}$$

8.4 Equation of Continuity

The equation of continuity states that the net flow of mass into unit volume per unit time is equal to the local rate of change of density:

$$\mathrm{Div}\,\rho\mathbf{V} = -\partial r/\partial t. \tag{33}$$

For an incompressible fluid

$$\mathrm{Div}\,\rho\mathbf{V} = 0. \tag{34}$$

8.5 Equation of State

For an ideal gas

$$p = \rho RT. \tag{35}$$

8.6 General Circulation

The broad pattern of air movement over the globe as a result of averaging winds over long time is known as the general circulation of the atmosphere. The basic pattern of surface winds and pressure fields over Earth due to solar heating gives rise to convection cells like a Hadley cell from the equator to the latitudes of $\pm 30°$ in the vertical north-south plane. There are two (or more) weaker cells from the latitude of 30° to the poles. The warm air at the equator rises and spreads out poleward at some high elevation. The rising warm air is replaced near the surface by cooler air, which is drawn towards the equator from north and south. The rotation of Earth deflects the air moving towards the equator towards west. Cold air over the poles descends to the ground and moves equatorward. Thus the tropical and polar cells are thermally driven and the middle latitude cell is frictionally driven by the two cells that border it. The Coriolis force deflects the poleward moving air to eastward and the equatorward moving air to westward.

At low latitudes winds are westerly for 0–20 km altitude. Between 20 and 40 km winds are westerly in winter and easterly in summer. For the region at 40–80 km altitude these are westerly throughout the year. Between 80 and 120 km winds are easterly in winter and westerly in summer.

8.7 Thermospheric Winds

The equation of motion for winds in the thermosphere is

$$\partial\mathbf{V}_n/\partial t + (\rho_i/\rho_n)n_{in}(\mathbf{V}_n - \mathbf{V}_i) = g - (\nabla p_n/\rho_n) + 2\Omega\times\mathbf{V}_n + (\eta/\rho_n)\nabla^2\mathbf{V}_n. \tag{36}$$

Here $\mathbf{V}_n$, $\mathbf{V}_i$ are the velocity of neutral air and the ion gas, ν_{in} is the ion-neutral collision frequency, η is the coefficient of viscosity. The second term on left is the ion drag that arises due to the interaction between the neutral atmosphere and ionosphere. The winds must also satisfy the conservation of mass through the continuity equation for the neutrals and for the ions. The global thermospheric wind system at

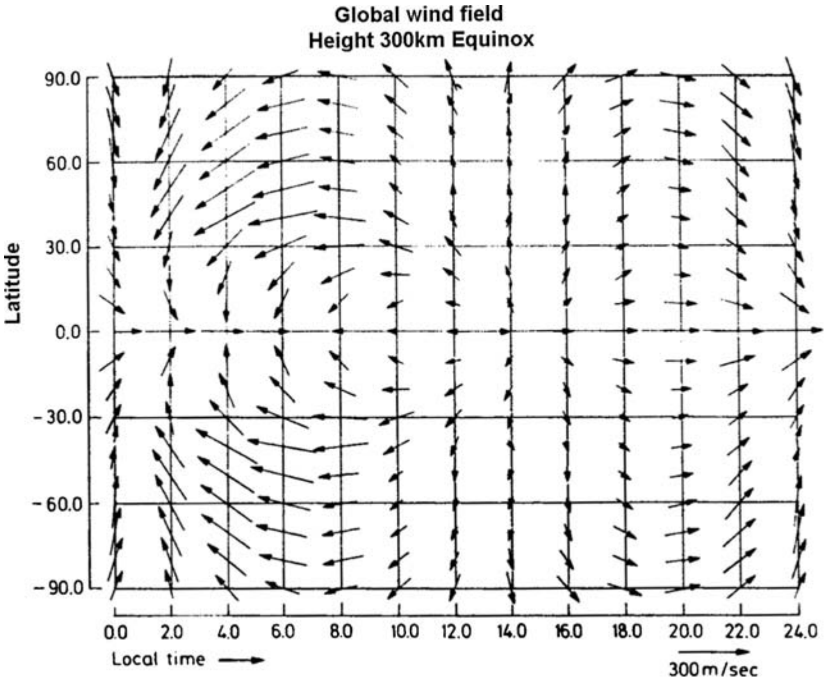


Fig. 10 Global thermospheric wind system at equinox using model values of $\nabla p_n/\rho_n$ based on satellite drag data and ion drag based on semi-empirical model ionosphere (after Blum and Harris 1975)

300 km for Earth is shown as a function of latitude and time for equinox condition in Fig. 10 using model values of $\nabla p_n/\rho_n$ based on satellite drag data and ion drag data based on a semi-empirical model ionosphere (Blum and Harris, 1975).

Atmospheric winds show a number of oscillations and wave-like motions superimposed on the general atmospheric circulation. These are classified, in terms of restoring forces, into (a) internal gravity waves, arising due to the stratification or buoyancy of the atmosphere with periods ranging from a few minutes (Brunt-Vaisala period) to the inertial period and (b) Rossby or planetary waves, arising due to the variation of the Coriolis force with latitude. Planetary waves are large-scale waves and periods near 2, 5, 10 and 16 days have been identified in the atmosphere. The inertio-gravity waves arise due to the combined stratification and Coriolis effects. The tides are the response of the atmosphere to some forcing like the lunar gravitational force with a periodicity of 12.4 h or to solar heating (fundamental period of 24 h with harmonics of 12 and 8 h). The waves propagate upward and in a lossless atmosphere the amplitude of the waves increases as the energy flux is conserved. The upward propagation is limited by the kinematic viscosity of the atmosphere and also by reflections from the thermal barrier ($dT/dz > 0$); so only certain waves propagate upward.

9 Other Planetary Atmospheres

A brief description of the atmospheres of other planets is given here based on published works (Banks et al. 1970; Barbato and Ayer 1981; Haider and Singhal 1994).

Mercury has an extremely rarified atmosphere because of the low gravity and the low escape velocity ($\sim$4 km s^{-1}). The surface pressure is 2×10^{-9} mb. Mariner 10 spectrometer results revealed the presence of hydrogen, helium, neon, argon, oxygen, zenon and carbon in Mercury's atmosphere. The presence of hydrogen and helium is due to solar wind fluxes from the Sun as Mercury has a weak magnetic field. Radioactive decay of elements within the crust is also a source of helium as well as sodium and potassium.

Venus is much denser and hotter than Earth. The surface pressure is 93 b. Spectrometric measurements from Earth had detected CO_2 atmosphere in Venus and also sulfur compounds. The gas chromatograph on Pioneer sampled atmosphere at 52, 42 and 22 km levels and found CO_2 about 96%. N_2 showed altitude variation. Venera 8 probe detected NH_3 in the lower atmosphere and a small concentration of water vapour (0.5%). The Venusian atmosphere also has sulfuric acid as a result of photochemical reactions that give rise to sulfur oxides that could combine with water vapor in the cooler atmospheric layer to form it. The turbopause is located at 140 km and the exobase at 160 km. The atmosphere of Venus is in state of vigorous circulation and super-rotation with winds blowing as fast as 100 m s^{-1}. Venus absorbs the bulk of the solar radiation within and above its layers of clouds. Also due to a slow rotation the Coriolis force is negligible. Based on wind measurements on Venera and Pioneer probes Taylor et al. (1979) derived a model circulation for the 70–110 km altitude region (Fig. 11). It shows the winds in the driving cell (70–90 km) and the return flow in the stratosphere (90–110 km). Venus lacks a magnetic field and its ionosphere separates the atmosphere from the outer space and the solar wind. The atmospheric pressure and temperature at about 50–65 km above the surface is nearly the same as that of Earth. The troposphere extends up to 65 km. The circulation in the troposphere follows the cyclostrophic approximation. The mesosphere extends from 65 to 120 km and the thermosphere begins at 120 km. The ionosphere extends from 120 to 300 km. It has an induced magnetosphere caused by the Sun's magnetic field carried by the solar wind. The ionopause (the boundary between the solar wind and the region of charged particles) varies between 250 and 1,500 km depending upon the ram pressure due to the solar wind.

Mars has a very thin atmosphere with the surface pressure of 7 mb. A neutral mass spectrometer aboard Viking 1 and 2 provided the first measurements of the neutral atmosphere of Mars. The number densities of different constituents in the upper atmosphere of Mars are shown in Fig. 12 (Singhal and Whitten 1988). Carbon dioxide is the major constituent below 180 km. Primary constituents are CO_2 (95.3%), nitrogen (2.7%), argon (1.6%), oxygen (0.13%), CO (0.07%) and water vapor (0.03%). Most of the CO_2 is lost. It also shows large seasonal variation associated with the formation of polar caps (frost point is 148 K). A small amount of ozone is also present. Mars also has a lot of dust in its atmosphere, and winds occasionally cause dust storms.

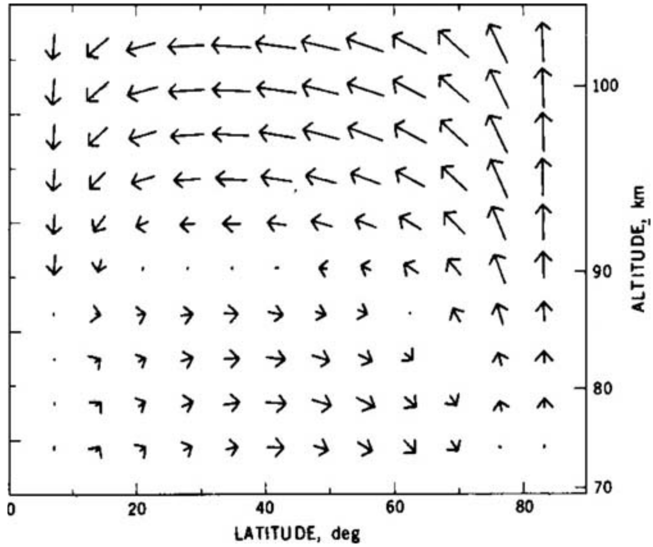


Fig. 11 Model of general circulation of Venus from 70 to 110 km (after Taylor et al. 1979)

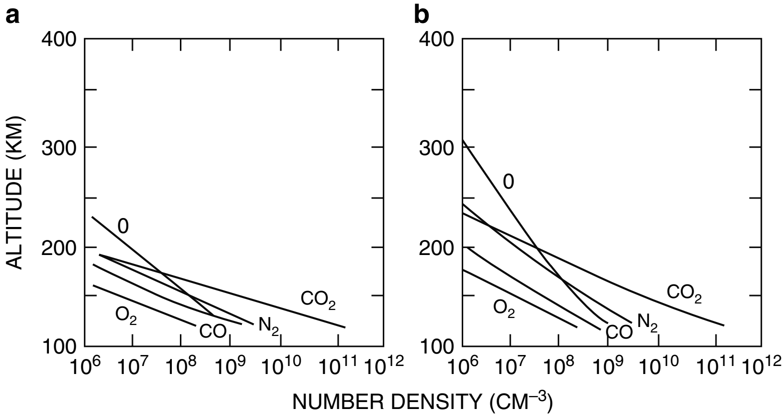


Fig. 12 Neutral upper atmospheric densities for Mars: (**a**) dayside and (**b**) nightside (after Singhal and Whitten 1988)

Outer planets like Jupiter and Saturn have an internal heat source; so the effective temperatures are far in excess of their equilibrium temperatures. In the absence of a planetary surface, pressure levels are referred to temperature levels with bottom of Jupiter defined as the point where the atmospheric pressure equals 100 Earth's atmosphere. For Jupiter the temperature at 1 atmosphere is $\sim$165 K and the tropopause temperature at 0.1 atmosphere is $\sim$113 K. Ground based optical telescopes and Pioneer 10 satellite data has shown Jupiter's atmosphere to consist of hydrogen, helium, ammonia, methane, phosphine and water vapor. Voyager probes

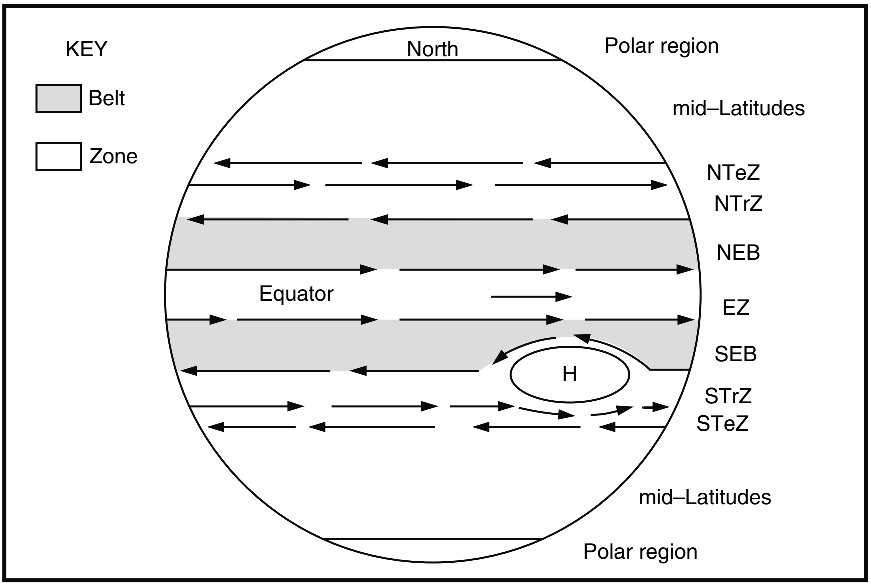


Fig. 13 The belts and zones of Jupiter. The arrows represent high velocity winds located at the boundaries (after Barbato and Ayer 1981)

have also detected acetylene and ethane. The Jupiter's atmosphere is divided into the troposphere, stratosphere, mesosphere and thermosphere. Troposphere extends from within the planet to a pressure of 100 mb (0.1 Earth atmosphere). Stratosphere extends from 100 to 0.1 mb. The temperature increase is due to the absorption of solar radiation by a layer of methane along with ethane, acetyline and phosphine. The clouds and wind pattern are predominantly zonal and divided as belts (high thermal temperature, low cloud height, cyclonic rotation, dark color clouds), zones (low thermal temperature, high cloud height, anti-cyclonic rotation, upward vertical motions and light color clouds) and Great Red Spot (low thermal temperature, high cloud height, anti-cyclonic rotation, upward vertical motions and thick reddish-orange clouds). The global distribution of cloud regions and the nomenclature is shown in Fig. 13 (Barbato and Ayer 1981). Winds are zonal with highest velocities of 130–150 $\mathrm{m\,s^{-1}}$.

The composition of Saturn's atmosphere is similar to Jupiter's atmosphere. The temperature at the reference level 1 bar is $\sim$140 K. The tropopause at 100 mb has a temperature of $\sim$85 K. It has a thin layer of ionosphere.

The atmospheres of Uranus and Neptune are primarily composed of methane and hydrogen. Uranus has a temperature of 100 K at 10^{-3} mb. Neptune shows a temperature of 104 K at 10^{-3} mb.

Acknowledgements Author would like to thank Director, Physical Research Laboratory, Ahmedabad for providing the facilities to work at PRL. Sincere thanks are also due to Dr. S A Haider for many discussions and to Som Sharma for help.

References

Banks, P.M., Johnson, H.E., Axford, W.I.: The atmosphere of mercury, comments astrophys. Space Phys. **II**/6, 214, (1970)

Barbato, J.P., Ayer, E.A.: Atmospheres. Pergamon, New York (1981)

Bauer, S.J.: Physics of planetary ionospheres. In: Roederer, J.G. (ed.) Physic and Chemistry in Space, vol. 6. Springer, Berlin (1973)

Blum, P.W., Harris, I.: J. Atmos. Terr. Phys. **37**, 193 (1975)

Degaonkar, S.S.: An introductory course on space science and Earth's environment. Gujarat University, Ahmedabad, India (1975)

Giraud, A., Petit M.: In: McCormac, B.M. (ed.) Ionospheric Techniques and Phenomena. Geophysics and Astrophysics Monographs. D. Reidal Publishing Company, USA (1978)

Haider, S.A., Singhal, R.P.: In: Varma, R.K. (ed.) Advances in Space Research in India. Indian National Science Academy, New Delhi, p. 69. (1994)

Hartle, R.E.: Phys. Fluids **14**, 2592 (1971)

Herman, J.R., Hartle, R.E., Bauer, J.: Planet. Space Sci. **19**, 443 (1971)

Houghton, J.T.: The Physics of the Atmosphere. Cambridge University Press, Cambridge (1977)

Hunten, D.M.: J. Atmos. Sci. **26**, 827 (1969)

Singhal, R.P., Whitten, R.C.: Icarus **74**, 357 (1988)

Taylor, F.W., Diner, D.J., Elson, L.S., McCleese, D.J., Martonchik, J.V., Delderfield, J., Bradley, S.P., Schofield, J.T., Gille, J.C., Coffey, M.T.: Science **205**, 65 (1979)

Planetary Ionospheres

Nanan Balan

Abstract The paper presents a summary of the lectures on *planetary ionospheres* given at NASA's 1st Asia Pacific School on International Heliophysical Year conducted at Indian Institute of Astrophysics, Kodaikanal, India during 10–22 December 2007. Following an introduction, the paper describes the structure of the ionospheres, theory of Earth's ionosphere including the effects of diffusion, neutral wind and electric field, and ionospheric electric fields, currents and variability.

1 Introduction

Ionosphere is the ionized part of the upper atmosphere. In the case of Earth, ionosphere extends from about 70 to 1,000 km height, with peak at around 300 km on the average. It is a weakly ionized plasma with ionized to neutral particle density ratio reaching a maximum of about 1:500. The region above ionosphere is plasmasphere, above which is magnetosphere. The ionosphere is embedded in neutral atmosphere (mesosphere-thermosphere region), the density of which decreases exponentially with height. There are several books on ionosphere (e.g., Rishbeth and Garriott 1969; Ratcliffe 1972; Mahajan and Kar 1988; Kelley 1989).

The ionosphere has been studied for its scientific and practical importance. It influences the propagation of electromagnetic waves and hence has advantages and disadvantages in communication. It is to exploit the advantages in radio communication and find remedies for the disadvantages in all branches of tele-communication that the ionosphere has been studied since Marconi's successful transmission of radio signals across the Atlantic. Since ionosphere varies from place to place and from time to time, it is important to study it at as many places as possible and for as long as possible. In this context, ionospheric variations during geomagnetic storms

N. Balan
Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, UK
e-mail: B.NANAN@sheffield.ac.uk

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_12,

form an integral part of *Space Weather*. Ionospheric studies are also important for radio astronomy, to account for the irregular changes in the amplitude and phase (or scintillations) of the radio signals that pass through the ionosphere.

On the scientific side, much of the basic science has been understood. Nevertheless, the ionosphere is being studied with renewed interest using radio and in-situ techniques to understand the science of many interesting phenomena that occur in it and also to understand its interactions with the regions above and below. In addition, the ionosphere is a naturally existing plasma. We can study it at our will and wish, and use the information to understand the problems associated with man-made plasmas, in fusion reactors. The existence of the ionosphere also results in the generation of natural resonators like the *ionospheric waveguide* formed around the plasma density maximum for the propagation of magnetosonic waves, *Alfven resonator* formed between the density maximum and an upper altitude at about 3,000 km for Alfven waves, and *Schumann resonator* formed between the nearly perfectly conducting terrestrial surface and ionosphere.

2 Ionospheric Structure

All planets (except Mercury) have ionospheres. The density of Earth's ionosphere increases with height, reaches a maximum (ionospheric peak) and then decreases for further increase in height (Fig. 1). The vertical structure also includes regions of enhanced density, called D, E and F regions during daytime. The D region occurs below 90 km, E region between 90 and 150 km and F region at heights above 150 km. At night, the ionosphere reduces to a single layer structure called F layer. A weak C region sometimes forms near 70 km. Comparatively strong ionization, populated by ion hydrates, sometimes occurring below 70 km is called cosmic ray layer. The regions below and above the ionospheric peak are referred to as bottomside and topside ionospheres. The electron densities in the E and F regions are high enough to reflect medium and short wavelength radio waves used for long distance radio communications. The densities in the D region and below are not sufficient to reflect radio waves for communication; instead, these regions absorb radio waves.

Figure 2 shows the electron density profiles of Venus's ionosphere at solar maximum and minimum (Knudsen et al. 1987). As shown, the ionosphere is nearly equally strong at solar maximum and minimum, with peak at around 150 km height, though topside ionosphere depletes significantly at solar minimum. The structures of the ionospheres of other planets can be found in Hinson (http://www.star.Stanford.edu/projects/), Atreya et al. (1984) and Mahajan and Kar (1988). Comparing all planets, the ionosphere is stronger in Earth than in other planets.

3 Conservation Laws

The behaviour of the ionosphere is controlled by chemical, dynamical and energy processes described by the conservation laws of mass, momentum and energy. The chemical processes determine the production and loss of ionization (ions and

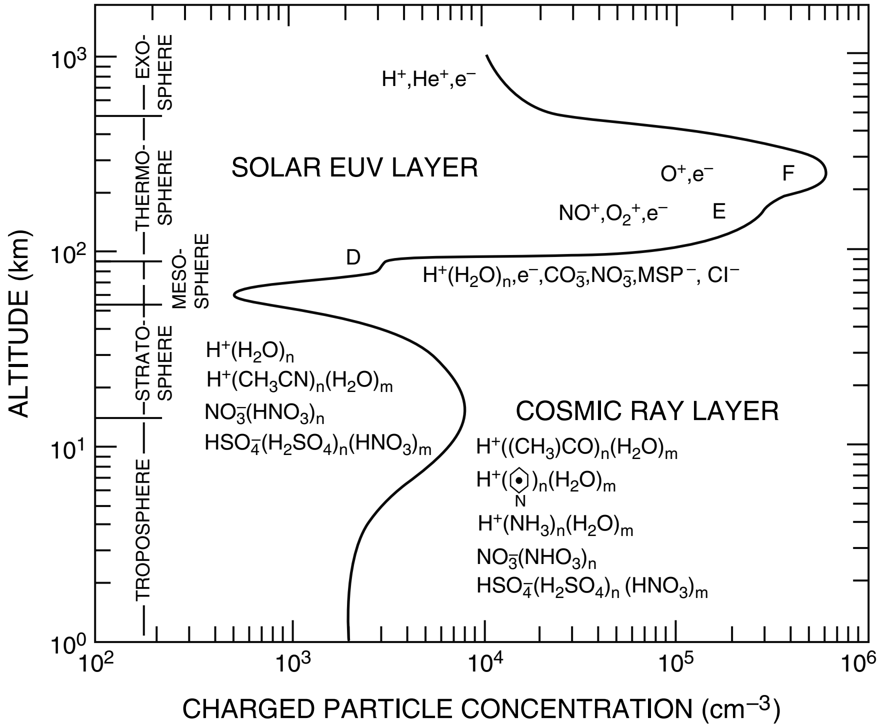


Fig. 1 Typical ionization density profile of Earth's atmosphere; major ion species at different altitudes are listed (Viggiano and Arnold 1995)

electrons) while dynamical processes determine movement of ionization. The rate of change of ionization in the ionosphere can be obtained from the *continuity equation* (or law of conservation of mass)

$$\frac{\delta n_e}{\delta t} = Q - L - \bigtriangledown \cdot (n_e \mathbf{V}), \tag{1}$$

where Q, L and n_e are the rate of production, rate of chemical loss and number density of electrons; $\bigtriangledown \cdot (n_e \mathbf{V})$ is the loss of ionization due to transport, $\mathbf{V}$ being transport velocity.

The motion of charged particles in the ionosphere is controlled mainly by the forces of pressure gradient, gravity, electric field and collisions. The *momentum equation*, which balances these forces, for the jth ion species can be written as

$$\rho_j \frac{d\mathbf{V}_j}{dt} = - \bigtriangledown p_j + \rho_j \mathbf{g} + \frac{q_j \rho_j}{m_j} (\mathbf{E} + \mathbf{V}_j \times \mathbf{B}) - \sum_k \rho_j \nu_{jk} (\mathbf{V}_j - \mathbf{V}_k), \tag{2}$$

where $\mathbf{V}$, ρ, q and m are the velocity, density, charge and mass of ion j, ν_{jk} is the collision frequency of ions j and k, $\mathbf{E}$ and $\mathbf{B}$ are electric field and magnetic field, and

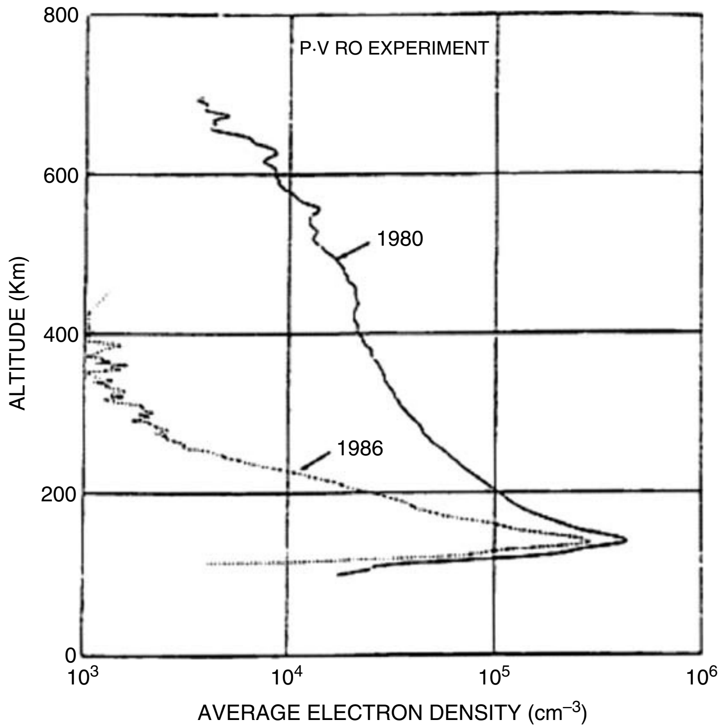


Fig. 2 Electron density profiles of Venus's ionosphere at solar maximum (1980) and solar minimum (1986) obtained by PVO radio occultation (Knudsen et al. 1987)

$j \neq k$. The first term in (2) represents the force due to pressure gradient ($\bigtriangledown p_j$). The terms involving g and $\nu_{jk}(\mathbf{V_j} - \mathbf{V}_k)$ are the forces due to gravity and collision (or friction) between ions j and k, and terms involving $\mathbf{E}$ and $\mathbf{B}$ are the forces due static electric field ($\mathbf{E}$) and Lorentz electric field ($\mathbf{V}_j \times \mathbf{B}$). Since there are different ion species, a complete picture of the plasma motion would require treatment of a large number of coupled equations. The forces of viscosity, Coriolis acceleration, centripetal acceleration, and tides and waves are not included in (2).

Theoretical values of electron and ion temperatures can be obtained by solving the *energy equations*. The electron energy equation is

$$\frac{3}{2}n_e k \frac{\delta T_e}{\delta t} = -n_e k T_e \bigtriangledown \cdot V_e - \frac{3}{2} n_e k V_e \cdot \bigtriangledown T_e - \bigtriangledown \cdot \mathbf{q}_e + \sum Q_e - \sum L_e, \quad (3)$$

where $\mathbf{q}_e = -\lambda_e \bigtriangledown T_e$ is the electron thermal heat flux, with λ_e being the thermal conductivity coefficient. The first term on the right-hand side represents adiabatic expansion, second term accounts for advection, third term is the divergence of electron heat flow, and $\sum Q_e$ and $\sum L_e$ are the sum of all the heating and cooling rates.

The ion energy equation, similar to the electron energy equation, is given by

$$\frac{3}{2}N_i k \frac{\delta T_i}{\delta t} = -N_i k T_i \bigtriangledown \cdot V_i - \frac{3}{2}N_i k V_i \cdot \bigtriangledown T_i - \bigtriangledown \cdot \mathbf{q}_i + \sum Q_i - \sum L_i , \quad (4)$$

where $\mathbf{q}_i$ is ion thermal heat flux, $\sum Q_i$ is the sum of all heating rates (mainly thermal ion heating due to Coulomb interactions with thermal electrons), and $\sum L_i$ is the sum of all the ion cooling rates (mainly to neutral gas).

4 Theory of Photoionization

The ionosphere is formed mainly from the photoionization of the atmosphere by solar X-rays and EUV radiations; high energy charged particles from the Sun also contribute to the ionization at high latitudes, which is not considered here. The intensity of solar radiations (X-rays and EUV) decreases due to absorption as they penetrate through the atmosphere while the concentration of the ionizable constituents increases. That causes the ionization to maximize at a height where the intensity of the radiation and concentration of the constituents optimize.

4.1 Simplified Production Function

Considering a plane horizontally stratified atmosphere composed of a single gas constituent on which a monochromatic solar radiation is incident at a zenith angle (angle with the vertical) χ, the rate of production of ionization Q is given by (Ratcliffe 1972)

$$Q = Q_m exp\left[1 + \frac{h_m - h}{H} - exp\left(\frac{h_m - h}{H}\right)\right], \quad (5)$$

where Q_m is the peak rate of production at height h_m and $H = \frac{kT}{mg}$ is the *scale height* of the atmospheric constituent; scale height of a gas is the height at which its density decreases to $1/e$ times its density at a given base level. Using reduced height $z = \frac{h-h_m}{H}$ measured above the peak of production, Q can be written as

$$Q = Q_m \exp\left[1 - z - \exp(-z)\right] . \quad (6)$$

The simplicity of the expression shows that the production layers corresponding to all wavelengths and solar zenith angles (χ) have same shape in terms of the normalized quantities z and Q/Q_m as shown in Fig. 3.

For vertical incidence of solar radiation ($\chi = 0$), Q becomes

$$Q = Q_0 \exp\left[1 - y - \sec\chi \exp(-y)\right] , \quad (7)$$

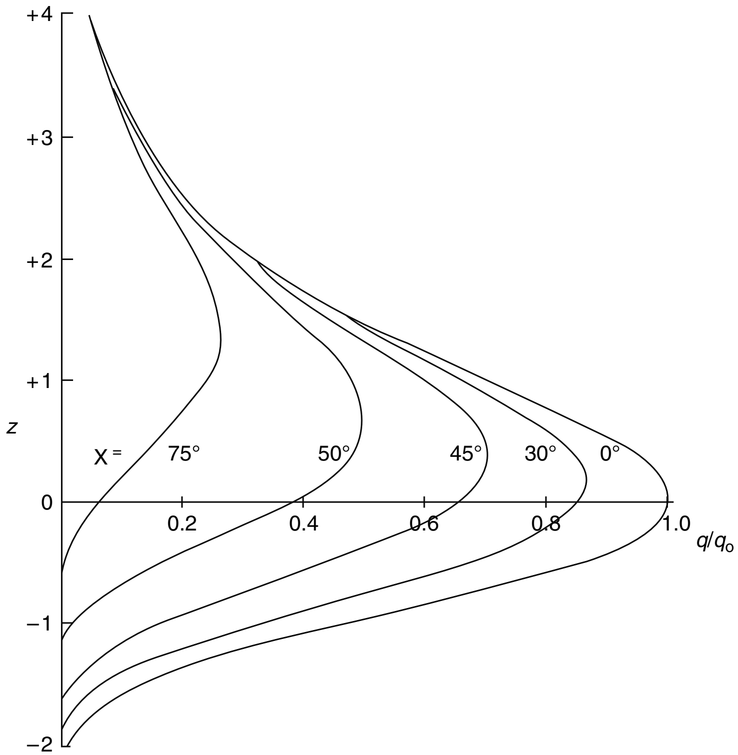


Fig. 3 Normalized Chapman production function versus reduced height z for different solar zenith angles χ

where $y = (z - z_0)/H$ with z_0 being the height of peak production Q_0. For very large y ($y \gg 0$ or $z \gg z_0$), Q takes the form $Q = Q_0 \exp\left(-\frac{z}{H}\right)$. That is, at heights well above the ionization maximum, the ionization decays at the rate determined by the scale height of the ionizable constituent. For very small y ($y \ll 0$ or $z \ll z_0$), Q becomes $Q = Q_0 \exp\left[-\sec \chi exp\left(-\frac{z-z_0}{H}\right)\right]$ which decreases rapidly toward lower altitudes.

4.2 Gracing Incidence and Generalized Production Function

The plane-earth approximation made above is not valid for χ near 90° because $\cos\chi$ reaches zero and therefore there should not be any production (7) while, in reality, significant production occurs even at $\chi = 100°$ when H is large. To overcome this difficulty, Chapman defined a *grazing incidence function* $Ch(x, \chi)$ as

$$Ch(x, \chi) = \left(\frac{1}{2}\pi x \sin\chi\right)^{1/2} \exp\left(\frac{1}{2}\cos^2\chi\right)\left[1 \pm erf\left(\frac{1}{2}x\cos^2\chi\right)^{1/2}\right], \quad (8)$$

where $x = (R_e + h)/H$ and the error function erf is readily available. The function $Ch(x, \chi)$ should replace sec χ near sunrise and sunset. It departs from sec χ when $\chi > 80°$. For example, at 90°, sec χ becomes infinite while $Ch(x, 90°)$ becomes $(\frac{1}{2}\pi x)^{1/2}$.

The other assumptions, such as monochromatic solar radiation and single constituent atmosphere, are also not valid. The atmosphere is composed of different gases of different scale heights and ionization potentials, and solar radiation spectrum consists of a myriad of lines and bands with different intensities. Further, the absorption cross section varies with gas species and photon wavelength. The approach is therefore to assume a neutral atmosphere model with height distributions of some of the major gases in concern, and then for a finite number of wavelength bands derive the production profiles for each gas, as used in mathematical models of the ionosphere (e.g., Schunk and Sojka 1996; Bailey and Balan 1996). A generalized production function Q_i can be defined as

$$Q_i = \sum_{\lambda} I(\lambda)\sigma_i(\lambda)n_i \exp\left(-\sum_{j} \sigma_j(\lambda)n_j H_j Ch_j(\chi)\right), \tag{9}$$

where $I(\lambda)$ is the intensity of solar radiation of wavelength λ, $\sigma_i(\lambda)$ and $\sigma_j(\lambda)$ are the photoionization and photoabsorption cross-sections of the ith and jth neutral gases of densities n_i and n_j; H_j is the scale height of the jth neutral gas and $Ch_j(\chi)$ is the Chapman function of the jth gas. The summation over λ is for the wavelength range of the ionizing radiation and summation over j is for the important absorbing gases O, O_2 and N_2.

4.3 Optical Depth and Ionization Potential

The intensity of solar radiation decreases due to absorption as it passes through the atmosphere. At the peak of ionization production, the intensity reduces to $1/e$ times the intensity at the top of the atmosphere ($I_m = I_\infty/e$); at that altitude the intensity is said to reach one *optical depth*. The optical depth τ at any altitude where the density is n and scale height is H is given by $\tau = \sigma n H$ sec χ. Since sec χ is least when χ is zero, unit optical depth is reached at the lowest altitude for an overhead Sun. The altitude of unit optical depth for an overhead Sun is illustrated in Fig. 4 for the solar spectrum below 3,000 Å. As shown, the different wavelengths have different heights of unit optical depth.

Production of ionization depends also on the ionization potentials V_p of the different species. The ionization and dissociation potentials (and equivalent wavelengths, according to $V_p = h\nu = h\frac{c}{\lambda}$, with h being Plank's constant and ν and λ being photon frequency and wavelength) of the most abundant species in the atmosphere are listed in Table 1. The species get ionized or dissociated by a photon if the photon energy ($h\nu$) is greater than V_p. Typical ionization cross sections in

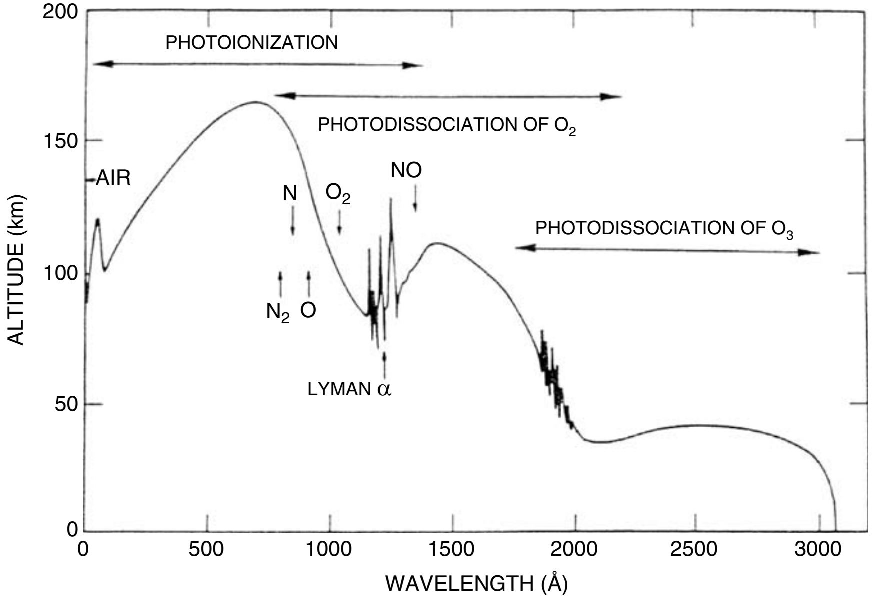


Fig. 4 The altitude of unit optical depth for wavelengths below 3,000 Å. The arrows indicate the wavelength regions of typical photoionization and photodissociation

Table 1 Ionization and dissociation potentials (and equivalent wavelengths, according to $V_p = h\nu = h\frac{c}{\lambda}$, with h being Plank's constant and ν and λ being photon frequency and wavelength) of most abundant species in Earth's atmosphere

Species	Ionization		Dissociation	
	V_p(eV)	λ(Å)	V_p(eV)	λ(Å)
N_2	15.58	796	9.76	1,270
O_2	12.08	1,026	5.12	2,422
O	13.61	911		
N	14.54	853		
NO	9.25	1,340	6.51	1,905
H	13.59	912		
He	24.58	504		

the EUV regime are of the order of 10^{-21}–$10^{-22}\,\text{m}^2$. For the low end of this range and an overhead Sun, unit optical depth of EUV radiation corresponds to $\frac{1}{\sigma} = nH = 10^{22}\,\text{m}^{-2}$, which is satisfied at about 110 km height. That explains why most of the solar EUV radiation is absorbed above 100 km altitude. Photodissociation has a smaller cross-section which satisfies down to 80 km height where $nH \approx 10^{24}\,\text{m}^{-2}$. The collisional cross-section for O_2 is reduced to $10^{-27}\,\text{m}^2$ in the Herzberg continuum (2,026–2,424 Å). This is the reason why these wavelengths can penetrate deeply into the stratosphere where $nH = 10^{27}\,\text{m}^{-2}$, where O_2 can dissociate to form O atoms.

4.4 Chemical Loss of Ionization

Once ionization is produced, a part of it is lost by chemical reactions and another part is removed (or added) by transport mechanisms. Important loss reactions are radiative and dissociative recombinations

$$X^+ + e \rightarrow X + h\nu, \tag{10}$$

$$XY^+ + e \rightarrow X + Y. \tag{11}$$

The radiative recombination (10) removes O^+ ions through $O^+ + e \rightarrow O + h\nu$, with the liberated photon (needed to conserve energy) having optical wavelength. It is a comparatively slow reaction because it is difficult to conserve momentum as it results in only one material particle. A more rapid loss process for O^+ is through a chain of reactions involving N_2 or O_2. The dissociative recombination (11) is important for the loss of N_2^+, O_2^+, and NO^+ ions (primary ions in D and E regions).

At heights where both reactions (10 and 11) are important, charge neutrality requires $n_e = [X^+]+[XY^+]$, and the loss rate L_i should be proportional to the product of electron and ion densities as $L_i = \alpha n_e^2$ where α is the recombination coefficient. Considering photochemical equilibrium (when production balances chemical loss), the continuity equation becomes $Q_i = L_i = \alpha n_e^2 = Q_0 exp[1 - y - \sec\chi\, exp(-y)]$. That gives

$$n_e(z) = \sqrt{\frac{Q_0}{\alpha}} exp\left[\frac{1}{2}\left(1 - y - \sec\chi exp(-y)\right)\right]. \tag{12}$$

This is referred to as α-Chapman profile, which is representative of E region where molecular ions are dominant and transport processes are unimportant.

However, when charge neutrality is maintained with the loss of one electron-ion pair, $L_i = \beta n_e$ where β is the loss coefficient. Photochemical equilibrium then becomes $Q_i = L_i = \beta n_e$, which gives the electron density profile called β-Chapman profile as

$$n_e(z) = \frac{Q_0}{\beta} exp\left[1 - y - \sec\chi exp(-y)\right]. \tag{13}$$

4.5 Formation of Ionospheric Regions

The ionizing radiations have a wide, continuous spectrum superposed by numerous intense emission lines. The atmospheric constituent have different ionization potentials and, in the thermosphere, they are distributed according to barometric law. Therefore, ionization of the different constituents maximizes at different altitudes. That accounts in part to the formation of the different ionospheric regions (Fig. 1).

D region is an ion-neutral collision-dominated region of molecular ions (O_2^+, N_2^+, NO^+) produced by X-rays (<8 Å), EUV (1,027–1,118 Å) and H Ly-α (1,216 Å) radiations. Galactic cosmic rays, the most energetic radiation, provides additional ionization for D region and below. The cosmic rays are so energetic that the secondary energetic particles associated with primary cosmic rays penetrate to ground level before reaching the ends of their paths. The primary and secondary cosmic ray ionization results in the cosmic ray layer (Fig. 1). The E region is produced mainly from the photoionization of O and O_2 by EUV radiation (796–1,027 Å), X-rays (8–140 Å) and H Ly β (1,026 Å) line. The chemical loss-process in the E-region is one of dissociative recombination (discussed below).

The F region, main ionospheric region, contains bulk of the ionosphere. It is produced mainly from the photoionization of O and partly form N_2 and O_2 by EUV (170–796 Å) radiation. The important chemical loss processes in the F region are those leading to the loss of O^+ ions through radiative and dissociative recombinations (discussed below). Although F region ionization production maximizes below 200 km height, the region splits into F_1 and F_2 layers during daytime due to the altitude variations of chemical loss processes and dynamical processes (discussed below), F_1 layer near the peak of ionization production and F_2 layer (main ionospheric layer) at higher heights.

4.6 Formation of F_1 Layer

The main ion (O^+) in the F region is lost through the ion-neutral rearrangement reactions $O^+ + N_2 \rightarrow NO^+ + N$ and $O^+ + O_2 \rightarrow O_2^+ + O$. The resulting molecular ions (NO^+ or O_2^+) are lost by comparatively rapid dissociative recombination process with electrons. The height at which the ion-neutral rearrangement and dissociative recombination becomes equal is called transition height h_t, below which rearrangement dominates and above which dissociation dominates. A clear F_1 layer is formed when the transition height h_t lies above h_0 (height of peak ionization production), which is satisfied during daytime away from noon in the summer of low solar activity. During other daytime periods h_t lies below h_0, and hence no clear F_1 peak is formed at h_0, instead an F_1 ledge is formed (Ratcliffe 1972).

4.7 Plasma Diffusion and F_2 Layer

Plasma diffusion can explain the formation of ionospheric peak (F_2 layer) at heights above the peak of ionization production (e.g., Yonezawa 1956; Ratcliffe. For simplicity, consider vertical plasma diffusion in the absence of magnetic field ($\mathbf{B} = 0$), neutral wind ($\mathbf{U} = 0$) and electric current. Though electrons are lighter, they diffuse together as a single system (ambipolar plasma diffusion) due to the polarization electric field between electrons and ions. The solution of the equations for the vertical motion of electrons and ions, gives the vertical plasma flux (nW) as

$$nW = -\frac{k\,(T_i + T_e)}{m_i \nu_i}\left[\frac{dn}{dz} + \frac{n m_i g}{k\,(T_i + T_e)}\right] \tag{14}$$

$$= -D_p\left(\frac{dn}{dz} + \frac{n}{H_p}\right), \tag{15}$$

where $D_p = \frac{k(T_i+T_e)}{m_i \nu_i}$ is ambipolar diffusion coefficient and $H_p = \frac{k(T_i+T_e)}{m_i g}$ is plasma scale height. If $T_e = T_i$, then $D_p = 2D$ and $H_p = 2H$ where D and H are the diffusion coefficient and scale height of a gas of neutral particles having the same mass and temperature as ions. The solution of plasma diffusion equation

$$\frac{\delta n}{\delta t} = -\frac{\delta}{\delta z}(nW) = \frac{\delta}{\delta z}\left[D_p\left(\frac{\delta n}{\delta z} + \frac{n}{H_p}\right)\right] \tag{16}$$

under equilibrium becomes $\frac{\delta n}{\delta z} = -\frac{n}{H_p}$, which for constant H_p gives

$$n = n_0 \exp\left(-\frac{z - z_0}{H_p}\right), \tag{17}$$

where n_0 is the plasma density at a reference altitude z_0, for example, at F region maximum. The equilibrium electron density profile above the ionospheric peak therefore decays exponentially with altitude with a scale height twice as large as the scale height of the neutral atmosphere (because $H_p = 2H$). The diffusion coefficient D_p increases exponentially with increasing altitude as

$$D_p = D_0 \exp\left(\frac{z - z_0}{H}\right), \tag{18}$$

where D_0 is the diffusion coefficient at z_0.

In the ionosphere; T_e is always higher than T_i (except during frictional heating). Let $T_e = c\ T_i$, with $c > 1$. Then D_p and H_p increase by the factor $(1 + c)/2$ when $T_e > T_i$. Thus, the shape of the electron density profile and plasma diffusion are strongly dependent on the electron-ion temperature ratio.

4.7.1 Effect of Geomagnetic Field

Plasma flux along geomagnetic field (due to diffusion, from 15) and its vertical component are

$$n v_{\|} = -D_p \sin I\left(\frac{dn}{dz} + \frac{n}{H_p}\right), \tag{19}$$

$$nW = n v_{\|} \sin I = -\sin^2 I D_p\left(\frac{dn}{dz} + \frac{n}{H_p}\right). \tag{20}$$

The vertical flux becomes zero at $I = 0$ (equator) where field lines are horizontal.

4.7.2 F_2 Layer

As discussed in the formation of F_1 layer (Sect. 4.6), the electron density above the F_1 layer increases (due to slow decrease of the production rate Q compared to chemical loss coefficient β). At high altitudes, where both Q and βn_e are unimportant and electron density is controlled mainly by diffusion, the electron density decreases exponentially with increasing altitude as discussed above. Thus, the electron density must maximize at an altitude in between these altitude regions. That is where the ionospheric peak (or the F_2 peak) occurs. Under steady state at low altitudes, $Q = \beta n_e$ (which is β Chapman relation). At high altitudes, where $Q \approx 0$ and $\beta \approx 0$, the steady state solution gives

$$n = n_{max}\left(-\frac{z - z_{max}}{H_p}\right), \tag{21}$$

$$z_{max} = \frac{H}{1+\lambda} ln \frac{H H_p \beta_0}{D_0 \sin^2 I}, \tag{22}$$

where H is the scale height of a mixture of N_2 and O_2 and $\lambda = m(O_2, N_2)/m(O)$ is their total mass compared to atomic oxygen mass.

4.8 *Effect of Neutral Wind and Electric Field*

As shown above, ionospheric peak occurs at altitudes above the peak of ionization production due to plasma diffusion. The peak can be shifted by neutral air winds and electric fields. A neutral wind of meridional component U_θ and zonal component U_ϕ produces a wind U in the magnetic meridian $U = U_\theta \cos\alpha \pm U_\phi \sin\alpha$ where α is magnetic declination angle and opposite signs correspond to opposite hemispheres. A horizontal wind blowing towards magnetic equator with speed U drives F region plasma up the geomagnetic field lines with a speed $U \cos I$, of which vertical component is $U \sin I \cos I$, with I being dip angle. This upward drift raises the ionospheric peak to altitudes of reduced chemical loss and so increases peak electron density. An opposite effect happens when the wind blows poleward. Very roughly, the rise in peak height $\triangle z$ for an equatorward wind U is

$$\triangle z \approx \frac{H}{D} U \sin I \cos I. \tag{23}$$

The altitude of the ionosphere peak can also be drifted by electric field $\mathbf{E}$ at places where magnetic field is not vertical. The drift velocity is $\mathbf{E} \times \mathbf{B}/B^2$, and its vertical component is upward for eastward $\mathbf{E}$ and downward for westward $\mathbf{E}$. This electromagnetic drift is most effective at equatorial latitudes, where the magnetic field is horizontal and $\mathbf{E} \times \mathbf{B}$ drift is totally vertical. The $\mathbf{E} \times \mathbf{B}$ drift is also effective at high latitudes where magnetic field is vertical and hence plasma drifts horizontally.

4.8.1 F_3 Layer

An additional stratification, called F_3 layer, can occur in the equatorial F region during daytime (Balan et al. 1998). As daytime F_2 layer drifts upward mainly due to upward $\mathbf{E}\times\mathbf{B}$ drift and partly due to equatorward neutral wind (and other dynamical sources such as tides and waves), a new layer develops at lower altitudes through the usual photochemical and dynamical processes of the equatorial F region. As time progresses, the original F_2 layer drifts upward and forms F_3 layer while the new layer develops into usual F_2 layer. Both layers can be detected by bottom-side ionosondes for sometime when the density of F_3 layer remains greater than that of F_2 layer. As time progresses further, both layers drift upward, and when the density of F_3 layer decreases below that of F_2 layer, the F_3 layer can be observed as topside ledges. Strong F_3 layer develops and rapidly drifts to the topside during strong daytime eastward electric field as during prompt penetration of electric field (Fig. 5).

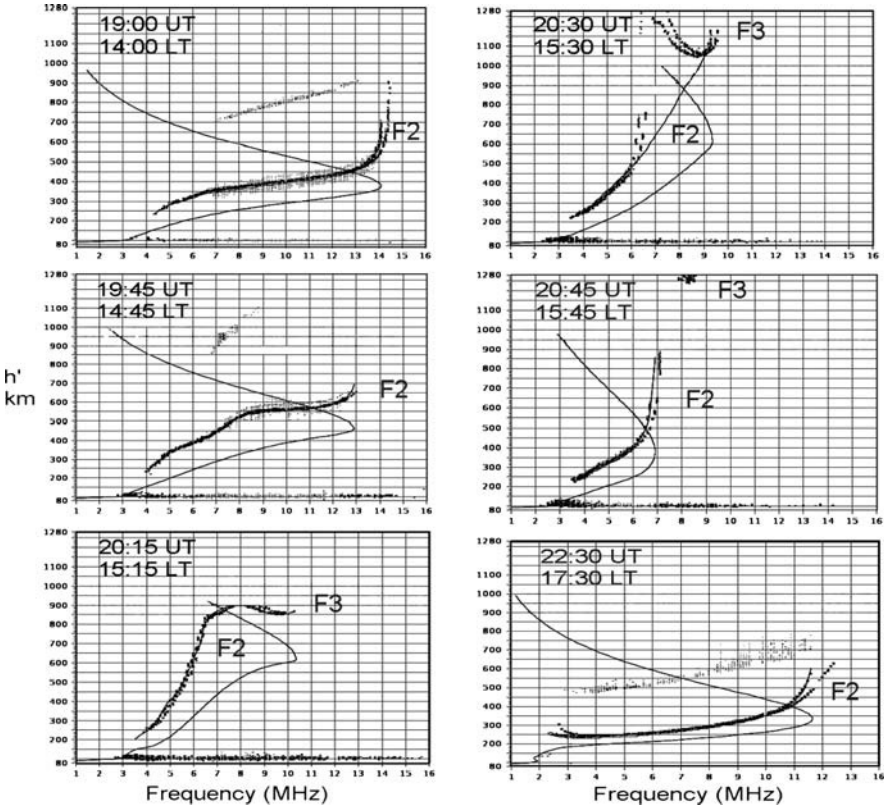


Fig. 5 Ionograms showing the development of F_3 layer at Jicamarca (11.9°S, 76.8°W; dip latitude 1°N) during eastward prompt penetration electric field on 09 November 2004

5 Ionospheric Electric Fields and Currents

When ions and electrons respond differently to the forces acting on them, there arises an electric current; this current diverges if there exists a spatial variation in electrical conductivity. Then a (polarization) electric field builds up quickly (within microseconds in the ionosphere) to modify the ion and electron velocities so that current divergence becomes zero. These physical process together with an ionospheric dynamo can explain the existence of electric fields and currents in the ionosphere.

5.1 *Dynamo Theory*

When air (due to tides, waves and winds) tries to move ionospheric plasma across geomagnetic field, there arises an electric field $\mathbf{U} \times \mathbf{B}$ (Lorentz field), which drives currents at levels in the ionosphere (mainly E region) where electrical conductivity is appreciable. The current density $\mathbf{j}$ is given by $\mathbf{j} = \sigma \cdot (\mathbf{U} \times \mathbf{B})$ where σ is conductivity. However, due to the vertical and horizontal variations of σ, currents cannot flow freely in all directions. That causes accumulation of electric charge, and a polarization electric field $\mathbf{E} = - \bigtriangledown \phi$ builds up, ϕ being electric potential. The total electric field $\mathbf{E}' = \mathbf{E} + \mathbf{U} \times \mathbf{B}$. The polarization field adjusts itself until the current flow becomes non-divergent and horizontal. The current density $\mathbf{j}'$ due to the total field is $\mathbf{j}' = \sigma' \cdot \mathbf{E}' = \sigma' \cdot (\mathbf{E} + \mathbf{U} \times \mathbf{B})$. As described below, the dynamo action is effective at E region heights because at these heights electrons and ions respond differently to neutral wind. At F region altitudes, the plasma motion is primarily along geomagnetic field lines.

5.1.1 Ion and Electron Velocities

The ion and electron velocities [from momentum (2)] are

$$\mathbf{V}_i = \mathbf{U} + \frac{1}{1+k_i^2}\left[\frac{k_i}{B}\mathbf{E}' + \left(\frac{k_i}{B}\right)^2 \mathbf{E}' \times \mathbf{B} + \left(\frac{k_i}{B}\right)^3 (\mathbf{E}'.\mathbf{B})\mathbf{B}\right], \tag{24}$$

$$\mathbf{V}_e = \mathbf{U} + \frac{1}{1+k_e^2}\left[-\frac{k_e}{B}\mathbf{E}' + \left(\frac{k_e}{B}\right)^2 \mathbf{E}' \times \mathbf{B} - \left(\frac{k_e}{B}\right)^3 (\mathbf{E}'.\mathbf{B})\mathbf{B}\right], \tag{25}$$

where $\mathbf{E}' = \mathbf{E} + \mathbf{U} \times \mathbf{B}$, and $k_i = \frac{\Omega_i}{\nu_{in}}$ and $k_e = \frac{\Omega_e}{\nu_{en}}$ are ion and electron mobility parameters, with $\Omega_i = eB/m_i$ and $\Omega_e = eB/m_e$ being ion and electron gyro frequencies; $\nu_{in} = 2.6 \times 10^{-15}(n_n + n_i)(M_n')^{1/2}$ and $\nu_{en} = 5.4 \times 10^{-16} n_n T_e^{1/2}$ are ion-neutral and electron-neutral collision frequencies, with M_n' denoting mean molecular mass, and densities in m^{-3}.

Altitude variations of collision frequencies (ν_{in} and ν_{en}), gyro frequencies (Ω_i and Ω_e), mobility parameters (k_i and k_e), and conductivities σ_p, $\sigma_\parallel$ and σ_H are shown in Fig. 6. As shown, $k_i \ll 1$ at altitudes below about 100 km, and therefore $\mathbf{V_i} \approx \mathbf{U}$ [from (24)]. Hence ion velocity in the bottom part of E region is determined purely by neutral wind (ions being closely coupled to neutral gas). At altitudes between about 100 and 115 km, k_i is still much less than 1, and so $\mathbf{V}_i \approx \mathbf{U} + k_i \frac{\mathbf{E}}{B}$ where we have used the fact that $(k_i/B) \;\mid \mathbf{U} \times \mathbf{B} \mid \;\ll\; \mid \mathbf{U} \mid$. Thus, the relative ion-neutral velocity in the middle part of E region (100–115 km) is parallel to the electric field $\mathbf{E}$.

At high altitudes (above about 180, F region), $k_i \gg 1$, and we can consider the relative ion-neutral velocity parallel to $\mathbf{B}$ and perpendicular to $\mathbf{B}$. The parallel velocity

$$(\mathbf{V}_i)_\parallel = (\mathbf{U})_\parallel + \left(\frac{k_i}{B}\right)\mathbf{E}_\parallel \tag{26}$$

becomes proportional to the $\mathbf{E}$ field component along $\mathbf{B}$, and perpendicular velocity becomes

$$(\mathbf{V}_i)_\perp = \mathbf{U} + \left(\frac{\mathbf{E}' \times \mathbf{B}}{B^2}\right)_\perp = \frac{\mathbf{E} \times \mathbf{B}}{B^2}. \tag{27}$$

For electrons, $k_e \gg 1$ at all altitudes, and therefore

$$(\mathbf{V}_e)_\perp = \frac{\mathbf{E} \times \mathbf{B}}{B^2}. \tag{28}$$

Hence the ion and electron velocities in the F region (27 and 28) are along the $\mathbf{E} \times \mathbf{B}$ direction and independent of neutral wind.

5.2 Ionospheric Conductivities

The current density $\mathbf{j}$ at a given height in the ionosphere is given by $\mathbf{j} = ne(\mathbf{V}_i - \mathbf{V}_e)$ in which singly charged ions and charge neutrality are assumed. Inserting the values of $\mathbf{V}_i$ and $\mathbf{V}_e$, and considering $\mathbf{E} = \mathbf{E}_\parallel + \mathbf{E}_\perp$, the current density becomes

$$\mathbf{j} = \sigma_0 \mathbf{E}_\parallel + \sigma_1 \mathbf{E}_\perp - \sigma_2 \frac{\mathbf{E} \times \mathbf{B}}{B}. \tag{29}$$

$$\sigma_0 = \frac{ne}{B}(k_e + k_i), \tag{30}$$

$$\sigma_1 = \frac{ne}{B}\left(\frac{k_e}{1 + k_e^2} + \frac{k_i}{1 + k_i^2}\right), \tag{31}$$

$$\sigma_2 = \frac{ne}{B}\left(\frac{k_e^2}{1 + k_e^2} - \frac{k_i^2}{1 + k_i^2}\right) \tag{32}$$

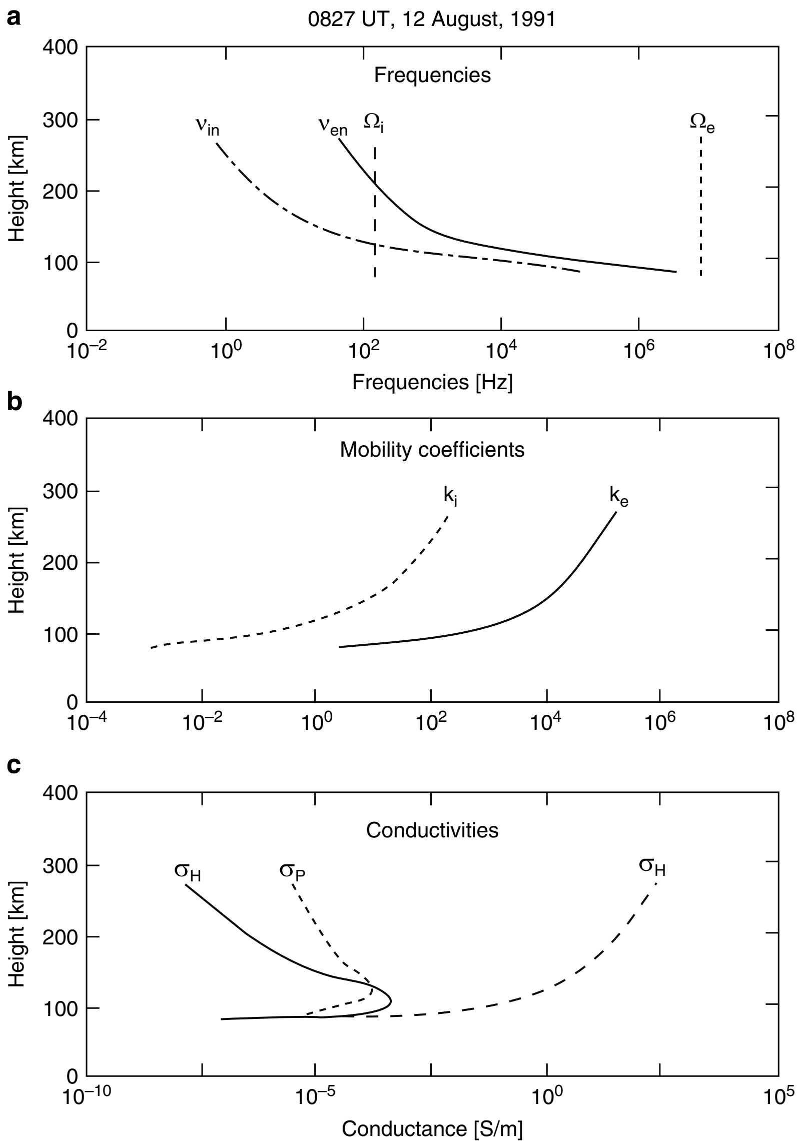


Fig. 6 Typical altitude profiles of the ion-neutral (ν_{in}) and electron-neutral (ν_{en}) collision frequencies, and ion and electron gyrofrequencies (Ω_i and Ω_e) (a), ion and electron mobility parameters (k_i and k_e) (b), and Pederson, Hall and parallel conductivities (σ_p. σ_H and $\sigma_{\|}$) (c) for an auroral location such as Tromso (69°, 19°E) (Brekke 1997)

are the parallel or direct conductivity (σ_0), transverse or Pedersen conductivity (σ_1) and Hall conductivity (σ_2). The parallel conductivity σ_0 corresponds to the conductivity when **E** is parallel to **B**. When **E** and **B** are perpendicular, the conductivity parallel to **E** (but perpendicular to **B**) is the Pedersen conductivity σ_1, and that perpendicular to both **E** and **B** is Hall conductivity σ_2. If the Hall current is inhibited, then the Pederson current (in the direction of the electric field) will be enhanced due to the polarization effect in the Hall direction. The Cowling conductivity (σ_3) then corresponds to the enhanced current in the direction of the electric field. The narrow current bands of the equatorial electrojet and those of the two auroral electrojets owe their existence to the Cowling conductivity acting in these regions.

5.2.1 Spatial Variation of Conductivity and Electrojets

Considering the cartesian coordinate system (x, y, z), with x toward magnetic north, y toward east and z toward Earth's center, the electric field, in general, has components in all directions and magnetic field has components in the x and z directions so that

$$\mathbf{E} = E_x\mathbf{x} + E_y\mathbf{y} + E_z\mathbf{z}, \tag{33}$$

$$\mathbf{B} = B\left(\cos I\mathbf{x} + \sin I\mathbf{z}\right), \tag{34}$$

where I is the dip angle. The current density has components in all directions and can be simplified to the tensor form

$$\begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix} = \begin{pmatrix} \sigma_1 \sin^2 I + \sigma_0 \cos^2 I & -\sigma_2 \sin I & (\sigma_0 - \sigma_1) \sin I \cos I \\ \sigma_2 \sin I & \sigma_1 & -\sigma_2 \cos I \\ (\sigma_0 - \sigma_1) \sin I \cos I & \sigma_2 \cos I & \sigma_1 \cos^2 I + \sigma_0 \sin^2 I \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}. \tag{35}$$

At high latitudes where I $\approx$ 90° and **B** is in the z direction, the above equation reduces to

$$\begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix} = \begin{pmatrix} \sigma_1 & -\sigma_2 & 0 \\ \sigma_2 & \sigma_1 & 0 \\ 0 & 0 & \sigma_0 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}. \tag{36}$$

Now suppose the northward current is inhibited ($j_x = 0$), then $E_x = \frac{\sigma_2}{\sigma_1} E_y$ and current in the east-west direction (j_y), which is auroral electrojet current, becomes

$$j_y = \left(\sigma_1 + \frac{\sigma_2^2}{\sigma_0}\right) E_y = \sigma_A E_y. \tag{37}$$

$\sigma_A = \left(\sigma_1 + \frac{\sigma_2^2}{\sigma_0}\right)$ is the auroral electrojet conductivity.

In equatorial region, where I is zero, the current tensor (35) reduces to

$$\begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix} = \begin{pmatrix} \sigma_0 & 0 & 0 \\ 0 & \sigma_1 & -\sigma_2 \\ 0 & \sigma_2 & \sigma_1 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}. \tag{38}$$

Now if the vertical current j_z is inhibited, then the above equation gives $E_z = -\frac{\sigma_2}{\sigma_1} E_y$ and the east-west current j_y (equatorial electrojet current) becomes

$$j_y = \left(\sigma_1 + \frac{\sigma_2^2}{\sigma_1}\right) E_y = \sigma_3 E_y. \tag{39}$$

$\sigma_3 = \sigma_1 + \frac{\sigma_2^2}{\sigma_1}$ is known as Cowling conductivity. Hence, the auroral electrojet and equatorial electrojet result from the polarization enforced by the restrictions $j_x = 0$ and $j_z = 0$.

6 Ionospheric Variations and Irregularities

As understood from theory, ionospheric density should vary with time of the day, day of year, season of year and phase of solar cycle. These variations are referred to as diurnal, day-to-day, seasonal and solar cycle variations. In addition, following geomagnetic storms, the ionospheric density increase/decrease very much from its average level, which are known as positive/negative ionospheric storms. The studies of geomagnetic and ionospheric storms are extremely valuable for *space weather* applications. The specific space weather effects include ionospheric density changes that affect GPS applications (communication and navigation), ionospheric currents that affect power supply and gas supply systems, atmospheric heating that affects satellite orbits, and heat and energy transfer that affect system dynamics.

Dense patches of ionization often occurring at E region heights (100–130 km), which is not related to the normal E region, is called *sporadic E* or E_s as it occurs sporadically. Sometimes, E_s appears in sheets which hide the overlying F layer; it may also be patchy and partially transparent to radio waves. The horizontal extend of E_s varies from tens to hundreds of kilometers, it drifts with velocities of the order of 50 m s^{-1} and has lifetimes of tens of minutes to several hours. For a theory of E_s see Whitehead (1989).

The most interesting physics in the equatorial F region takes place during post-sunset hours when F layer is temporarily raised to high altitudes where collisions are low, E region density is reduced rapidly by chemical recombination, and bottomside F region density gradient is increased by loss of photoionization. In short, the post-sunset F layer is in a state of delicate equilibrium; it is like lifting and lowering of a system consisting of a heavy fluid resting on top of a light fluid. Any disturbance cause by background noise, neutral winds, gravity waves, or some other source can disturb the equilibrium and generate instabilities (or irregularities) in the plasma. If conditions (especially high altitude of the ionosphere) are favorable, the irregularities will grow and manifest as *spread F and plasma bubbles* (e.g., Kelley 1989), which can disturb communication. The need for understanding the characteristics and origin of these irregularities, which cover scale sizes from centimeter to hundreds of kilometers, has increased as demands on communication and navigation have grown through satellite systems such as GPS network.

References

Atreya, S.K., Waite, J.H., Donahue, T.M., Nagy, A.F., McConnell, J.C.: Theory, measurements and models of the upper atmosphere and ionosphere of Saturn. In: Gehrels, T., Mathews, M.S. (eds.) Saturn, University of Arizona Press, Tucson, AZ, pp. 239–280 (1984)

Bailey, G.J., Balan, N.: A low-latitude ionosphere-plasmasphere model. In: Schunk, R.W. (eds.) STEP handbook, Utah State Univ. Logan, pp. 173–206 (1996)

Balan, N., Batista, I.S., Abdu, M.A., Macdougall, J., Bailey, G.J.: Physical mechanism and statistics of occurrence of an additional layer in the equatorial ionosphere. J. Geophys. Res. **103**, 29169 (1998)

Balan, N., Thampi, S.V., Lynn, K., Otsuka, Y., Alleyne, H., Fejer, B.G., Abdu, M.A.: F3 layer during penetration electric field. J. Geophy. Res. (2008). doi:10.1029/2008JA013206

Brekke, A.: Physics of the upper polar atmosphere. Wiley, New York (1997)

Kelley, M.C.: The Earth's ionosphere, plasma physics and electrodynamics. Academic, New York (1989)

Knudsen, W.C., Kliore, A.J., Whitten, R.C.: Solar cycle changes in the ionization source of the nightside Venus ionosphere. J. Geophys. Res. **92**, 13391 (1987)

Mahajan, K.K., Kar, J.: Planetary ionospheres. Space Sci. Rev. **47**, 303 (1988)

Ratcliffe, J.A.: An introduction to the ionosphere and magnetosphere. Cambridge University Press (1972)

Rishbeth, H., Garriott, O.K.: Introduction to ionospheric physics. Academic, New York (1969)

Schunk, R.W., Sojka, J.J.: Ionosphere-thermosphere space weather issues. J. Atmos. Terr. Phys. **58**, 1527 (1996)

Viggiano, A.A., Arnold, F.: Ion chemistry and composition of the atmosphere. In: Volland (ed.) Vol 1, (1995)

Whitehead, J.D.: Recent work on mid-latitude and equatorial sporadic-E. J. Atmos. terr. Phys. **51**, 401 (1989)

Yonezawa, T.: A new theory of formation of the F2 layer. J. Radio Res. Labs. **3**, 1 (1956)

Planetary Magnetospheres

Thomas Earle Moore

Abstract This is a brief, equation-free introduction to the physics of planetary magnetospheres, emphasizing their relationship to the heliosphere at large, over the life of our solar system. It will focus on unifying principles, rather than providing a taxonomy of the many different magnetospheres in our own solar system. A planetary *magnetosphere* is a closed cell of plasma bound together by a permeating magnetic field, which may originate from any combination of internal or external currents. The essential features are thus an ionized atmosphere (*ionosphere* or *plasmasphere*), created by intrinsic and/or extrinsic sources of energy, and a magnetic field. At planetary scales, parcels of plasma sharing a common magnetic flux tube must move together as a unit distributed along that flux tube. The magnetic field thus plays the role of a connective tissue threading plasmas and binding them together via *Maxwell stresses* analogous to those of surface tension, but distributed throughout the volume. More powerfully than surface tension acts to confine water in droplets, magnetic fields confine plasmas in magnetic cells, or magnetospheres. Rotation and relative motion are important factors in the character of magnetospheres, while magnetic linkages within and between them exert strong control over their interactions via the process known as *reconnexion*, which acts as a plasma pump and energizer.

1 Past

Planetary magnetospheres come into existence with the planets and other bodies as they form. Prior to stellar ignition, the accreting material is thought to form a disk containing a mixture of states of matter, spinning somewhat like a gigantic cement mixer, but with the most rapid rotation at the center, according to *Keplerian* motion, but modified by gas (or plasma) pressure gradients. The disk transports material

T.E. Moore
NASA's Goddard Space Flight Center
e-mail: thomas.e.moore@nasa.gov

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_13,

inward to form the central star, while transporting angular momentum outward, and dissipating energy in the flow shear between the differentially orbiting inner and outer portions of the disk. As the star ignites, it irradiates the stellar system with ionizing photons, creating plasmas within a few optical depths. It also develops a magnetic *dynamo*, coupling itself to ionized material it has created throughout the translucent cloud, and enhancing transport of angular momentum outward. The ignition also launches a mass outflow, mainly at high latitudes outside the massive and inward migrating proto-planetary disk near the equator. As planets form within that disk, they may produce their own magnetic *dynamos* where conditions are appropriate. Plasma pressure and Maxwell stresses modify the disk dynamics significantly, in ways that are not fully understood.

With gravitational *accretion* of matter and release of energy and radiation, materials are vaporized and ionized until the solar wind outflow works its way into the equatorial disk regions, driving the less dense gas and plasma from the system, and leaving behind the greater condensations of matter that become planets and other bodies. In our solar system, magnetic coupling of the Sun's rapid rotation throughout the partially ionized *accretion* disk, as illustrated in Fig. 1, is thought to account for the excess of angular momentum in the outer solar system. This coupling may also have spun up ionized material beyond its proper *Keplerian* orbital speed, causing much of it to flow away from the Sun and become entrained in the solar wind, while braking the spin of the Sun in its polar regions and contributing to the differential rotation that has roiled the solar magnetic dynamo ever since.

The effect of solar radiation and magnetized plasma expansion was to transport the more volatile components out of the inner solar system, evolving them from ices by heating them, and then blowing them away, just as we see happening to comets that stray into the inner solar system. The solar wind has been much diminished by now, but in the its T Tauri phase, it was powerful enough to open up the spaces among the planets by removing all but the largest and most dense objects remaining. To this day, the solar wind does its best to ablate volatile materials from everything in the solar system.

2 Present

As the fourth state of matter, in which electrons are freed from atoms, plasmas are electrically conductive. They are also distinguished, especially in space, by their low densities and lack of particle collisions. This makes it tempting to consider treating them as a sum over individual charged particle motions, but plasma charges and currents determine the electric and magnetic fields, so these must be consistent. There are also turbulent electromagnetic fields that lead to randomization and the need for statistical techniques.

Classical fluid (conservation) equations, augmented with electromagnetism and known collectively as magnetohydrodynamics (MHD), provide an effective numerical method of simulating the behavior of planetary-scale magnetized plasmas. These

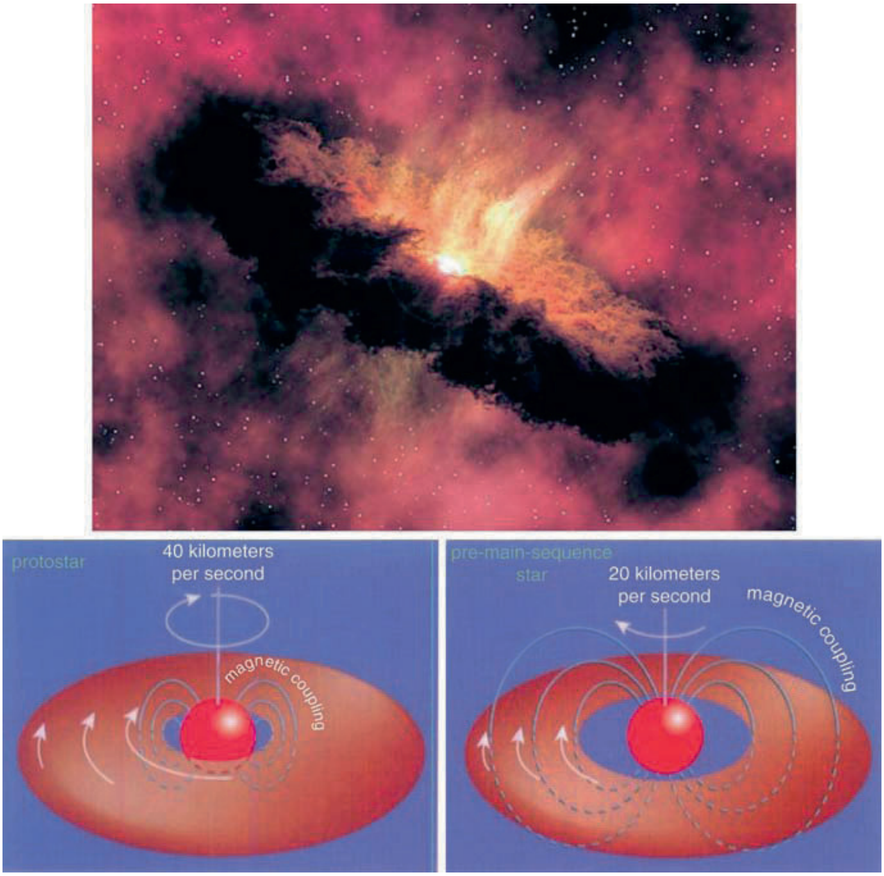


Fig. 1 *Upper panel*: Artist conception of a protoplanetary accretion disk after stellar ignition but before planet formation [Robert Hurt/NASA JPL-Caltech]. *Lower panel*: Schematic of an *accretion* disk coupling with magnetized protostar, enhancing outward transport of angular momentum [courtesy of Joe Nuth/NASA/Goddard]

equations describe plasma flows and resultant magnetic fields, with implicit electric and current flow fields, which are responsible for enforcing the common plasma motions along the magnetic field, known as the "*frozen-in flux*" condition. The basic idea is given in the inset:

> Flows of conducting plasmas are driven by pressure gradients and Maxwell stresses (field tension and pressure), created by energy stored or released within, or flowing through a plasma system. Plasma motions require a consistent electric field, and currents flow to spread the electric field along magnetic flux tubes so that connected plasmas are required to share the same motion. Reconnexion is the opening or closing of magnetic connexions between plasma cells. [UK spelling is used here in recognition of many UK contributions to the concept, especially as applied to magnetospheres].

In the past 25 years, large-scale three-dimensional MHD simulations have progressed from their infancy to their current status as realistic global circulation models for space weather, that successfully describe many of the observed features of magnetospheres. Such models can be set up to compute the interactions between multiple fluids, mainly ionospheric plasmas interacting with the solar wind plasmas. MHD simulations cannot treat smaller spatial scales, but hybrid simulations (kinetic ions, fluid electrons) do capture ion *gyro*-scale structures, which are especially prominent in smaller magnetospheres, such as those of comets or smaller moons and asteroids. Electron-scale features, such as the very thinnest *current sheets*, must still be described by fully kinetic particle in cell simulations, which must necessarily have local, rather than global scope.

2.1 Magnetization

Magnetospheres are usually nested, or embedded within larger magnetospheres. For example, the terrestrial magnetosphere is embedded within the heliosphere, or magnetosphere of the Sun. The heliosphere is in turn embedded within the magnetosphere of our galaxy, which in turn is embedded within the field and plasma of intergalactic space, and so on.

A typical magnetosphere forms when magnetized plasma flows past a magnetized object. Some of the solar planets and moons have molten metallic cores that churn with overturning motions as they seek to lose the heat of their formation, of radioactive decay of their constituents, or of tidal deformations when they orbit a larger body at close distance. Strong convective overturning of conducting fluids often develops into a magnetic dynamo, an example of which is the Earth's core. Rotation is also a significant factor in dynamo development, with the most rapid rotators having the most powerful dynamos, such as Earth, Jupiter, Saturn, Neptune, and Uranus. Despite its slow rotation, Mercury has a small intrinsic field, while Mars has lost its dynamo. Venus, which actually is in an extremely slow retrograde rotation, has no sensible dynamo at all.

Smaller, less conductive, and slowly spinning objects may not excite a magnetodynamo, but a conducting object with no intrinsic magnetization and embedded within a strong flow of magnetized plasma forms what is called an "induced" magnetosphere, as shown in Fig. 2. The object may be conducting at multiple levels, because of surface or core conductivity, or because it is evolving gas that is being ionized by exposure to an ultraviolet radiation field. As always, magnetic fields couple together all conductors linked by magnetic flux tubes. The relative flow induces currents that deform the magnetic field and generate Maxwell stresses that decelerate the fast medium and accelerate the slow medium. The magnetic field linking the two regions becomes draped over the slower "obstacle" and is stretched out as the flow proceeds. Equilibrium is reached when the deceleration of the fast medium is matched by the acceleration of a tail of material extending downstream of the emitting source. A current sheet separates the two ends of the draped magnetic field

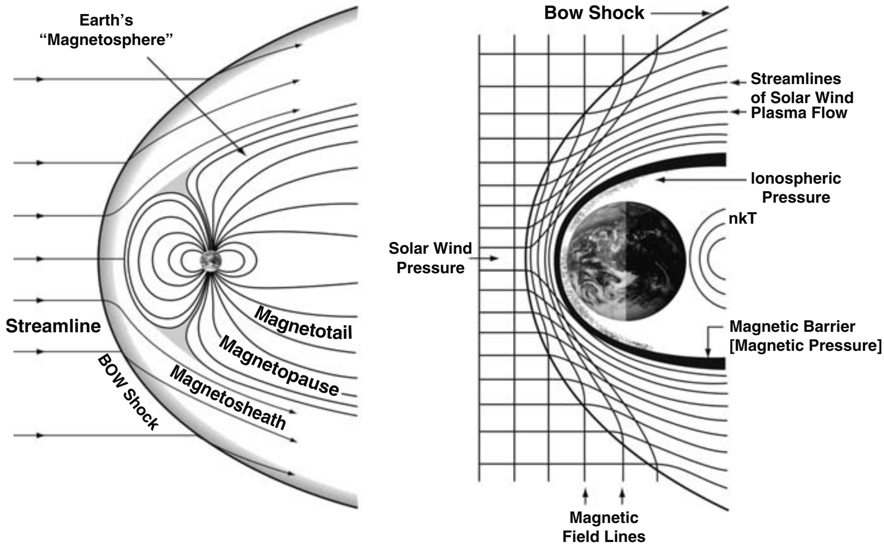


Fig. 2 Comparison of an *intrinsic* and an *induced magnetosphere* [courtesy of Janet Luhmann, UCB]

lines, and is subject to both internal instabilities and dynamics owing to perturbations in the surrounding solar wind. Reconnexion across the current sheet may occur, creating a closed plasma cell (*plasmoid*) around the body, as described further below.

2.2 Magnetosphere in 3D

In 3D, a magnetosphere resembles a blunt object in the solar wind, the obstacle being the energy density (pressure) of its intrinsic magnetic field, which deflects the solar wind around it. Since the solar wind is generally super-*Alfvénic* (and also faster than the fastest magnetohydrodynamic wave mode), a *bow shock* must decelerate and compresses the solar wind before it can be deflected around Earth, in a region called the *magnetosheath*, bounded by the *magnetopause* on the interior. Just inside the *magnetopause* is a layer of slowed solar wind known as the *mantle*. Polar magnetic flux tubes are swept back to form the north and south *lobes*, while plasma pressure trapped in the *plasma sheet* inflates closed flux tubes sandwiched between them. Within the more nearly dipolar inner magnetosphere are the *radiation belts* and *plasmasphere*, the latter being an extension of the ionosphere and atmosphere into a donut-shaped region encircling Earth. The radiation belts co-exist with this *cold plasma*. It can readily be seen from Fig. 3 that cells of magnetized plasma soon acquire nearly as much structure as living cells.

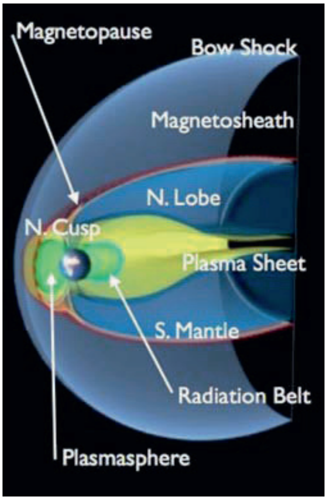


Fig. 3 A 3D visualization of the terrestrial magnetosphere illustrates the principal regions and structures and their relationships (courtesy of UCAR Comet / MetEd program http://www.meted.ucar.edu)

2.3 Reconnexion

When two magnetized plasma cells interact, the relative orientation of their magnetic fields determines the details of the interaction. This involves reconnexion of the magnetic fields, opening loops of closed flux within each cell to form linked flux tubes that connect and couple the cells (or vice versa). The cohesive, connective tissue of magnetic flux tubes is then extended between the cells (or pinched off), acting to accelerate each of the cells toward the velocity of the other, producing a merger (or a separation). When two magnetic cells approach, the shear between their fields causes them to connect together or "merge" by forming new flux tubes that connect the two cells, along which the two plasmas vie for influence as they mix and are pumped together into the newly connected region. The magnetic forces draw plasma flows from the two intake regions through the X line region, to the two exhaust regions. The result can be flows that are very different from those of normal gas dynamics of flow around a blunt object, as illustrated here in Fig. 4.

In the Earth's magnetotail, slowed solar wind and accelerated ionospheric plasmas form errant plasma cells in the tail of the magnetosphere, moving too fast to be restrained by the magnetic tension force. These episodically pinch off and escape from the main cell, forming new cells of mixed plasma called "plasmoids," which escape downstream in the solar wind. The result is a continual *ablation* of the plasma from the small cell. Conversely, an appreciable amount of solar wind is slowed down and incorporated into the magnetosphere, with excess energy being either *thermalized* or transferred to the ionospheric plasmas, leading to increased rate of ionospheric loss downstream.

In summary, reconnexion acts like a pump, moving plasma to/from the separate cells as they are disconnected/connected by reconnexion. The pump is powered by Maxwell stresses of the sheared magnetic fields, which are relieved in the process of

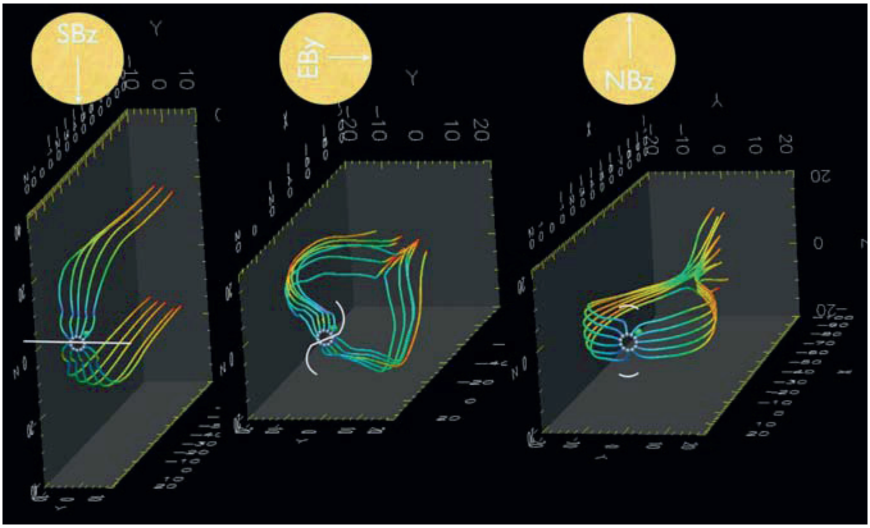


Fig. 4 Distortion of boundary layer streamlines from aerodynamic cylindrical symmetry by dayside reconnexion for south, east and north interplanetary magnetic field (IMF), as simulated in ideal MHD. Inferred X lines are indicated by white curves

the pumping action, generating fast flows that approach the Alfvén speed, a measure of the acceleration of plasma density by the Maxwell stresses.

2.4 Turbulence and Acceleration

Reconnexion relieves shear stresses in the magnetic field and energy is released into the plasmas as they are accelerated by those stresses, as part of the pumping action. This action is often intense, sporadic and localized, such that it generates turbulent flows of both matter and electrical current. The result of *turbulence* is a cascade of energy from larger (MHD) scales to smaller scales as organized shear layers break up into disorganized, tumultuous eddies and vortices. As scale sizes decrease, the effects can involve kinetic scales of the ion or electron, gyro-motion. A variety of plasma wave-particle interactions then serve to dissipate the motions into thermal energy.

There is also a hypothesis, supported by laboratory measurements and space observations of reconnexion, which states that turbulent dissipation reverses the normal cascade to smaller scales in a reconnexion layer. Plasma turbulence is thought to impede the flow of current in a shear layer, and this enhances the local electric field near the X-line, thereby increasing the rate of reconnexion (magnetic flux transfer per unit time), so that microscopic turbulent features thereby affect the macroscopic field topology.

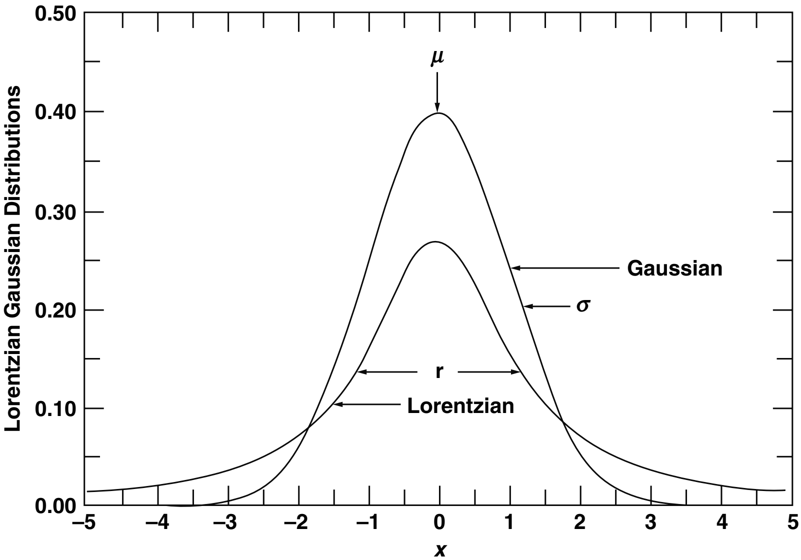


Fig. 5 Comparison of normalized Lorentzian (or kappa) and Gaussian (or Maxwellian) distributions, with full width at half measure FWHM=2.354 sigma

Magnetized space plasmas tend to acquire high velocity particles, with a subset gaining energies well in excess of the mean or thermal energy. The acceleration of such particles often produces a power law "tail" extending to high energy from the thermal core, whose mathematical description is given by the Lorentzian or "kappa" function, named for a parameter that describes the power law slope and prominence compared with the thermal core, as shown in Fig. 5. Depending on the effectiveness of the acceleration and the sensitivity of a particular detector, particles are observed with energies extending up into the relativistic range of energies.

In planetary magnetospheres, populations of electrons and ions with non-thermal energies are a persistent feature (Van Allen radiation belts) that ebbs and swells in prominence, and can do serious damage to the best-protected circuits or physiologies. The mechanisms that generate such particles are perplexingly distinct from those that control the thermal plasma distributions. The temperature range in the ionosphere begins at 1,000 K while in the solar wind it begins at about 1,000,000 K. But after processing by the magnetosphere, the storm plasmas reach thermal energies in the neighborhood of 1,000,000,000 K. These populations are important as seed populations for the non-thermal radiation belts, which appear as a tail with characteristic energies extending to 1–10 MeV. Space storm plasmas increase with unusually high solar wind density, whereas radiation belts increase with unusually high solar wind velocities, possibly implicating turbulence as an agent.

2.5 Rotation and Circulation

Rotation is a critically important factor in the behavior of magnetospheres, because of the electromagnetic coupling of plasmas threaded by magnetic fields. A magnetized plasma that rotates must do so as a solid body, or its magnetic field must undergo stretching and shear that will eventually produce reconnexion to relieve the excessive magnetic stresses. In the context of our solar system, the premier case is that of the solar magnetic field, which enforces solid body rotation close to the solar equator. Beyond that, the solar magnetic field becomes tightly wound up as the sun fails to enforce co-rotation with increasing distance (and therefore at increasing latitude of connection with the solar atmosphere. The resultant shears lead to continuing or episodic reconnexion, both in the solar atmosphere, where the rotation rate difference between poles and equator produces solar active regions, and in the solar wind where the equatorial current sheet is often subject to reconnexion. The solar *coronal* plasma is flung outward once it is no longer constrained by the solar magnetic field.

Planetary magnetospheres behave in an analogous fashion, if they rotate sufficiently fast. Jupiter's magnetosphere has both a strong magnetic field and rapid rotation. Moreover, Jupiter's magnetosphere is filled with plasma both from its ionosphere, and from planetary satellites. The latter small bodies, with their weak gravity, are copious sources of gas and plasma, as they are as well for Saturn and Uranus. Owing to their gravitation, only lighter species escape from the ionospheres of the gas giant planets.

Jupiter's internal plasma sources load its magnetosphere with plasmas that are flung outward by Jupiter's rotation, inflating its magnetic field until it can no longer restrain them. At the radius where the magnetic field can no longer enforce co-rotation, reconnexion leads to a more or less steady disconnection of the plasma to form plasma cells (plasmoids) that fly off downstream through Jupiter's magnetotail, eventually joining the solar wind, as illustrated in Fig. 6.

The Earth's ionosphere fills its magnetosphere with nearly co-rotating light ion plasma, to a radius of about 4–5 R_E. Beyond that, the plasma is more often influenced by solar wind driven boundary layer circulation, rather than by co-rotation. Dayside magnetic reconnexion removes plasma and magnetic flux from the dayside and transport it into the magnetotail, where flux tubes form a sunward flowing plasma sheet. The magnetotail also sheds plasma by ejecting plasmoids at times. Thus, the rotation of the central body's ionosphere competes with external plasma flow for control of the plasma within a magnetosphere. Mercury and Jupiter are examples of solar wind and planetary rotation-dominated magnetospheres, respectively, with Earth at midrange.

As the solar wind seeks to erode away the plasma and magnetic fields of Earth and carry them off downstream, the ionosphere responds to circulation of the inner magnetosphere by releasing plasmaspheric *cold* plasma to the dayside magnetopause region in the form of *plasma plumes*. Also, it responds to energy dissipated in the *auroral* zones by releasing O^+ plasma into the magnetosphere. Energy goes into heating the neutral gas as well as the topside ionospheric plasma. The situation

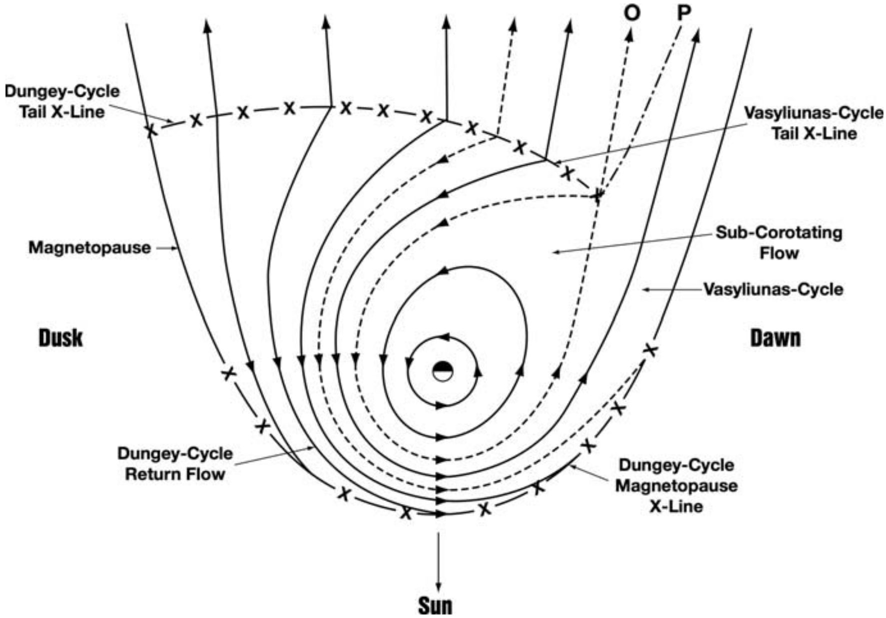


Fig. 6 Circulation (*Dungey cycle*) and rotation (*Vasyliunas cycle*) effects in Jupiter's magnetosphere. The indicated X-line leads to disconnection of the plasma above in the figure. If the X-line is transient, plasmoids are released down the Jovian magnetotail [courtesy of S W H Cowley/U Leicester]

is analogous to a water droplet suspended in a supersonic gas flow, for which the frictional interaction is so intense that it heats the droplet contents, turns them to vapor, and then carries them off downstream, with net ablation.

2.6 Magnetotail and Space Storms

The magnetotail forms through reconnexion between the solar and terrestrial magnetic fields, as illustrated in Fig. 4. As flux tubes with one end embedded in Earth and the other embedded in the solar wind are formed by low latitude dayside reconnexion, their lower segment curves away from the Sun so as to drag ionospheric plasmas in the same direction, while their higher segment curves toward the Sun so as to exert drag on the solar wind, leading to the "S" shape of long and drawn out magnetotail lobe flux tubes.

As lobe flux tubes are rejoined with solar wind flux tubes by reconnecting at high dayside latitudes, newly closed flux tubes are formed that are filled with solar wind plasmas and then expand rapidly along the flanks of the magnetosphere and into the magnetotail, supplying stretched flux tubes with solar plasma on them to the plasma sheet. The mix of high and low latitude reconnexion (therefore lobe and flank flux tubes) is controlled by the orientation of the IMF, with low latitude reconnexion (open lobe field) dominant for the southward IMF and high latitude reconnexion (closed flank field) dominant for the northward IMF.

The location of reconnexion in the magnetotail varies substantially, between distant tail reconnexion where the solar wind reaches the midplane, and episodic reconnexion closer to Earth, ejecting plasmoids. The latter allows the closed flux tubes in the earthward *exhaust* region to relax, or "dipolarize". Plasmas are further heated and compressed as they are transported from the plasma sheet into the inner magnetosphere. The energy dissipation that goes into magnetospheric plasma clouds raises the pressure sufficiently to inflate the magnetic field that confines them. In fact, the peak pressure contained inside the magnetosphere often exceeds the driving solar wind *dynamic pressure* by a factor of 10–20. Equivalently, these plasmas carry currents in the same sense as the Earth's magnetic dynamo, adding to the total dipole moment of the planet consistent with inflating the geomagnetic field. This is called the "ring current," though it is often an asymmetric or incomplete "ring". The ring current is actually a *magnetization* or edge current supported by the inward plasma pressure gradient at the outer boundary of the plasma cloud. It is not a bulk current carried by the differential drifts of ions and electrons, a serious but common initial misconception by students of this field of study (Fig. 7).

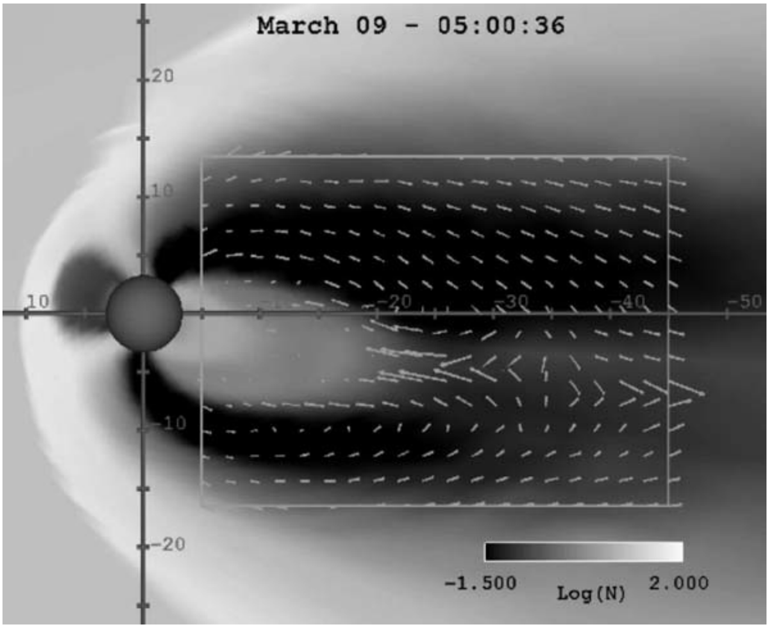


Fig. 7 MHD simulation of the magnetotail illustrating the plasma flows associated with reconnexion in the mid tail. Density N is plotted on the gray scale indicated, while the flow vectors indicate the global flows, in which the exhaust regions send plasma both earthward and tailward

The dominant plasma pressure in the inner magnetosphere is observed to be *thermal* ions in the energy range of 3–300 keV, including some super-thermal *hot tail* populations that shade into the radiation belts. Hot electrons carry at most 20% of the pressure or current. Beginning in the 70's, ion composition observations showed that the quiet *ring current* is dominated by H^+ (indeterminate origin) and He^+ (ionospheric), but acquires a substantial component of oxygen ions when it becomes strong. It also contains minor components of He^{++}, O^{++}, and even O^{5-6+}, the multiple charge states most likely originating mainly in the solar *corona* and wind. The O^+ ions have certainly come from the ionosphere, illustrating the point made above that solar and terrestrial plasmas mix when the solar wind interaction is very strong. The largest storm plasma events may consist almost entirely of ionospheric O^+ plasmas. Multiply charge ions that most likely come from the solar wind are also observed in the hot storm plasmas.

3 Future

The Sun is variable over the 11 or 22 year cycle of sunspot activity in its magneto-dynamo, accompanied by EUV variations that modulate planetary atmosphere temperature, scale heights, and ablation rates. The activity cycle is driven by differential rotation of the Sun. There are also longer-term variations of that cycle, such as the Maunder minimum of the late seventieth century. On an even longer scale, our geo-dynamo undergoes relatively abrupt polarity reversals on an irregular basis that averages five per Myr. These unfold over a few thousand years, and there are indications that we have been entering such a reversal during recent millennia. Assuming our planet is not impacted by a large bolide, we should survive as a species to witness the effects of the current reversal, if it is completed rather than aborted as some of them are. Though such reversals are difficult to reconstruct in detail from paleomagnetic records, the likely scenario, based on dynamo simulations, is that the geo-dipole moment will fall to approximately one tenth of its normal value as it rotates erratically to the opposite pole through the equator, and then intensifies again to its normal steady value, but with north magnetic pole toward the north geographic pole for the duration of this cycle (Fig. 8).

During such a reversal period, rotation will cause the magnetosphere to oscillate daily from one that is not so different from the present to one that is pole-onward to the solar wind, like that of Uranus during parts of its orbit. *Auroras* will then be seen encircling the magnetic poles, which will lie near the equator. Humankind will be treated to a very different sort of auroral light show than at present, with possible impacts upon the upper atmosphere, and other changes in exposure to solar energetic particles or even low energy cosmic rays. The tools are in hand to predict all aspects of our magnetosphere's interaction with the solar wind, though this has not yet been attempted for a reversal period.

Fig. 8 Photographer's concept of aurora in the tropics. Image courtesy of Trey Ratcliff, http://www.stuckincustoms.com

Acknowledgements I'm indebted to all of my mentors and collaborators, to whom go the credit for any insights expressed here. The blame for any mistakes lies squarely with the author.

Further Reading

Hultqvist, B., Oieroset, M., Treumann, R.A.: Magnetospheric plasma sources and losses. Kluwer, Dordrecht, Holland (1999)

Kivelson, M., Russell, C.T.: Introduction to Space Physics. Cambridge University Press, Cambridge, UK (1995)

Moore, T.E., Horwitz, J.L.: Stellar ablation of planetary atmospheres. Rev. Geophys. **45**, RG3002 (2007) doi:10.1029/2005RG000194.

Parks, G.K.: Physics of Space Plasmas, an introduction. Westview Press, Boulder, Colorado (2003)

Glossary

Ablation
Loss of surface material from a body through melting, sublimation, or evaporation into an unbound state by friction.

Accretion
Gain of material onto a body through recombination, condensation, liquefaction, or solidification.

Alfvén speed
The natural conducting fluid transverse mode wave speed defined by the fluid density and magnetic field tension.

Aurora
Excitation of atmospheric light emissions by super-thermal particles.

Bow shock
A strong fluid jump wave needed to slow the fluid flow relative to an obstacle to less than the fastest wave mode speed, so it can be deflected around the obstacle.

Cold plasma
Plasma with a temperature taken to be either irresolvably low or uninterestingly low in comparison with other higher temperatures of interest.

Corona
The rarified gaseous or plasma atmosphere around a celestial body, which is so hot as to be barely confined by the gravitational potential of the body.

Current sheet
A spatially-distributed electrical current flow that is much thinner than its extent in the other two dimensions, usually attributed to the demagnetization of charge carriers by the magnetic field perturbation produced by the current sheet itself.

Dynamic pressure
The energy density of a flow, approximately equal to the stagnation pressure that is produced when the flow runs into an obstacle.

Dungey cycle
The raindrop-like circulation flow produced in a magnetosphere by the relative flow of the solar wind, combined with reconnexion at the dayside that produces downstream flow along the boundary layers, accompanied by reconnexion in the magnetotail (or at high latitudes), producing sunward return flow to close the circulation cycle.

Dynamo
Any machine that converts mechanical energy of motion into electrical energy of current flow.

Exhaust
For reconnexion, the two quadrants of the region around a magnetic X line in which flow is pumped away from the X line by Maxwell stresses.

Frozen-in flux
The necessity for all conducting fluid parcels connected by a magnetic flux tube to share the same motion, thereby giving rise to identification of the magnetic flux with the plasma parcels themselves.

Gyro motion
The circular gyration of charged particles in a magnetic field.

Hot tail
A statistical population of particles with energies and abundances that lie above the normal curve of the thermal distribution for such particles.

Induced magnetosphere
One in which the tail like magnetic field originates with the external fluid through the drag exerted upon it by the internal conducting fluids.

Intrinsic magnetosphere
One in which the tail like magnetic field originates from motion of conducting fluids internal to the central object.

Intake
For reconnexion, the two quadrants of the region around a magnetic X line in which flow is supplied to the X line.

Ionosphere
Gaseous atmosphere that is partially ionized owing to the presence of an ionizing radiation field or other source of ionizing energy.

Magnetosphere
A closed cell of plasma bound together by a permeating magnetic field, which may originate from any combination of internal or external currents.

Magnetosheath
The region of flowing plasma between a bow shock and a magnetic obstacle.

Magnetopause
The boundary between the magnetic field of a magnetosphere and that of its external plasma environment. When connected, field lines pass through this boundary, making its definition somewhat vague in such regions.

Magnetization current
An effective electric current produced by a gradient in the magnetization density of a magnetized solid, or the pressure of a plasma.

Mantle
Layer of solar wind plasma inside the magnetopause owing to flow along flux tubes that pass through that boundary, but may have been closed by reconnexion.

Maxwell stress
The net force produced by currents in a magnetized fluid, given by the cross product of the current density with the local magnetic field vector. Consists of terms corresponding to tension along the magnetic field and pressure transverse to the magnetic field.

Plasma sheet
Similar to a current sheet, in the case where the current is carried by a gradient of the plasma pressure in the sheet.

Plasmasphere
The region in which ionospheric *cold* plasma density reaches equilibrium values owing to trapping of the plasma on closed flux tubes with closed circulation.

Plasma plume
A long cloud of plasma drawn from a reservoir into a region of normally lower density.

Plasmoid
A cell of plasma defined by closed magnetic flux tubes without an enclosed magnetized object, in which all currents flow within the plasma.

Radiation belt
A donut shaped region of enhanced energetic particle fluxes that area stably trapped in a nearly dipolar magnetic field region.

Ring current
A donut shaped region of enhanced plasma pressure that is temporarily trapped in a nearly dipolar magnetic field region.

Reconnexion
A change in the connection of magnetic flux between plasma cells or magnetospheres, through the operation of a plasma pump powered by the release of magnetic energy as sheared fields are relaxed.

Solar wind
Radially outward flow of the magnetized solar atmosphere through the solar system, usually at speeds that are faster than any wave mode speed.

T Tauri phase
An early stage of stellar development in which a very intense stellar wind flows with very large mass flux.

Thermal plasma
Plasma with a well-defined temperature or statistical average particle energy.

Turbulence
Disordered fluid motion, usually in regions of strong flow shear (gradients), with a cascade from macroscopic to microscopic scales of fluctuation.

Vasyliunas cycle
Rotation-driven plasma circulation flow in a dipolar magnetic field region, leading to reconnexion at the radius where the plasma can no longer be confined by the magnetic field, leading to release of the plasma by reconnexion in the magnetotail, in a series of plasmoids.

The Sun and Space Weather

Arnold Hanslmeier

Abstract In this chapter we will briefly review the basic interactions between particles and magnetic fields, the processes that occur on the Sun which are relevant for space weather as well as their influences on Earth and the space environment of Earth. The strong societal impact of space weather to our complex world of telecommunication will be stressed.

1 Definition of Space Weather and Some Examples

1.1 Definition

The US National Space Weather Programme gave the following definition of Space Weather:

Conditions on the Sun and in the solar wind, magnetosphere, ionosphere and thermosphere that can influence the performance and reliability of space-borne and ground-based technological systems and can endanger human life or health.

This definition can be extended to:

All influences on Earth and near-Earth space.

Space weather plays an important role in modern society since we strongly depend on communication which is mostly based on satellites and thus influenced by the propagation of signals throughout the atmosphere. Moreover, satellites themselves are vulnerable to space weather:

- Space Shuttle: numerous micrometeoroid/debris impacts have been reported.
- Ulysses: failed during the peak of Perseid meteoroid shower.

A. Hanslmeier
Institute of Physics, Department of Geophysics, Astrophysics and Meteorology, Univ.-Platz 5, A-8010 Graz, Austria
e-mail: arnold.hanslmeier@uni-graz.at

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_14,

- Pioneer Venus: several command memory anomalies related to high-energy cosmic rays.
- GPS satellites: photochemically deposited contamination on solar arrays.
- 1989 power failure in Quebec due to magnetic storms.
- On Earth: radio fadeouts. The HF communication depends on the reflection of signals in the upper Earth's atmosphere which is strongly influenced by the Sun's shortwave radiation.

An introduction to space weather with much more references was given by Hanslmeier (2007).

The short term effects are summarized as space weather, but also long term effects have to be taken into consideration: the solar luminosity has evolved from about 75% four billion years ago to its present value and will slightly increase in the future. The galactic environment changes due the rotation of the solar system around the galactic center. These long term effects are summarized as space climate.

The Sun, including interactions between solar magnetosphere and solar wind with the Earth's magnetosphere, is the main driver for space weather, but also other sources like space debris, asteroids, meteoroids, radiation and particles, cosmic rays, nearby supernova explosions and even the galactic environment are included in this definition.

1.2 Space Weather Customers

Modern communication systems, based on Earth or on satellites as well as power systems on Earth, and health of astronauts can be strongly influenced by space weather. Therefore, the aim of space weather research is to try to avoid negative consequences, mainly by efficient system design and warning.

Customers of space weather predictions are satellite designers and operators, manned space flight missions, telecommunication companies and power suppliers.

One of the main aims of solar research is to understand better the mechanisms that produce space weather relevant phenomena on the Sun (like flares or coronal mass ejections) and to predict the occurrence of energetic solar events that influence space weather.

There are many organizations throughout the world which are strongly devoted to space weather research, for example the US Space Weather Program, US-NASA's Living With a Star program, ESA's space weather program, SWENET, Space Weather European Network, SIDC, Solar influences data center at the Royal Observatory in Belgium, Lund space weather center, the Australian IPS Radio and Space Services , the Australian Space Weather Agency, and many others (such as the Space Research Group in Oulu, Finland).

2 Some MHD Basics

In this chapter we shortly review basic equations of plasma physics to understand the complex interactions that occur during eruptive processes on the Sun or when solar magnetic fields and particles interact with the geomagnetic field.

2.1 *The Lorentz Force*

Magnetic fields are generated by currents (flow of charged particles) and changing electric fields. Magnetic fields always have two poles. The Lorentz force is the force on a point charge q due to electromagnetic fields. Considering only a magnetic field B, the Lorentz force is given by:

$$\mathbf{F} = q(\mathbf{v} \times \mathbf{B}). \tag{1}$$

F in Newtons, q in Coulombs, **v** in m/s, **B** in Teslas.

In the general Lorentz equation the electric field **E** enters as

$$m\frac{d\mathbf{v}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{2}$$

The Lorentz force can be split into components of motion perpendicular and parallel to the magnetic field.

2.2 *Charged Particles and Fields*

Consider a magnetic field line and a charged particle. In case of equilibrium between centrifugal force and Lorentz force we can derive the radius of gyration, r:

$$\frac{mv^2}{r} = qvB \Rightarrow r = \frac{mv}{qB}. \tag{3}$$

Therefore, charged particles circulate around magnetic field lines and the sense of rotation is clockwise for negative particles. More generally, motion is circular about an imaginative guiding center. The gyrofrequency or Larmor frequency is given by:

$$\omega = \frac{qB}{m} \tag{4}$$

in rad s^{-1}.

Example: Radius of gyration in Earth's magnetosphere; consider an electron e^- with 100 keV kinetic energy, $E = mv^2/2$; its radius of gyration is 100 m.

Under the action of some external force, **F**, a drift of particles occurs, the drift velocity is given by:

$$\mathbf{v}_{\mathrm{df}} = \frac{\mathbf{F} \times \mathbf{B}}{q\mathbf{B}^2}. \tag{5}$$

As an example we consider the gradient of the magnetic field around a planet where the field strength decreases with increasing distance:

$$\mathbf{F} = -\mu_b \nabla B. \tag{6}$$

This force causes electrons and protons drift in opposite directions (electrons eastwards). Therefore, a ring current system develops which strengthens the Earth's magnetic field outside the current system.

Other examples are the field curvature drift, the gravitational field drift and the electrical field drift

$$\mathbf{F} = q\mathbf{E}, \tag{7}$$

where the drift velocity becomes

$$\mathbf{v}_E = \frac{c\mathbf{E} \times \mathbf{B}}{B^2}. \tag{8}$$

Particles move perpendicular to **E** and **B** and protons and electrons move in the same direction. The $\nabla\mathbf{B}$ and $\mathbf{E} \times \mathbf{B}$ drift dominate in the magnetosphere.

2.3 Magnetic Mirror

What happens if the field strength changes along the motion of a particle? The particles are bounced back from the strong field strength region. In the magnetosphere, electrons and ions will bounce back and forth between the stronger fields at the poles. B_{max} and B_{min} denote the maximum and minimum field strengths, and particles with a pitch angle greater than

$$(v_\perp / v_\parallel)_{\mathrm{crit}} = \sqrt{B_{\mathrm{max}}/B_{\mathrm{min}} - 1} \tag{9}$$

will be reflected!

Note also, that charged particles cannot cross magnetic field lines. Therefore, the magnetic field of Earth provides a shielding against charged particles (mainly coming from the Sun).

2.4 Earth's Magnetic Field

The geomagnetic field can be characterized as non-static with the surface field strength between 30 and 60 microteslas (at the poles). Close to the poles, the

Table 1 Summary of geomagnetic indices

aa	3-hr range index, derived from two antipodal stations
AE, AU, AL	1- , 2.5-min, or hourly auroral electrojet indices
am, an, as	3-hr range (mondial, northern, southern) indices
Ap	3-hr range planetary index derived from Kp
C, Ci, C9	Daily local (C) or international (Ci) magnetic character; C9 was first derived from Ci, then from Cp
Cp	Daily magnetic character derived from Kp
Dst	Hourly index mainly related to the ring current
K	3-hr local quasi-logarithmic index
Km	3-hr mean index derived from an average of K indices (not to be confused with the Km of the next item)
Km, Kn, Ks	3-hr quasi-logarithmic (mondial, northern, southern) indices derived from am, an, as
Kp, Ks	3-hr quasi-logarithmic planetary index and the intermediate standardized indices from which Kp is derived (not to be confused with the Ks of the preceding item)
Kw, Kr	3-hr quasi-logarithmic worldwide index and the intermediate from which Kw is derived
Q	Quarter hourly index
R	1-hr range index
RX, RY, RZ	Daily ranges in the field components
sn, ss	3-hr indices associated with an and as
U, u	Daily and monthly indices mainly related to the ring current
W	Monthly wave radiation index

geomagnetic field appears as unipolar (like the gravitational field) and its strength varies as $\sim r^{-2}$, far from the poles it behaves like a dipole $\sim r^{-3}$. The K-index describes its long term stability. A list of different geomagnetic indices which measure the state of the magnetosphere/ionosphere system is given in Table 1.

The shape of the geomagnetic field is strongly influenced by the interaction with the interplanetary magnetic field (IMF) and the solar wind particles. The magnetopause is a region which is defined by the equality of solar wind pressure and magnetic pressure. Reconnection occurs when the interplanetary magnetic field is antiparallel to the Earth's magnetic field. The particles become thermalized in the region between the bow shock and the magnetopause. Shocks occur when an object is travelling faster than the local speed of sound, and in shock fronts properties like density, pressure, velocity change almost instantaneously. In the bow shock, supersonic solar wind hits the magnetosphere.

Both the Alfvén and sound velocity are supersonic (Mach number >1):

$$v_A = \frac{B}{\sqrt{\mu_0 \rho}}, \qquad v_s = \sqrt{\frac{c_p}{c_v}\frac{p}{\rho}}. \tag{10}$$

The solar wind stretches the dipole field, compressing it on the side toward the Sun and stretching it into a long tail where the field lines close at very long distances ($\sim$3000 R_E, R_E Earth's radius).

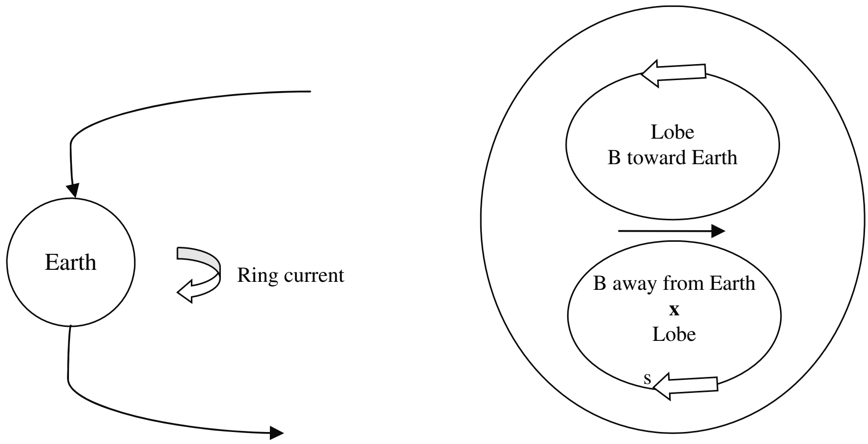


Fig. 1 Global structure of the Earth's magnetic field. In the two lobes the magnetic fields are opposite and the lobes are separated by a plasmasheet. For the southern half of the magnetosphere the current is clockwise. In the middle both systems add to form a neutral sheet. The right drawing is a cross section of the left at a distance of 20 R_E

The plasmasheet is a sheet of plasma in the tail region dividing the two lobes of the Earth's magnetic field (see Fig. 1). For both electrons and protons the particle density is $0.5\,\mathrm{cm}^{-3}$. The lobes in the magnetotail have opposite field directions and are separated by the plasmasheet – otherwise they would cancel. The plasmasphere is a torus shaped region, surrounding Earth. It was detected in 1963 and has a very sharp edge at the plasmapause extending to 4–6 Earth radii, down to the ionosphere. Inside the plasmapause geomagnetic field lines rotate with Earth. Outside the plasmasphere, magnetic field lines are unable to corotate, as the solar wind influence is too large. The plasmasphere is mainly composed of hydrogen.

The Van Allen radiation belts were discovered in 1958 by Van Allen; like the plasmasphere they are toroidally shaped. The outer belt extends from 5 to 10 Earth-radii, R_E and the particles are found in the energy range between 0.1 and 10 MeV. The inner belt extends from 0.1 to 1.5 R_E and particles have energies $\sim$100 MeV.

The variations of the geomagnetic field can be classified into short-term variations, e.g., flares, geomagnetic storms, and more long-term variations (some quasi-periodicity of 250,000 years). When IMF has a southward oriented component ($B_z < 0$), reconnection occurs, i.e., the field lines of the IMF cancel with those of the geomagnetic field. Reconnection can also occur in the magnetotail when strongly enhanced flow of particles result in compression of that region. When the field is cancelled by reconnection, charged particles can penetrate. Therefore, the number of particles increases because of the reconnection at the magnetopause and in the tail (see Fig. 2). Yermolaev et al. (2005), reviewed geomagnetic storm effectiveness of coronal mass ejections (CMEs) and solar flares.

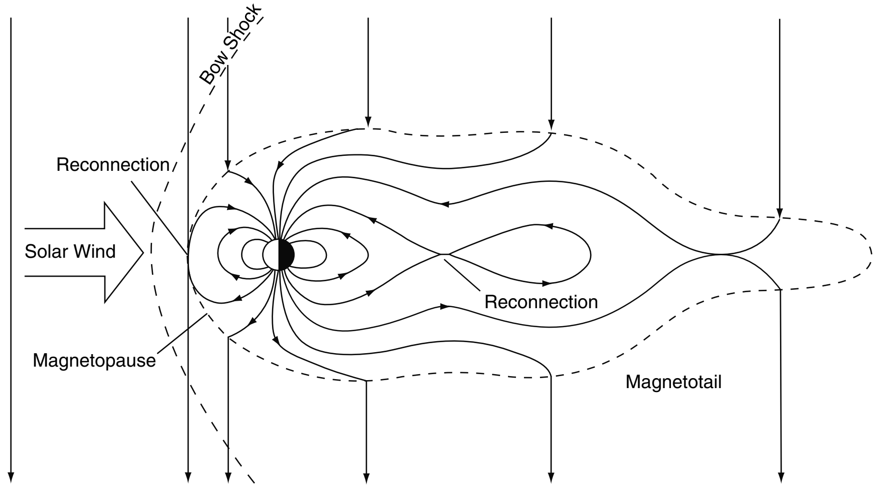


Fig. 2 Reconnection in the bow shock region of the geomagnetic field

3 Solar Energetic Phenomena

We overview briefly the most relevant phenomena on the Sun for space weather. More details can be found in other chapters of this book.

3.1 Active Regions

Big sunspots can be seen even with the naked eye on the solar disk and were already known to the ancients. Sunspots are dark spots associated with magnetic fields that inhibit convective energy transport and thus cool them. They consist of a dark umbra at a temperature $T \sim 4{,}500$ K which is less than the ambient temperature in the photosphere ($T \sim 6{,}000$ K). More than 150 years ago it was found that their number varies with a period of about eleven years. This sunspot cycle is nowadays called solar activity cycle because all energetic phenomena on the Sun vary with this cycle (at least their occurrence).

The solar irradiance is reduced by about 0.02% when a large spot group passes across the Sun. However, this sunspot deficit is overcompensated by faculae that are brighter and hotter regions often seen in the vicinity of spots. In the photosphere they are seen near the solar limb. Faculae can be observed on the whole disk using filtergrams (observations of the Sun at a particular wavelength).

An active region can be defined as a region where magnetic flux penetrates through the photosphere, therefore spots often appear as a bipolar group. From the photosphere, magnetic fields propagate to the above lying chromosphere and corona, with expanding field lines. In the photosphere, because of the larger density, plasma motions can carry the magnetic field lines that are said to be frozen in.

Because of the motions of their footpoints in the photosphere, higher in the chromosphere and corona the field lines can reconnect and magnetic energy is released. This gives rise to a solar flare.

3.2 Flares and CMEs

Flares occur in the corona and produce X-rays and UV-radiation, which indicates high temperatures and energies during a flare outburst. Also energetic particles are emitted and a small fraction of them is accelerated to high energies. As a result, synchrotron radiation, produced by electrons moving in helical paths around magnetic field lines, is generated (this can be observed as bursts in the radio part of the electromagnetic spectrum). During an intense flare, the flux of high energy particles is increased at Earth. Magnetic storms at Earth often occur with a delay of about 36 h after the flaring event was observed on the Sun. Flares occur in regions where there is a rapid change in the direction of the local magnetic field and their energy is a consequence of magnetic reconnection (see Fig. 3).

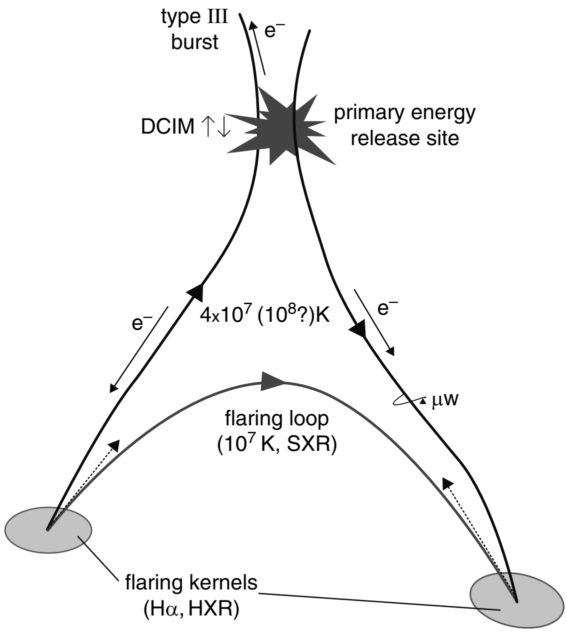


Fig. 3 Two oppositely directed magnetic field lines interact due to a compression – reconnection occurs – and the resulting flaring loop is shown by bold arrow-lines (*grey*). Electron beams are marked by e^-. Chromospheric evaporation from the flare kernels is indicated by thick dotted arrows. The primary energy release occurs in the corona at heights between 10^4 and 10^5 km. DCIM indicates fast drifting bursts in the 200–2,000 MHz range. HXR (hard X-ray) emission is related to radio features. HXR emission at successively lower energies indicates delays of slower electrons relative to faster ones. Courtesy: B. Vrsnak

Table 2 Optical and X-ray classification scheme of solar flares

Importance class	Area A at disk 10^{-6} sol. hemisphere	Energy (erg)	SXR class	Energy 0.1–0.8 nm
S	$A < 100$	10^{28}	A	−8
1	$100 \leq A < 250$	10^{29}	B	−7
2	$250 \leq A < 600$	10^{30}	C	−6
3	$600 \leq A < 1200$	10^{31}	M	−5
4	$A \geq 1200$	10^{32}	X	−4

The power of flares is related to the height of the energy release site. Flares are more powerful and impulsive when the energy release site is located at low heights. This can be explained by the weakening of the magnetic fields with height.

Flares can be classified either by their area on the disk (importance class, see Table 2) or by their soft X-ray emission based on soft X-ray observations of the Sun in the wavelength range 0.1–0.8 nm (peak flux measured in powers of $10\,\mathrm{W\,m^{-2}}$).

Example: A B5 flare has a peak flux of $5 \times 10^{-7}\,\mathrm{Wm^{-2}}$. Flares smaller than C1 can only be detected during a solar minimum phase when the general X-ray background is low. Occasionally, flares exceed class X9 in intensity and are referred simply to as X10, X11, and so on.

Physical processes of particle acceleration, injection, propagation, trapping, and energy loss in solar flare conditions are reviewed by Aschwanden (2002).

CMEs are linked to long duration flares but recent observations also showed that some short duration flares may have ejecta with speeds up to 2000 km s^{-1}. During CMEs the coronal magnetic field lines are opened. There occurs a closing down or reconnection within several hours. The eruption ‘opens’ the magnetic field into interplanetary space, forming a large scale current sheet in the wake of the CME. The field closes down by magnetic reconnection in the current sheet, which results in a prolonged energy release (typical for gradual or eruptive flares). During solar activity minimum about $\sim$0.5 CMEs day^{-1} and during solar activity maximum $\sim$4.5 day^{-1} are observed. The CME mass shows no solar cycle dependence. During solar minimum CMEs are concentrated around the equator, during maximum they originate from a wide range of latitudes.

The SECCHI instruments on each of the two STEREO spacecraft are observing CMEs from their initiation, through the corona, and into interplanetary space beyond the Earth’s orbit. Extreme solar winds that are caused by CMEs produce geomagnetic storms and ionospheric storms. Type II radio bursts are associated with shock acceleration of particles in very fast and wide CMEs and can be used as a proxy for them. Gopalswamy et al. (2005), found that the majority (78%) of the m-to-km type II bursts were associated with solar energetic particle (SEP) events.

3.3 Solar Wind and Interplanetary Magnetic Field

The solar wind is a stream of charged particles expelled from the corona and it varies in strength through the solar activity cycle. The particles arrive at the orbit of Earth at an average speed of about $400\,\mathrm{km\,s^{-1}}$. The total mass loss is a few $10^{-14}\,\mathrm{M}_{\odot}\,\mathrm{year}^{-1}$. The charged particles flow along the open field lines[1] which are cospatial with coronal holes that are regions of lower density and temperature in the corona.

If the solar wind mass loss was constant over the solar evolution the total mass loss of the Sun over that period would be in the order of $10^{-4}\,\mathrm{M}_{\odot}$.
Interplanetary magnetic field (IMF): as the Sun rotates various streams rotate as well (co-rotation) and produce a pattern in the solar wind much like that of a rotating lawn sprinkler. At the orbit of Earth, one astronomical unit (AU) or about 1.5×10^{8} km from the Sun, the interplanetary magnetic field makes an angle of about 45° to the radial direction. Further out[2] the field is nearly transverse (i.e. about 90°) to the radial direction.

The Sun's magnetic field that is carried out into interplanetary space is called the interplanetary magnetic field . When the orientation of the IMF is antiparallel to the Earth's magnetic field, the disturbances are enhanced because of reconnection.

The whole solar system is embedded in the heliosphere. The extent of the heliosphere is determined by the level of solar activity.

4 Solar Radiation and Particles

In this chapter we review shortly the effect of solar radiation and particles on space weather.

4.1 Solar Irradiance

The solar constant S is the amount of the Sun's incoming electromagnetic radiation (solar radiation) per unit area, measured on the outer surface of Earth's atmosphere:

$$S = 1.36\,\mathrm{kWm^{-2}}. \tag{11}$$

The solar irradiance is $S/4$ because the cross section of the Earth is πr^2 and the energy is distributed over the entire surface $4\pi r^2$. The total solar irradiance,

[1] The term open magnetic field lines does not imply magnetic monopoles but means that they are closed very far from the Sun in interplanetary space.

[2] For example, beyond the orbit of Saturn.

TSI, has been monitored from space for 25 years. The solar irradiance is variable due to dark sunspots and bright faculae. Over the cycle 22 the value ranged roughly from 1366.0 to 1367.5 $\mathrm{W m^{-2}}$. The cycle 23 had a similar maximum (Mekaoui and Dewitte 2008). Fröhlich (2006) gave an overview of solar irradiance measurements since 1978.

About half of the solar radiation is in the visible, the other half is in the IR, only a very small fraction is in the UV and at shorter wavelength or in the radio.

4.2 Solar Energetic Radiation

The solar radiation varies with (a) the 11-year solar activity cycle and (b) when large active regions pass across the disk due to the solar rotation. The variation is up to 0.3% in the visible part of the spectrum but strongly increases towards shorter wavelengths. The variation of solar constant between cycles 21 and 22 was 0.1%. Because of this variation, the globally averaged equilibrium surface temperature on the Earth's surface varied about 0.2°C.

Solar radiation shorter than 320 nm represents only 2% of the total solar irradiance; 0.01% of the incident flux is absorbed in the thermosphere at about 80 km and 0.2% in the stratosphere above 50 km (see Table 3). It controls the thermal structure and photochemical processes above the troposphere; for example the stratosphere is controlled by absorption and dissociation of O_2 in the 175–240 nm range. The 205–295 nm range is predominantly absorbed by ozone O_3.

The influence of solar radiation is larger for shorter wavelengths that are absorbed higher in the atmosphere: the exospheric temperature changes from about 700 K (solar minimum) to 1200–1500 K (solar maximum).

4.3 Particles

The particle population in the Earth's atmosphere is maintained by three main contributions:

Table 3 Effects of solar radiation at different wavelengths on the middle and upper atmosphere. Variab denotes the variability within the wavelength range

Wavelength range (nm)	Variab. middle Atm.	Variab. upper Atm.	Effect	Height (km)
1–10, SXR		Sporadic	Ion. all	70–100
10–100, XUV	2 ppm	2 x	Ion. N_2, O, O_2	
100–120, EUV	6 ppm	30%	Ion. NO	80–100
120–200, VUV	150 ppm	10%	Diss. O_2	40–130
200–240, UV	0.12%	5%	Diss. O_2, O_3	20–40
240–300, UV	1.0%	<1%	Diss. O_3	20–40

- Electrons: they reach the high latitude thermosphere after interaction with the geomagnetic field and acceleration;
- High energy solar protons: their flux is enhanced during periods of large flares;
- Galactic cosmic rays: they originate from outside the heliosphere but their input on Earth is partly controlled by solar activity.

During large flares, intense fluxes of energetic protons (10–10^4 MeV) penetrate the Earth's polar cap regions. Ionization between 100 and 20 km is increased and can last for a few hours to a few days. Large numbers of NO_x molecules are produced leading to a subsequent ozone depletion.

There are hints that the Earth's cloud cover follows the variation in galactic cosmic rays (GCRs). The number of these particles hitting the Earth's atmosphere is modulated by the strength of solar activity – the strength of the IMF and solar wind. Cosmogenic isotopes such as ^{14}C are anticorrelated with solar activity. During solar maximum less GCR reach Earth because of the stronger shielding by the heliosphere. These GCRs and the secondary particles they produce may act as condensation nuclei (a recent study was made by Arnold 2006). A variation in cloud cover of 3% during an average 11-year solar cycle could have an effect of 0.8–1.7 W m^{-2}. Besides other influences (e.g. change of the Earth's orbit) the different well known periods of warm and cold climate on Earth can also be explained by the long-term variation of solar activity. Between 1645 and 1715 the Little Ice Age occurred which coincided with almost no solar activity. During that period the production of the cosmogenic isotope ^{14}C was enhanced.

4.4 Space Climate

Table 4 summarizes the effect of different factors on the Earth's climate. As we have demonstrated, the variation of the solar irradiance is low at least in the course of a solar activity cycle. Solar activity however influences on the flux of cosmic ray particles (GCRs), which may act as condensation nuclei for clouds (Svensmark and Friis-Christensen 1997).

A reconstruction of the climate (e.g. by analysis of fossil pollen, cosmogenic isotopes ($^{16}O/^{18}O$, ^{14}C) shows that several periods of warmth and cold have occurred.

Table 4 Various influences on the climate

S, solar constant (at 1 AU)	1360 W m^{-2}
S/4, top of atmosphere	340 W m^{-2}
S/4(1-a), a=0.3 Earth's albedo	235 W m^{-2}
1% change in the albedo	1 W m^{-2}
Estimated rad. effect due to CO_2 increase since 1750	1.5 W m^{-2}
Doubling of CO_2	4 W m^{-2}
Radiative effect of clouds (cooling)	17–35 W m^{-2}

There occurred only episodes of warmth, cold periods lasted much longer. Presently, we are in one of such episodes.

- Little Ice Age for the northern hemisphere from 15th to 19th centuries.
- In the Middle Ages from ~9th to 14th centuries a period of warmth.
- Mid-Holocene Warm Period (approx. 6,000 years ago); this seems to be in connection with changes of the Earth's orbit (theory of Milankovich).
- Penultimate interglacial period (approx. 125,000 years ago). It appears that temperatures (at least summer temperatures) were slightly warmer than today (by about 1 to 2°C), caused again by the changes in the Earth's orbit (Hughes and Diaz 1994).
- Mid-Cretaceous Period (approx. 120–90 million years ago): Breadfruit trees apparently grew as far north as Greenland (55°N), and in the oceans, warm water corals grew farther away from the equator in both hemispheres.

It is clear that there exist other than solar influences. However, solar variations on the long term certainly influence the climate.

During its main sequence evolution, the Sun has become more luminous:

$$L(t) = [1 + 0.4(1 - t/t_0)]^{-1} L_0. \tag{12}$$

In this equation L_0 is the present solar luminosity and t_0 the present age of the Sun (4.6 Gyr). The early Sun had only 70% of its present luminosity but was more active. This is also called the faint young Sun problem. Because of this low solar luminosity, Earth would have been totally covered with ice which would have increased the albedo and therefore the ice would never have melted. There are also strong indications of liquid water present on Mars several 10^8 years ago. Thus there must have been other mechanisms to keep these planets at higher surface temperature, e.g. a higher concentration of greenhouse gases like CO_2 in the primitive Earth's atmosphere.

All calculations show that during the early life of the Sun, the UV flux was much higher than today. The early Sun was similar to a T Tauri star. At the age of about 40 million years the Sun became a main sequence star.

5 Space Weather and Damage

5.1 Effects of Radiation on Biological Systems

The passage of ionizing radiation in a living organisms can result in direct effects on DNA leading to single strand breaks (SSB), double strand breaks (DSB), associated base damage (BD), or clusters of these damage types. Water in the body tends to absorb a large fraction of radiation and becomes ionized forming highly reactive molecules which are called free radicals. Those react with and damage the DNA molecules.

Table 5 Radiation dose limits in mSv for astronauts

Time period	Blood forming organs	Eyes	Skin
30 days	250	1,000	1,500
Annual	500	2,000	3,000
Career for males	2,000 mSv+75(age[years]−30)	4,000	6,000
Career for females	2,000 mSv+75(age[years]−38)	4,000	6,000

Table 6 Total average annual radiation dose in the US

Radon in the air	2 mSv (56%)
Rocks, building material	0.28 mSv (8%)
Cosmic rays	0.28 mSv (8%)
Natural radioactive material in body	0.39 mSv (11%)
Medical and dental rays	0.39 mSv (11%)
Nuclear medicine tests	0.14 mSv (4%)

Typical effects of radiation sickness are severe burns that are slow to heal, sterilization and cancer.

To minimize the risk for astronauts and people involved with radiation in general, the US National Council on Radiation Protection and Measurements has established radiation dose limits. For a comparison, a computed tomography (CT) of the body corresponds to 10 mSv which is comparable to the annual average background radiation for three years (see Table 5).

The radiation dose limits for ordinary citizens are much lower (see Table 6). The annual dose is about 50 mSv, the lifetime dose is age [years]×10 mSv. In the US the total average annual dose is about 3.6 mSv.

5.2 *Surface Charging*

Explosive solar particle events (SPEs) are usually associated with solar flares and coronal mass ejections. Protons and electrons are emitted at high energies which can cause problems in orbiting satellites. In January 1994 three geostationary satellites suffered failures of their momentum wheel control circuity under the long duration influence of high-energy electron fluxes associated with high-speed solar wind streams. Such events may occur also during times of sunspot minimum. The strength of shocks plays a key role in shock-related phenomena, such as radio bursts and solar energetic particle (SEP) generation and this is discussed by Shen et al. 2007.

According to USAF: Damaging conditions are assumed when the daily electron flux (which is given by the number of high-energy >2 MeV electrons per cm^2 per steradian per day) meets either of the following conditions[3]: greater than 3×10^8 per day for 3 consecutive days; or greater than 10^9 for a single day. Such conditions often occur about 2 days after the onset of a large geomagnetic storm.

[3] see also http://www.ips.gov.au/Educational/1/3/7

The K-index is a measure for geomagnetic storms. The values of K (3-hr averages) range from 0–9. $K = 0$ means quiet; $K \geq 4$ surface charging effects could begin, $K \geq 6$ surface charging is probable. Surface charging usually does not cause big problems. Particles with energy $\geq$1 MeV cause deep dielectric charging.

Single-event upsets (SEUs) are random errors in semiconductor memory that occur at a much higher rate in space than on the ground. They are non-destructive, but can cause a loss of data if left uncorrected. SEUs are often associated with heavy ions from the galactic cosmic radiation.

Because of a miss-alignment of the radiation belts due to the tilt of magnetic axis to rotation axis, the Earth's surface magnetic field is weakest in the South Atlantic Anomaly. Therefore particles drifting around Earth travel much closer to Earth than at other latitudes and longitudes.

5.3 Solar Activity and Satellite Lifetimes

Satellites in low Earth orbit (LEO), with perigee altitudes below 2,000 km, are subject to atmospheric drag. This force very slowly circularizes the orbits and the orbital altitude is reduced. The rate of decay of these orbits becomes extremely rapid at altitudes below 200 km. As soon as the satellite is down to 180 km it will have a few hours to live and after several revolutions around Earth it will re-entry down to Earth. Enhanced solar radiation at short wavelengths causes the upper atmosphere to expand and the drag on satellites in LEOs becomes larger.

Solar activity has also influences on solar panels. The most efficient solar panels are the DS1 solar panels which convert about 22% of the available energy into electrical power. The solar panels lose about 1–2% of their effectiveness per year. After a five year mission, the solar panels will still be working with more than 90% power. The two major dangers are: (a) short-wavelength radiation emitted during solar flares can damage the electronics inside the panels, (b) micrometeorites, which are tiny, gravel-sized bits of rock and other space junk floating in space can scratch or crack solar panels.

5.4 Some Examples of Space Weather Damage

The solar proton event in August 1972: This event occurred between the manned Apollo 16 and 17 missions. Even inside of a spacecraft the astronauts would have absorbed a lethal dose of radiation within 10 hr after the onset of the event. At 6:20 UT an optical flare was observed on the Sun. At 13:00 UT the astronauts' allowable 30-day radiation exposure to skin and eyes was exceeded. At 14:00 UT the astronauts' allowable 30-day radiation exposure for blood forming organs and the yearly limit for eyes was exceeded. At 15:00 UT the yearly limit for skin was exceeded. At 16:00 UT the yearly limit for blood forming organs and the career limit for eyes was exceeded. At 17:00 UT the career limit for skin was exceeded.

October 1989 event: A major power failure occurred in Quebec; an astronaut on the Moon would have been exposed to a lethal dose (protection possible by lunar soil). Astronauts on board of Mir received the full year radiation dose within few hours (Table 5).

The Bastille Day event: An extremely powerful solar flare occurred on 14 July 2000. An X5-class solar flare caused a so called S3 radiation storm on Earth and fifteen minutes after the outburst of the flare the ionosphere was bombarded with energetic protons. Such storms cause the following effects on satellites: single-event upsets, noise in imaging systems, permanent damage to exposed components/detectors, and decrease of solar panel currents. It can also expose air travellers at high latitudes to low levels of radiation, the equivalent of a brief chest X-ray (Table 7). Also a full halo coronal mass ejection was observed and $>10^{10}$ t of plasma was ejected into space towards Earth at speeds between 1,300 and 1,800 km s^{-1}. As a consequence of the geomagnetic disturbances, aurorae could be seen even at El Paso, power companies reported on damage of a transform and GPS accuracy was degraded for several hours. The K_p-index was larger than 9.

Mesospheric temperatures were found to be increased by 200 K because of this event (Dymond et al. 2005; Tsurutani et al. 2005).

5.5 *Geomagnetically Induced Currents*

As we have discussed, the magnetic field of Earth is compressed by the interacting with solar magnetic clouds and particles. The changing magnetic field induces currents at Earth and these induced currents produce magnetic fields that again disturb the Earth's surface magnetic field. The magnitude of the induced currents and electrical fields depends on electrical conductivities of the different ground layers. Magnetic variations with lower frequencies penetrate deeper.

To simulate such ground effects, the real time GIC (geomagnetic induced currents) simulator is available at http://www.spaceweather.gc.ca/gic_simulator_e.php.

Table 7 Single dose effects

0.1 mSv	Dental X-rays
0.3 mSv	5 hr transcontinental flight
0.5 mSv	Chest X-ray
100 mSv	Risk of cancer later in life 5 in 1,000
000 mSv	Risk of cancer later in life 5 in 100
2,500 mSv	Sterility in females
3,500 mSv	Sterility in males
4,000 mSv	Average lethal dose (without any treatment)

There are several damaging effects of GICs like enhanced corrosion and power-failures. We mention a few well documented examples:

- On 30 October 2003 50 000 customers in southern Sweden had no electricity due to a power failure caused by a GIC.
- Research on historical geomagnetic storms can help to create a good data base for intense and super-intense magnetic storms. For the event on March 13, 1989 the Dst $= -640$ nT.
- Evidence for a superstorm that occurred on Sep 1–2 1859 with a Dst $= -1\,760$. This event was observed as a white light flare by Carrington. Tsurutani et al. (2006), described the Halloween 2003 solar flares and the Bastille Day event in 2000 and resultant extreme ionospheric effects. The October 28, 2003 flare caused a 30% increase in the local noon equatorial ionospheric column density within 5 min. Interplanetary coronal mass ejection (ICME) electric fields acted with a delay (due to solar wind propagation) with the ionosphere and the total electron content was enhanced by a factor of 300%.

Acknowledgements I want to cordially thank the organizers for inviting me to give the lectures and their kind hospitality.

References

Arnold, F.: Space Sci. Rev. **125**, 169 (2006)
Aschwanden, M.: Space Sci. Rev. **101**, 1 (2002)
Dymond, K. F., Budzien, S. A., McCoy, R. P., Crowley, G.: American Geophysical Union, Fall Meeting 2005, abstract # SA51A–1116, (2005)
Fröhlich, C.: Space Sci. Rev. **125**, 53 (2006)
Gopalswamy, N., Aguilar-Rodriguez, E., Yashiro, S., Nunes, S., Kaiser, M.L., Howard, R.A.: J. Geophys. Res. **110**, 12 (2005)
Hanslmeier, A.: The Sun and Space Weather, ASSL vol. 347. Springer (2007)
Hughes, M.K., Diaz, H.F.: The medieval warm period, Proc. Workshop Tucson 1991 (1994)
Mekaoui, S., Dewitte, S.: Sol. Phys. **247**, 203 (2008)
Shen, C., Wang, Y., Ye, P., Zhao, X.P., Gui, B., Wang, S.: Astrophys. J. **670**, 849 (2007)
Svensmark, H., Friis-Christensen, E.: J. Atmos. Sol.-Terr. Phys. **59**, 11, 1225 (1997)
Tsurutani, B.T., Mannucci, A.J., Iijima, B., Guarnieri, F.L., Gonzalez, W.D., Judge, D.L., Gangopadhyay, P., Pap, J.: Adv. Space Res. **37**, 1583 (2006)
Tsurutani, B.T., Judge, D.L., Guarnieri, F.L., Gangopadhyay, P., Jones, A.R., Nuttall, J., Zambon, G.A., Didkovsky, L., Mannucci, A.J., Iijima, B., Meier, R.R., Immel, T.J., Woods, T.N., Prasad, S., Floyd, L., Huba, J., Solomon, S.C., Straus, P., Viereck, R.: Geophys. Res. Lett. **32**, 3 (2005)
Yermolaev, Yu.I., Yermolaev, M.Yu., Zastenker, G.N., Zelenyi, L.M., Petrukovich, A.A., Sauvaud, J.-A.: Planet. Space Sci. **53**, 189 (2005)

Sun: Climate Coupling on Sub-Decadal to Multi-Millennial Time Scales

Manish Tiwari and R. Ramesh

Abstract Several studies have indicated that the Sun appears to control the terrestrial climate on several timescales that range from less than a decade to several thousands of years. Here we discuss various techniques used to reconstruct past solar activity and present a critical review of the existing theories that attempt to correlate past solar and climatic variabilities with special reference to the phenomenon of South Asian monsoon. We also outline existing problems that need attention to further elucidate the Sun-climate connection.

1 Introduction

The subject of relative contributions of the natural and anthropogenic causes to the presently experienced global warming (IPCC AR4 2007) is hotly debated. This has given a renewed impetus to the research on the effect of solar variability on the climate on various time scales; numerous studies elucidating various causal mechanisms have appeared in the literature during the past couple of decades. Here we attempt to review the existing hypotheses and discuss solar variability and its effect on the Earth's climate. Solar energy governs the Earth's climate, although the latter is further modulated by internal processes such as oscillations in ocean-atmosphere system, volcanic eruptions, atmospheric content of greenhouse gases, cloud cover, ice and vegetation extent and so on. The solar energy being received by Earth in turn varies because of changes in the SunEarth geometry (Milankovitch cycles) and the variation in the incoming solar irradiance at various wavelengths due to changes in the solar activity itself. The Milankovitch cycles have nothing to do with solar variability (processes taking place inside the Sun) and affect the Earth's climate due to changes in the Earth's eccentricity, obliquity and precession (although only the

M. Tiwari (✉)
National Centre for Antarctic and Ocean Research, Vasco da Gama, Goa 403 804, India
e-mail: manish@ncaor.org

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_15,

change in eccentricity affects the solar energy reaching Earth while the other two just redistribute heat among different latitudes). Eddy (1976) was the first to suggest that long-term changes in solar activity can affect climate (colder climate during periods of lower solar activity), which was further reinforced by Reid (1987), who found a remarkable similarity between the globally averaged sea surface temperature (SST) and solar activity. But in spite of two decades of intense research, many questions that deal with the exact nature of the processes taking place inside the Sun that causes variability in its energy output and how this slight variation in the energy output leads to large changes in Earth's climate remain unanswered (Beer et al. 2000; Bard and Frank 2006).

2 Variable Sun

The Sun is variable and this became evident to European astronomers in early seventeenth century through the changing sunspot numbers of which they kept a reasonably good record using the then newly invented telescope. Although the incident solar radiation or the solar energy reaching Earth ($\sim$1,367 W m^{-2}) was thought to be unvarying (hence called "solar constant") as it was not possible to determine the slight fluctuations in it due to the presence of the atmosphere. But since 1978, radiometers are being sent in space (satellites), which revealed that "solar constant" was not a constant but varied with root mean square amplitude of 0.1% as shown in Fig. 1 (Willson 1997; Willson and Mordvinov 2003, Fröhlich and Lean 2004; Bard and Frank 2006); this has been revised to 0.08% (IPCC AR4 2007). Therefore a more precise quantity – "Total Solar Irradiance" (TSI) – was formulated, which is defined as the value of the integrated solar energy flux over the entire electromagnetic spectrum through an area of 1 m^2 oriented towards the Sun arriving at the top of the terrestrial atmosphere at a distance of 1 astronomical unit (AU) from the Sun.

The Sun is composed of many layers that include the core, radiation zone, convection zone, photosphere, chromosphere and the corona. Several processes taking place in these layers have been proposed to cause variability in the solar output on timescales ranging from minutes to billions of years (e.g., Beer et al. 2000):

- *Nuclear Fusion in the Sun's Core* (conversion of hydrogen to helium) is the source of the Sun's energy (luminosity) and it has been estimated that about 4 billon years ago, solar output was 75% of what is observed today and it will be 35% more than today after another 4 billion years (Newman and Rood 1977; Gough 1977). But the rate of change on billion year time scale is exceedingly small (10^{-4} Wm^{-2} kyr^{-1}); this has negligible effect on climate change on human/historical time scales.
- *Changes in Convective Pattern* – In the radiation zone, energy is transferred by interaction with the surrounding atoms and is considered very stable on a million year time scales. But the energy transport through the convection zone is considered to be a significant source of variability on a 100 kyr timescale (Hoyt and Schatten 1993; Nesme-Ribes et al. 1994).

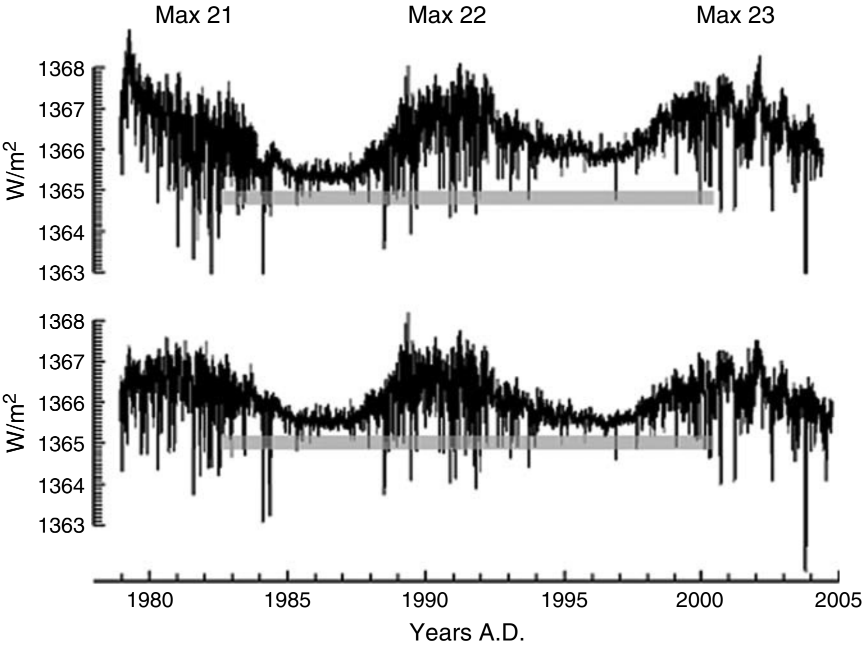


Fig. 1 TSI from satellite borne radiometers; as the raw data is processed by applying different correction factors that may yield different scenarios (*grey* bars exhibit long term trends) – the top panel exhibits a slight increasing trend (Willson and Mordvinov 2003) while the bottom one stays uniform (Fröhlich and Lean 2004); Max 21, 22 and 23 depict maxima in solar irradiance in the respective solar cycles (Reprinted from Earth and Planetary Science Letters, 248, Bard and Frank, Climate change and solar variability: What's new under the sun?, 1–14, Copyright 2006, with permission from Elsevier)

- *Changes in Photospheric Features* – TSI variability is a composite of various temporal and spatial changes occurring on the photosphere that include features such as changes in its temperature (Kuhn and Libbrecht 1991; Gray and Livingston 1997), and changes in the amount and distribution of magnetic flux on the solar surface (Foukal and Lean 1988; Lean et al. 1998). The thermal and magnetic perturbations originating at the bottom of the convection zone is manifested on the photosphere in the form of sunspots and associated faculae and so on. During higher solar activity, the number of sunspots increases. But the Sun, on an average, is brighter during the higher sunspot numbers as the brightening due to the faculae and the network outweighs the darkening due to enhanced sunspots (Lean et al. 1997). Changes in photospheric features have yielded variability on timescales ranging from minutes to decades (longer periodicities could not be ascertained as space borne radiometer data is available only for the past few decades).

Another source of TSI variability has been attributed to the changes in solar diameter (Sofia and Fox 1994; Sofia and Li 2001), although recent studies have shown that solar diameter changes are a few kilometre per year during a solar cycle contributing only 0.001% to the TSI variability (Dziembowski et al. 2001). Low frequency, millennial scale fluctuations have also been proposed (Beer et al. 2000) that may affect the climate change on longer timescales.

The Sun exhibits periodicities at various timescales; the most obvious being the 11-year Schwabe cycle (first noted in 1843 AD by an amateur solar astronomer H. Schwabe based on the changing number of sunspots), which is related to the 22-year Hale cycle (described by George Hale in the early twentieth century) in which the magnetic polarity of the Sun reverses and then returns back to the original state and the 33-year Bruckner cycle. As instrumental data is available only for the past three decades, longer cycles are hard to recognize; based on sunspot numbers and other proxy records (e.g., cosmogenic radionuclides), the 88-year Gleissberg cycle, the 208-year Suess cycle,the ∼2,300-year Hallstatt cycle (Mayewski et al. 1997) and even a 100,000-year cycle (Sharma 2002) along with the following periods of low solar activity have been documented (Beer et al. 2000): Oort minimum (980–1120 AD), Wolf minimum (1282–1342 AD), Spörer minimum (1416–1534 AD), Maunder minimum (1645–1715 AD) and Dalton minimum (1790–1820 AD) during the past millennium.

3 Solar Spectral Variability

Solar irradiance is composed of various wavelengths (corresponding to a black body radiation at 5,770 K) and the variability in the energy distribution among these wavelengths contributes to the solar spectral variability as shown in Fig. 2. The TSI variability is mainly limited to wavelengths below 500 nm and is considerably more for the shorter wavelength, i.e. in the ultraviolet region rather than in the visible region. The spectral irradiance in the UV region is continuously being monitored from space since 1991 onwards (Woods et al. 1996). In the UV region, solar cycle irradiance changes of ∼20% near 140 nm, 8% near 200 nm and 3% near 250 nm have been observed (Lean et al. 1997) whereas the TSI variability during a solar cycle is ∼0.08% (IPCC 2007) which equals a radiative climate forcing of 0.22 Wm^{-2} (Lean and Rind 1999). Spectral variability in the case of wavelengths larger than 400 nm (visible and infrared) is not well known due to the lack of long term measurements and is theoretically estimated (Solanki and Unruh 1998). Recently, space-borne instruments have also measured variability in the visible and infrared region and results confirm that variability occurs at all wavelengths (Harder et al. 2005; Lean et al. 2005) primarily in response to the varying number of sunspots and faculae. The short wave part of the spectrum (<400 nm) contributes only 7% of the total irradiance (Beer et al. 2000) but can be an important parameter affecting the climate if there is a mechanism that is exceptionally sensitive to even minor variations in it. But the IPCC AR4 (2007) clearly states that these observations are too limited temporally to make any definite statement about the total irradiance variability during a solar cycle.

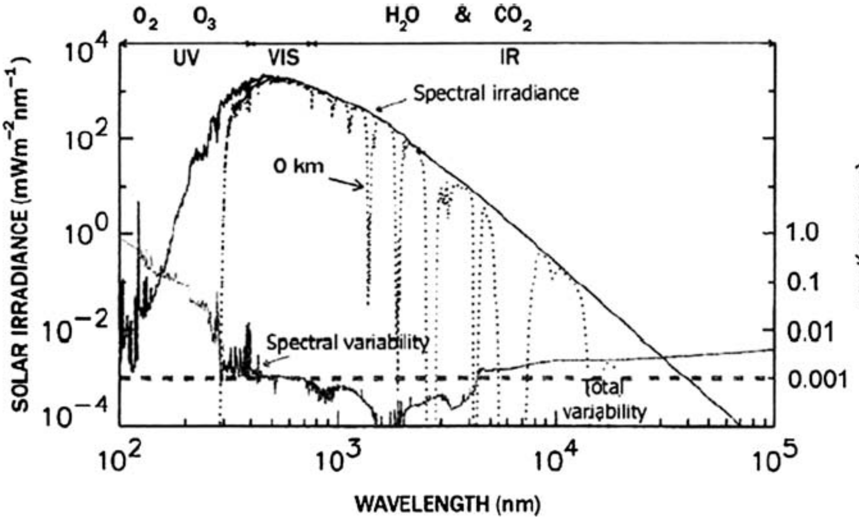


Fig. 2 The solar irradiance reaching the top of the atmosphere as a function of wavelength, the *dotted line* represents the irradiance reaching the Earth's surface (0 km) after absorption by ozone and other atmospheric gases in the UV and near-IR region respectively; also shown is the TSI variability in different spectral bands (spectral variability) and the horizontal dashed line depicts variability (total variability) in the solar irradiance along the 11-year solar cycle (right hand side scale; Reprinted from Space Science Reviews, 94, Lean, Short term, direct indices of solar variability, 39–51, Copyright 2000, with permission from Springer)

4 Records of Past Solar Activity

The Sun may exhibit much higher variability and cyclicity than what is found in the instrumental or telescopic record. It is necessary to reconstruct a long term record (on millennial timescale) for which cosmogenic radionuclides (e.g. ^{14}C, ^{10}Be etc.) are employed as proxies for solar activity. These radionuclides are produced in the atmosphere by interaction with the galactic cosmic rays (GCR, composed of protons, alpha particles, electrons, gamma rays and other atomic nuclei), the influx of which in the Earth's atmosphere is modulated by the solar wind and the associated magnetic field. During higher solar activity, the deflection due to the solar magnetic field is enhanced; hence less galactic cosmic rays reach the Earth's atmosphere, resulting in the lower production of cosmogenic nuclides. This has been inferred by comparing the precisely assigned ages of tree rings by counting the rings with that estimated by ^{14}C dating. The limitations in such an approach include ignoring (a) the strength and direction of the geomagnetic field that influences the production rate of cosmogenic radionuclides, and (b) changes in transport and storage of these nuclides in various Earth reservoirs caused by changing climate (explained below).

4.1 ^{14}C Record

Radiocarbon (^{14}C) is produced in the atmosphere by the interaction of cosmic ray produced neutrons with nitrogen nuclei via (n,p) reaction:

$$^{14}N_7 + ^{1}n_0 \rightarrow ^{14}C_6 + ^{1}H_1 \tag{1}$$

This radiocarbon atom is oxidized to $^{14}CO_2$ rapidly and is taken up by plants and trees via photosynthesis. The ^{14}C decays by β emission to stable ^{14}N with a half-life of ~5,730 years. To reconstruct the past ^{14}C activity, materials such as annual growth rings of trees, annual bands of corals and varve sediments have been used (Chakraborty et al. 1994; Chiu et al. 2007). Past concentration of the radiocarbon (^{14}C) can be obtained if the calendar and ^{14}C activity of the sample are known, and is denoted by $\Delta^{14}C$ (‰). It indicates the departure from modern pre-industrial, pre-nuclear atmospheric ^{14}C concentration (by convention, the ^{14}C activity of the 1,850 wood) after correction for a mass dependent fractionation and can be calculated as follows (Stuiver and Polach 1977):

$$\Delta^{14}C = [(e^{\lambda 1t1}/e^{\lambda ctc}) - 1]1000(‰), \tag{2}$$

where λ_1 is the decay constant for the correct ^{14}C half-life (5,730 years, Godwin 1962);

t_1 is calendar age (obtained by the U-Th dating of corals, dendrochronology etc.);

λ_c is the decay constant for the conventional ^{14}C half-life (5,568 years, Libby 1955);

t_c is the reservoir age corrected radiocarbon age based on the conventional half-life.

4.1.1 Geomagnetic Influence on ^{14}C Production

Another factor affecting the ^{14}C production in a non-linear way, apart from the solar wind and the associated magnetic field, is the Earth's magnetic field (as it deflects away the incoming galactic cosmic rays) as shown in Fig. 3. It has been shown that there is a long-term change (2,000–3,000 year time scale) in the ^{14}C production due to the changing geomagnetic field intensity, which has been approximated and taken into account to obtain the corrected ^{14}C activity, called the "residual $\Delta^{14}C$ value (Stuiver and Quay 1980). It has been recently proposed that the variability in the geomagnetic field intensity on the centennial time scale can also contribute significantly to the ^{14}C production rate (Snowball and Sandgren 2002; St-Onge et al. 2003). But as evident from Fig. 3, the short-term variability in the geomagnetic field intensity is very small (depicted by arrows), hence it will not affect the ^{14}C production much, although the low temporal resolution of the paleointensity data and the differences between various reconstructions remain a major source of uncertainty (Muscheler et al. 2007).

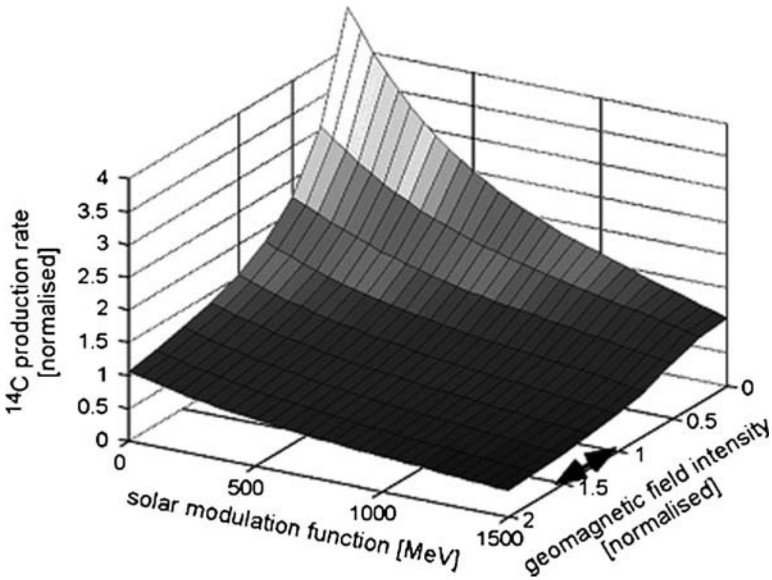


Fig. 3 Normalised ^{14}C production rate as a function of solar modulation and geomagnetic field intensity (Masarik and Beer 1999) that is normalised with respect to the present value; arrows represent range of variability in the geomagnetic field intensity for the past 1,000 years (Reprinted from Quaternary Science Reviews, 26, Muscheler et al., Solar activity during the last 1000 yr inferred from radionuclide records, 82–97, Copyright 2006, with permission from Elsevier)

4.1.2 Effect of Climatic Variations on Transport and Storage of ^{14}C

^{14}C, produced in the atmosphere, is distributed between various reservoirs such as biosphere, ocean's mixed layer and the deep ocean. If the production of deep ocean water reduces, then the activity of ^{14}C in the atmosphere increases. For example, during the last glacial period, the atmospheric Δ^{14}C increased by 100‰ due to the reduced oceanic (thermohaline) circulation (Bard 1998). Thus variation in the oceanic ventilation (and global carbon cycle) with a changing climate is the largest single source of uncertainty to the ^{14}C content of the atmosphere (Siegenthaler et al. 1980). Also, the annual ^{14}C production rate is much smaller than the ^{14}C content in the global carbon reservoir due to which short-term variations in ^{14}C production gets diminished. Therefore to use ^{14}C as a proxy for the past solar activity, the redistribution of ^{14}C in different carbon reservoirs needs to be accounted for by using global carbon cycle models (Muscheler et al. 2007).

To overcome these limitations, recently attempts have been made to reconstruct solar activity changes for the past 32,000 years using cosmogenic in-situ ^{14}C in ice of Greenland (Lal et al. 2005). The advantage is that at such high latitudes the effect of geomagnetic field fluctuation can be neglected to a first order. More importantly, climatic variations that affects the distribution of ^{14}C between different reservoirs (such as atmosphere, surface and deep-ocean etc.) is insignificant as in the case of polar ice, the ^{14}C is produced in-situ by nuclear interactions of fast neutrons with oxygen nuclei.

4.2 ^{10}Be Record

^{10}Be is produced in the atmosphere due to cosmic ray induced spallation of nitrogen and oxygen atoms by high-energy particles (Lal and Peters 1967). ^{10}Be can also be produced in the exposed rocks and soils by in-situ production. It decays by β emission to ^{10}B with a half-life of 1.51 Myr (Hoffman et al. 1987). ^{10}Be sticks to aerosols and is scavenged via precipitation. Consequently, the residence time of ^{10}Be is very short (1 week to 2 years; McHargue and Damon 1991), which prevents its homogenization in the atmosphere, and it is deposited at the latitude at which it is produced. This may lead the local atmospheric circulation patterns to dominantly affect its distribution. The ^{10}Be record reconstructed from the polar ice may under-estimate solar activity as poles are strongly shielded from the solar wind and the associated magnetic field that would lead to higher production rate of ^{10}Be compared to the global production. Modelling results by Field et al. (2006) have shown that the solar induced production variation at the poles should be $\sim$20% more than the globally averaged value. Varying precipitation may also add to the uncertainty of the ^{10}Be reconstruction (Wagner et al. 2001). Here the advantage lies with ^{14}C, which has a relatively longer residence time ($\sim$5 years) in the atmosphere that leads to its global homogenization. Due to the lack of proper understanding of the global transport and deposition of ^{10}Be, a simple assumption is made that the concentration of ^{10}Be in the ice cores represent the production rate of ^{10}Be in the atmosphere (Muscheler et al. 2007). The concentration of ^{10}Be in marine sediments also depends on the (sedimentation) rate at which ^{10}Be atoms are removed from the sea surface to the bottom sediments. In spite of this limitation, a globally stacked record of ^{10}Be concentration from marine sediments has been reconstructed that yields relative variations in the ^{10}Be production rate for the past 200,000 years (Frank et al. 1997). Based on this record, Sharma (2002) has shown that solar activity exhibits a 100,000-year cycle that is strongly correlated to the 100 ka glacial-interglacial cycle and proposed that solar activity is controlling the climatic variations on a 100,000-year timescale.

Muscheler et al. (2007) compared the solar modulation function reconstructed from a stacked ^{10}Be record (using data from several ice cores from Greenland and Antarctica) with that from the ^{14}C record (Fig. 4), and found a reasonably good agreement. This implies that the averaging of the ^{10}Be data from various records from polar ice can provide good estimates for the global production rates of ^{10}Be. Although there are dissimilarities as well; the most pronounced being that the solar modulation function based on ^{10}Be shows higher short term variability than that based on the ^{14}C record. It could be due to local effects in transport and deposition of ^{10}Be.

Solanki et al. (2004) reported that the current level of solar activity (during the past 70 years) has been higher than what was experienced during the past 8,000 years. But after taking into account the the ^{14}C record corrected for fossil fuel burning, Muscheler et al. (2007) has shown that the mean value of solar activity over the last 55 years has been reached or exceeded several times in the past 1,000 years. This means that the current value is certainly high but such episodes were not uncommon in the past.

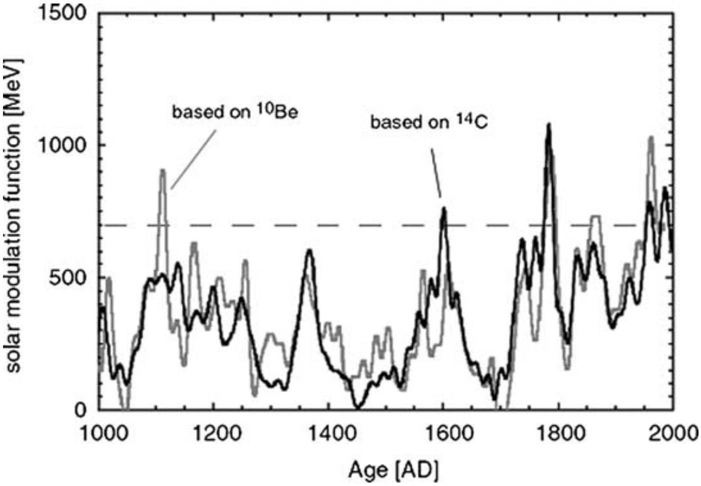


Fig. 4 Solar modulation function based on ^{10}Be (*grey line*) *vs.* ^{14}C (*black line*); normalized and obtained after low-pass filtering (cut of frequency of $1/20\,\mathrm{yr}^{-1}$); Reprinted from Quaternary Science Reviews, 26, Muscheler et al., Solar activity during the last 1000 yr inferred from radionuclide records, 82–97, Copyright 2006, with permission from Elsevier

5 Sun and Terrestrial Climate Variations: Causal Mechanisms

Earth receives $\sim$1,367 Wm^{-2} of solar irradiance that gets distributed over the planet and thus the top of the atmosphere receives $\sim$342 Wm^{-2} (one-fourth of 1,367 Wm^{-2}, i.e., the area Earth projects normal to the Sun divided by the total surface area of Earth). After accounting for the Earth's albedo ($\sim$30%), it further reduces to $\sim$239 Wm^{-2}. The effect of variability in this irradiance reaching the Earth's surface has been estimated using several techniques (Mendoza 2005). For example, Lean et al. (1995) explored the linear correlation between the TSI and the northern hemisphere temperature anomalies since 1610 AD. They observed a high correlation between them but found that only one-third of the warming observed since 1970 can be explained by solar forcing, which is further supported by Solanki and Krivova (2003). The total radiometrically observed TSI variation is 0.08% (IPCC AR4 2007; corresponds to a change in radiative forcing of $\sim$0.24 Wm^{-2}), but it is estimated (by studying Sun-like stars) that TSI might be lower by 0.25% during the Maunder minimum than at present, corresponding to a change in radiative forcing of $\sim$0.5–0.7 Wm^{-2} respectively (Lean et al. 1995; Hoyt and Schatten 1993; Solanki and Fligge (1999). However, according to the latest IPCC AR4 report (2007), the total increase in TSI from Maunder minimum to the minimum of the present solar cycle is only 0.04%, corresponding to an increase in radiative forcing of 0.1 Wm^{-2}. This estimate is based on flux transport simulations of the total magnetic flux evolution from the Maunder minimum to the present (Wang et al. 2005a). It is further argued that the earlier reconstruction quoting higher values of increase in radiative forcing (up to 0.7 Wm^{-2}) were founded on assumed changes in TSI during solar cycle minima based on brightness fluctuation in Sun-like stars, which are not considered valid anymore. Lean et al. (2005) also agreed that the decrease in irradiance during the Maunder minimum was overestimated by earlier studies. The increase in the mean irradiance since the Maunder minimum to the present cycle mean, although, is 0.08% (double of the Maunder minimum to the present minimum value) as the amplitude of the 11-year solar cycle has increased since Maunder minimum. In spite of such a low variability in TSI, its apparent ability to affect the climate indicates that some indirect mechanisms with a positive feedback are involved, some of which are discussed below.

The first involves the heating of the Earth's stratosphere by the increased absorption of the solar ultraviolet (UV) radiation by ozone during periods of enhanced solar activity (Haigh 1994). As discussed earlier, the solar irradiance varies much more in the UV spectral band (up to a maximum of 15%) during a solar cycle as compared to TSI that varies only by $\sim$0.08% (IPCC AR4 2007). Several workers have noted that changes in the UV radiation during the solar cycle causes changes in the stratospheric ozone production (Fioletov et al. 2002; Geller and Smyshlyaev 2002; Hood 2003), temperature and circulation (Ramaswamy et al. 2001; Haigh 2003; Crooks and Gray 2005), and quasi-biennial oscillation of zonal winds (McCormack 2003; Salby and Callaghan 2004). In fact, Gray et al. (2005) has reported with enhanced statistical confidence that stratospheric ozone concentration, temperature and wind circulation have signals of the 11-year solar cycle. During the maximum of a solar cycle the stratospheric ozone production increases by 2–3% along with a temperature increase in the upper stratosphere (a rise of 1°C has been reported at 50 km). This heating is transferred to the troposphere as shown by the theoretical models (Schneider 2005), mostly through the action of large scale atmospheric waves. It further affects the tropospheric circulation, e.g. the increased Hadley circulation reported during enhanced solar activity (Haigh 1994), and thus affects the distribution of atmospheric moisture.

The second mechanism, which has generated considerable interest as well as criticism, is that during periods of higher solar activity, the flux of galactic cosmic rays to the Earth is reduced, generating less cloud condensation nuclei, resulting in less cloudiness (Svensmark and Friis-Christensen 1997; Svensmark 1998), which in turn results in a lower albedo. Svensmark and Friis-Christensen (1997) have reported a reasonably good correlation coefficient (0.93) between the 12 months running means of the total cloud cover (data obtained from the International Satellite Cloud Climatology Project) and cosmic ray intensity measured at Climax Neutron Monitor Station, Colorado. The reduction in cloud cover is significant [a decrease of 3% from 1987 (solar minimum) to $\sim$1990 (increasing solar activity)], which would result in global warming corresponding to 1–1.5 Wm^{-2} (Rossow and Cairns 1995). But lately this hypothesis has come under fire (Laut 2003; Damon and Laut 2004) as the above mentioned correlation disappears on a period of several decades (Bard and Frank 2006). Questions were raised about the type of the cloud affected by this phenomenon – high or low altitude clouds – because high-altitude clouds tend to warm the climate as they trap the outgoing radiation whereas low-altitude clouds cool Earth as they reflect the solar irradiance. During high solar activity, when less

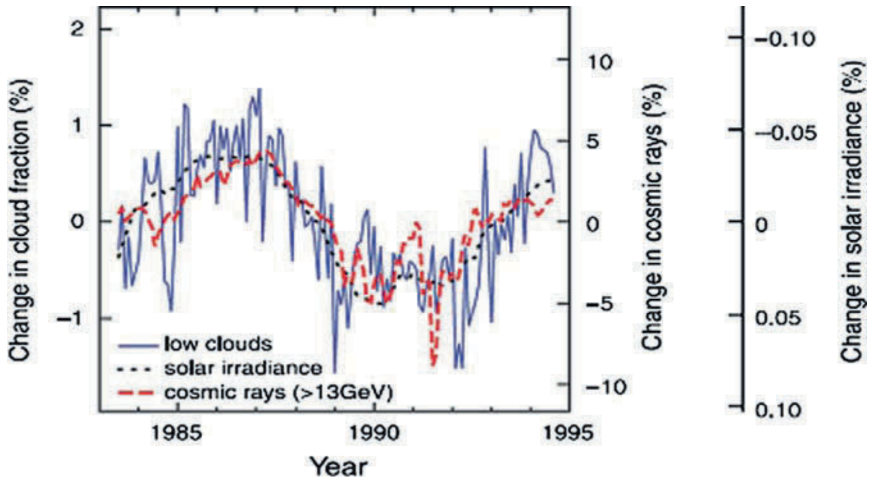


Fig. 5 Correlation between low-altitude clouds (*blue line*), TSI (*dotted line*) and cosmic rays (*red broken line*); note that the solar irradiance scale is reversed – during periods of lower solar activity, there is more of cosmic rays and low-altitude clouds (from (1) Carslaw et al. Cosmic rays, clouds, and climate. Science 298, 1732–1737, 2002, Reprinted with permission from AAAS; (2) Reprinted figure with permission from Marsh & Svensmark, Physical Review Letters, 85, 5004–5007, 2000, Copyright (2000) by the American Physical Society)

cloud condensation nuclei are produced, the reduction in high altitude clouds will tend to cool the climate – opposite to what proposed earlier. In response, Marsh and Svensmark (2000) revised the hypothesis to be limited only to low-altitude clouds as they exhibit better correlations with solar variation between 1983 and 1995 (Fig. 5).

This correlation between low-altitude clouds and solar activity is further criticized: why does the variation in solar activity not produce cloud condensation nuclei in the upper parts of the atmosphere, and why solar activity affects only the lowest part where already sufficient cloud condensation nuclei are present (Bard and Frank 2006). It has also been reported that even the correlation between low-altitude clouds and solar activity breaks down once the cloud cover data beyond 1995 is considered (Laut 2003). Kristjánsson et al. (2004) also observed a correlation between globally averaged low-latitude clouds and solar activity, though not statistically significant. Pallé (2001) concludes that the observed correlation is an artifact of the satellite observing perspective and possibly the spatial distribution patterns of the high-altitude clouds are changing, rather than the areal extent of clouds. Therefore even if a correlation exists, it does not suggest any direct relationship between solar activity and clouds as high latitude clouds are getting redistributed in response to changing atmospheric circulation due to the changing UV radiation and TSI – the first mechanism discussed above. Despite the criticism, this theory attracts considerable research as it provides a very elegant way by which the weak solar signal gets amplified and affects the climate system; probably a longer time series of solar as well cloud data and improved signal to noise ratio, especially in cloud data, can help test this hypothesis (Carslaw et al. 2002).

Another proposed mechanism involves coupling between the ionosphere and the troposphere (Crowley 1983). The ionosphere is the region in the atmosphere (at the altitude of ∼80–250 km) where the atoms absorb the incoming solar energy in the UV band and produce ions and free electrons. The varying solar activity also varies the ionospheric charge, which is intimately coupled with tropospheric thunderstorms. Variations in the ionospheric electric field could affect the electric charge on the clouds, which in turn could play an important role in coalescence of droplets and condensation of water vapour (Markson and Muir 1980). Another line of study indicates that variations in solar wind disturbances (dynamic pressure) may cause changes in climate via a global electric circuit and thermal regime changes of the stratosphere at polar and mid-latitudes (Makarova and Shirochkov 2000, 2004). Makarova and Shirochkov (2002) noted enhanced stratospheric temperature at the North and South poles during periods of enhanced solar wind dynamic pressure (while the opposite effect was observed at Vostok, located at a distance of 1,500 km from the South Pole). Makarova and Shirochkov (2004) conclude that the solar wind energy has a direct affect on the Earth's stratosphere and exhibits periods of 10–12 years (related to the solar flare production) and 30–40 years (similar to that exhibited by TSI) that needs to be taken into account in modeling efforts to understand Earth's climate variability.

A related mechanism has been proposed that tries to correlate the Earth's magnetic field and climate via changes in the cloud cover (e.g., Courtillot et al. 2007). About 100 yr long periods of increase in the geomagnetic field intensity by 15–30%, called as "Archeomagnetic Jerks" have also been reported extending back to 4,000 yr BP that show a very good correlation with cool episodes in the Northern Hemisphere (Gallet et al. 2005). Such jerks could temporarily alter the geometry of the Earth's magnetic field (the dipole tilts to lower latitudes) causing cosmic rays to interact with the more humid troposphere generating more intense cloud condensation (Feynman and Ruzmaikin 1999; Gallet et al. 2005; Courtillot et al. 2007) and hence a higher albedo leading to cooling events.

6 Solar Influence on Monsoon

An earlier study by Mehta and Lau (1997) did not find any consistent relationship at the 11-year time scale between the Indian summer monsoon – the southwest (SW) monsoon – and solar activity. However recent studies analyzing instrumental rainfall time series have shown that there exists a significant positive correlation with the sunspot activity during 1871–2000, and FFT and wavelet analyses of the southwest monsoon rainfall series show periods that match with those exhibited by sunspot numbers (Hiremath and Mandi 2004; Bhattacharya and Narasimha 2005). There have been several paleoclimatic studies using different proxies from different regions that exhibit periodicities similar to those of solar activity (Table 1).

The presence of similar periodicities in two phenomena does not necessarily mean that they are related in general. However, in the case of the Sun-monsoon association several common decadal to millennial periodicities do occur (reported

Table 1 A select list of paleoclimatic studies that indicate that the Sun plays a major role in governing the variations as observed in the Asian monsoon on decadal to millennial timescales

S. No.	Reference	Proxies, material & area of study	Span (Year BP)	Average resolution	Solar periodicities (years)
1.	Castagnoli et al. (1990)	$CaCO_3$ in core from Ionian Sea	0 to ∼1, 820	∼3.87 yr per sample	∼200
2.	von Rad et al. (1999)	Thickness of varve sediments off the Karachi coast	0 to ∼5, 000	∼7 yr per sample	∼250, 125, 95, 56, 45, 39, 29-31, 14
3.	Hong et al. (2000)	$\delta^{18}O$ in peat cellulose, NE China	0 to ∼6, 000	∼20 yr per sample	[a]86, 93, 101, 110, 127, 132, 140, 155, 207, 245, 311, 590, 820, 1,046
4.	Hong et al. (2001)	$\delta^{13}C$ in peat cellulose, NE China	0 to ∼6, 000	∼20 yr per sample	[a]70, 80, 90, 107, 110, 123, 134, 141, 162, 198, 205, 249, 278, 324, 389, 467, 584, 834, 1,060
5.	Neff et al. (2001)	$\delta^{18}O$ of stalagmite, Oman	∼6, 100 to 9,600	∼4.1 yr per sample	∼1,018, 226, 28, 10.7, 9
6.	Agnihotri et al. (2002)	Org. carbon, N and Al content in a sediment core off Gujarat coast	0 to ∼1, 200	∼20 yr per sample	∼200, 105, 60
7.	Leuschner and Sirocko (2003)	Wind and upwelling indices in a sediment core, Oman margin	∼5, 000 to 135,000	150–500 yr per sample	1,100, 1,450, 1,750, 2,300
8.	Fleitmann et al. (2003)	$\delta^{18}O$ of stalagmite, Oman	∼400 to –10,300	∼4.5 yr per sample	∼220, 140, 90, 18, 11
9.	Staubwasser et al. (2002)	$\delta^{18}O$ of G. ruber in a marine sediment core off Pakistan	0 to ∼10, 000	∼17 yr per sample	∼220
10.	Wang et al. (2005b)	$\delta^{18}O$ of stalagmite, Southern China	0 to ∼9, 000	∼5 yr per sample	∼512, 206, 148, 24, 22-17,15-16
11.	Ji et al. (2005)	Iron oxide content in a lacustrine core, Tibet	0 to ∼17,500	∼ 38 yr per sample	∼293, 200, 163, 123
12.	Tiwari et al. (2005)	$\delta^{18}O$ of 3 planktic foraminiferal species in a sediment core off Mangalore coast, India	∼450 to 1,200	∼50 yr per sample	∼200

(*Continued*)

Table 1 (Continued)

S. No.	Reference	Proxies, material & area of study	Span (Year BP)	Average resolution	Solar periodicities (years)
13.	Gupta et al. (2005)	%G. bulloides in sediment core off Oman Margin	∼0 to 11,000	∼30 yr per sample	∼1,550, 152, 137, 114, 101, 89, 93, 79
14.	Thamban et al. (2007)	Elemental conc. & magnetic susceptibility in core from eastern Arabian Sea OMZ	∼350 to 10,000	∼18 yr per sample	2,200, 1,350, 950, ∼700, 470, 320, 220, 156, 126, 113, 104, 100, 92
15.	Yadava and Ramesh (2007)	$\delta^{13}C$ & $\delta^{18}O$ in speleothem from Karnataka, India	0 to 330	∼1 yr per sample	∼132, 21, 18, 2.4

[a]Significance level not specified, many of these peaks likely to be spurious

by many workers in diverse proxies and regions), which implies a consistent relationship between the two on various time scales. The changes in solar radiative heating in the troposphere are considered too weak to affect the monsoon system but indirect mechanisms have been proposed that may amplify these changes. For example, Kodera (2004) proposes a stratospheric origin using modern meteorological datasets. It relates to the mechanism discussed earlier wherein variation of the solar UV irradiance creates circulation change in the stratosphere that finally gets transmitted to the troposphere. During the boreal summer, convective activity builds over the Indian Ocean sector at two places – one over land (in the Northern Hemisphere) and another over the ocean in the Southern Hemisphere that can get excited through external forcing (Chandrasekar and Kitoh 1998). Stratospheric perturbations modulate the equatorial convective activity, which may manifest itself as a north-south seesaw of convective activity in the Indian Ocean sector; causing higher precipitation during higher solar activity (Kodera 2004).

7 Conclusion

It has been shown that the Sun exerts control over climate through direct (radiative forcing of the troposphere) that gets amplified via several indirect mechanisms as discussed earlier. Climate models have shown that variations in TSI can explain the pre-industrial increase in temperature (i.e. 0.3°C surface warming from 1650 to 1790) but it accounts for only 0.25°C of the 0.6°C warming experienced from 1900 to 1990 (Lean and Rind 1999; Mendoza 2005). It signifies that other forcings (most probably the anthropogenic emission of greenhouse gases) are responsible for the rapid increase in the surface temperature observed during the twentieth century. As far as the monsoonal circulation (a major part of the global climate) is concerned, solar variability seems to play a significant role in governing it on decadal to centennial time scales *via* direct and indirect radiative forcing (Kodera 2004).

More focused modeling efforts are needed to fine tune our understanding of the Sun-climate relationship particularly in finding the exact link. Level of scientific understanding quoted IPCC AR4 (2007) for solar forcing on climate is "low" (i.e., chances of being correct are only 2 out of 10). Several hypotheses could benefit from longer data sets, e.g. to comprehend the correlation between solar activity and cloud cover (Bard and Frank 2006). The foremost aim is to understand the total solar variability on different time scales: what is the total amplitude of the TSI variation on a long term and is it sufficient to explain the observed climatic changes? For this a more accurate and precise reconstruction of the past variations in cosmogenic nuclide concentrations must be carried out. How do different spectral bands of TSI behave with varying solar activity? One area that needs urgent attention is the various internal amplification mechanisms, which might significantly alter the climate, even when the total variability in TSI is not much as discussed in Sect. 1.5.

Acknowledgements This is NCAOR contribution no. NCAOR-R-44.

References

Agnihotri, R., Dutta, K., Bhushan, R., Somayajulu, B.L.K.: Evidence for solar forcing on the Indian monsoon during the last millennium. Earth Planet. Sci. Lett. **198**, 521–527 (2002)

Bard, E.: Geochemical and geophysical implications of the radiocarbon calibration. Geochim. Cosmochim. Acta **62**, 2025–2038 (1998)

Bard, E., Frank, M.: Climate change and solar variability: What's new under the Sun? Earth Planet. Sci. Lett. **248**, 1–14 (2006)

Beer, J., Mende, W., Stellmacher, R.: The role of the Sun in climate forcing. Quat. Sci. Rev. **19**, 403–415 (2000)

Bhattacharya, S., Narasimha, R.: Possible association between Indian monsoon rainfall and solar activity. Geophys. Res. Lett. **32**, L05813 (2005)

Carslaw, K.S., Harrison, R.G., Kirkby, J.: Cosmic rays, clouds, and climate. Science **298**, 1732–1737 (2002)

Castagnoli, G.C., Bonino G., Caprioglio, F., Serio, M., Provenzale, A., Bhandari, N.: The CaCO3 profile in a recent Ionian Sea core and the tree ring radiocarbon record over the last two millennia. Geophys. Res. Lett. **17**, 1545–1548 (1990)

Chakraborty, S., Ramesh, R., Krishnaswami, S.: Air-Sea exchange of CO_2 in the Gulf of Kutch, northern Arabian Sea based on bomb-carbon in corals and tree rings. Proc. Ind. Acad. Sci. **103**, 329–340 (1994)

Chandrasekar, A., Kitoh, A.: Impact of localized sea surface temperature anomalies over the equatorial Indian Ocean on the Indian summer monsoon. J. Meteorol. Soc. Jpn. **76**, 841–853 (1998)

Chiu, T.-C., Fairbanks, R.G., Cao, L., Mortlock, R.A.: Analysis of the atmospheric ^{14}C record spanning the past 50,000 years derived from high-precision $^{230}Th/^{234}U/^{238}U$, $^{231}Pa/^{235}U$ and ^{14}C dates on fossil corals. Quat. Sci. Rev. **26**, 18–36 (2007)

Courtillot, V., Gallet, Y., Le Mouël, J.-L., Fluteau, F., Genevey, A.: Are there connections between the Earth's magnetic field and climate? Earth Planet. Sci. Lett. **253**, 328–339 (2007)

Crooks, S.A., Gray, L.J.: Characterization of the 11-year solar signal using a multiple regression analysis of the ERA-40 dataset. J. Clim. **18**, 996–1015 (2005)

Crowley, T.J.: The geologic record of climate change. Rev. Geophy. Space Phys. **21**, 828–877 (1983)

Damon, P.E., Laut, P.: Pattern of strange errors plagues solar activity and terrestrial climate data. Eos **85**, 370–374 (2004)

Dziembowski, W.A., Goode, P.R., Schou, J.: Does the Sun shrink with increasing magnetic activity? Astrophys. J. **553**, 897–904 (2001)

Eddy, J.A.: The Maunder minimum. Science **192**, 1189–1202 (1976)

Feynman, J., Ruzmaikin, A.: Modulation of cosmic ray precipitation related to climate. Geophys. Res. Lett. **26**, 2057–2060 (1999)

Field, C.V., Schmidt, G.A., Koch, D., Salyk, C.: Modeling production and climate-related impacts on ^{10}Be concentration in ice cores. J. Geophys. Res. **111**, D15107 (2006)

Fioletov, V.E., et al.: Global and zonal total ozone variations estimated from ground-based and satellite measurements: 1964–2000. J. Geophys. Res. **107**, 4647 (2002)

Fleitmann, D., Burns, S.J., Mudelsee, M., Neff, U., Kramers, J., Mangini, A., Matter, A.: Holocene forcing of the Indian monsoon recorded in a stalagmite from southern Oman. Science **300**, 1737–1739 (2003)

Foukal, P., Lean, J.: Magnetic modulation of solar luminosity by photospheric activity. Astrophys. J. **328**, 347–357 (1988)

Frank, M., Schwarz, B., Baumann, S., Kubik, P.W., Suter, M., Mangini, A.: A 200 kyr record of cosmogenic radionuclide production rate and geomagnetic field intensity from Be-10 in globally stacked deep-sea sediments. Earth Planet. Sci. Lett. **149**, 121–129 (1997)

Fröhlich, C., Lean, J.: Solar radiative output and its variability: evidence and mechanisms. Astron. Astrophys. Rev. **12**, 273–320 (2004)

Gallet, Y., Genevey, A., Fluteau, F.: Does Earth's magnetic field secular variation control centennial climate change? Earth Planet. Sci. Lett. **236**, 339–347 (2005)

Geller, M.A., Smyshlyaev, S.P.: A model study of total ozone evolution 1979–2000: The role of individual natural and anthropogenic effects. Geophys. Res. Lett. **29**, 5–8 (2002)

Godwin, H.: Half-life of radiocarbon. Nature **195**, 984 (1962)

Gough, D.O.: Theoretical prediction of variations in the solar output. In: White, O.R. (ed.) The solar output and its variation, pp. 451–473. University Press Boulder, Boulder (1977)

Gray, D.F., Livingston, W.C.: Monitoring the solar temperature: Spectroscopic temperature variations of the Sun. Astrophys. J. **474**, 802–809 (1997)

Gray, L.J., Haigh, J.D., Harrison, R.G.: Review of the influences of solar changes on the earth's climate. Hadley Centre Technical Note No. 62. pp. 82. Met Office, Exeter (2005)

Gupta, A.K., Das, M., Anderson, D.M.: Solar influence on the Indian summer monsoon during the Holocene. Geophys. Res. Lett. **32**, L17703 (2005)

Haigh, J.D.: The role of stratospheric ozone in modulating the solar radiative forcing of climate. Nature **370**, 544–546 (1994)

Haigh, J.D.: The effects of solar variability on the Earth's climate. Phil. Trans. R. Soc. Lond. Ser. A **361**, 95–111 (2003)

Harder, J., et al.: The spectral irradiance monitor I. Scientific requirements, instrument design and operation modes. Sol. Phys. **230**, 141–167 (2005)

Hiremath, K.M., Mandi, P.I.: Influence of the solar activity on the Indian monsoon rainfall. New Astron. **9**, 651–662 (2004)

Hoffman, H.J., Beer, J., Bonani, G., Von Gunten, H.R., Raman, S., Suter, M., Walker, R.L., Wolfli, W., Zimmermann, D.: ^{10}Be half-life and AMS-standards. Nucl. Instr. Meth. Phys. Res. B **29**, 32–36 (1987)

Hong, Y.T., Jiang, H.B., Liu, T.S., Zhou, L.P., Beer, J., Li, H.D., Leng, X.T., Hong, B., Qin, X.G.: Response of climate to solar forcing recorded in a 6000-year δ^{18}O time-series of Chinese peat cellulose. Holocene **10**, 1–7 (2000)

Hong, Y.T., Wang, Z.G., Jiang, H.B., Lin, Q.H., Hong, B., Zhu, Y.X., Wang, Y., Xu, L.S., Leng, X.T., Li, H.D.: A 6000-year record of changes in drought and precipitation in northeastern China based on a δ^{13}C time series from peat cellulose. Earth Planet. Sci. Lett. **185**, 111–119 (2001)

Hood, L.L.: Thermal response of the tropical tropopause region to solar ultraviolet variations. Geophys. Res. Lett. **30**, 2215 (2003)

Hoyt, D.V., Schatten, K.H.: A discussion of plausible solar irradiance variations, 1700–1992. J. Geophys. Res. **98**, 18895–18906 (1993)

IPCC: In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (eds.) Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge (2007)

Ji, J., Shen, J., Balsam, W., Chen, J., Liu, L., Liu, X.: Asian monsoon oscillations in the northeastern Qinghai–Tibet Plateau since the late glacial as interpreted from visible reflectance of Qinghai Lake sediments. Earth Planet. Sci. Lett. **233**, 61–70 (2005)

Kodera, K.: Solar influence on the Indian Ocean monsoon through dynamical processes. Geophys. Res. Lett. **31**, L24209 (2004)

Kristjánsson, J.E., Kristiansen, J., Kaas, E.: Solar activity, cosmic rays, clouds and climate – an update. Adv. Space Res. **34**, 407–415 (2004)

Kuhn, J.R., Libbrecht, K.G.: Nonfacular solar luminosity variation. Astrophys. J. Lett. **381**, L35–L37 (1991)

Lal, D., Peters, B.: Cosmic ray produced radioactivity on the Earth. In: Flügge, S. (ed.) Handbuck der physik, vol. 46(2), pp. 551–612. Springer, Berlin (1967)

Lal, D., Jull, A.J.T., Pollard, D., Vachera, L.: Evidence for large century time-scale changes in solar activity in the past 32 kyr, based on in-situ cosmogenic ^{14}C in ice at Summit, Greenland. Earth Planet. Sci. Lett. **234**, 335–349 (2005)

Laut, P.: Solar activity and terrestrial climate: An analysis of some purported correlations. J. Atmos. Sol. Terr. Phys. **65**, 801–812 (2003)

Lean, J., Beer, J., Bradley, R.: Reconstruction of solar irradiance since 1610: Implications for climate change. Geophys. Res. Lett. **22**, 3195–3198 (1995)

Lean, J., Rottman, G.J., Kyle, H.L, Woods, T.N., Hickey, J.R., Puga, L.C.: Detection and parameterization of variations in solar-mid-and-near-ultraviolet radiation (200–400 nm). J. Geophys. Res. **102**, 29939–29956 (1997)

Lean, J., Cook, J., Marquette, W., Johannesson, A.: Magnetic sources of the solar irradiance cycle. Astrophys. J. **492**, 390–401 (1998)

Lean, J.: Short-term direct indices of solar variability. Space Science Reviews **94**, 39–51 (2000)

Lean, J., Rottman, G., Harder, J., Kopp, G.: SORCE contribution to new understanding of global change and solar variability. Solar Phys. **230**, 27–53 (2005)

Lean, J., Rind, D.: Evaluating sun-climate relationships since the little ice age. J. Atmos. Sol. Terr. Phy. **61**, 25 (1999)

Leuschner, D.C., Sirocko, F.: Orbital insolation forcing of the Indian monsoon – a motor for global climate change. Palaeogeogr. Paleoclimatol. **197**, 83–95 (2003)

Libby, W.F.: Radiocarbon Dating. University of Chicago Press, Chicago (1955)

Makarova, L.N., Shirochkov, A.V.: Magnetopause position as an important index of the space weather. Phys. Chem. Earth **C 25**, 495–498 (2000)

Makarova, L.N., Shirochkov, A.V.: Long-term variability of stratospheric temperature above central Antarctica. Phys. Chem. Earth **C 27**, 449–453 (2002)

Makarova, L.N., Shirochkov, A.V.: The Sun and the solar wind variability of different time-scales and the climate dynamics. Adv. Space Res. **34**, 432–435 (2004)

Marsh, N.D., Svensmark, H.: Low cloud properties influenced by cosmic rays. Phys. Rev. Lett. **85**, 5004–5007 (2000)

Markson, R., Muir, M.: Solar wind control on the earth's electric field. Science **208**, 979–990 (1980)

Masarik, J., Beer, J.: Simulation of particle fluxes and cosmogenic nuclide production in the earth's atmosphere. J. Geophys. Res. **104**, 12099–12111 (1999)

Mayewski, P.A., Meeker, L.D., Twickler, M.S., Whitlow, S., Yang, Q., Lyons, W.B., Prentice, M.: Major features and forcing of high-latitude northern hemisphere atmospheric circulation using a 110,000 year-long glaciochemical series. J. Geophys. Res. **102**, 26345–26366 (1997)

McCormack, J.P.: The influence of the 11-year solar cycle on the quasi-biennial oscillation. Geophys. Res. Lett. **30**, 2162 (2003)

McHargue, L.R., Damon, P.E.: The global beryllium-10 cycle. Rev. Geophys. **29**, 141–158 (1991)
Mehta, V.M., Lau, K.M.: Influence of solar irradiance on the Indian monsoon-ENSO relationship at decadal-multidecadal time scales. Geophys. Res. Lett. **24**, 159–162 (1997)
Mendoza, B.: Total solar irradiance and climate. Adv. Space Res. **35**, 882–890 (2005)
Muscheler, R., Joos, F., Beer, J., Müller, S.A., Vonmoos, M., Snowball, I.: Solar activity during the last 1000 yr inferred from radionuclide records. Quat. Sci. Rev. **26**, 82–97 (2007)
Neff, U., Burns, S.J., Mudelsee, M., Mangini, A., Matter, A.: Strong coherence between solar variability and the monsoon in Oman between 9 and 6 kyrs ago. Nature **411**, 290–293 (2001)
Nesme-Ribes, E., Ferreira, E.N., Sadoumy, R., Le Treut, H., Li, Z.X.: Solar dynamics and its impact on solar irradiance and the terrestrial climate. J. Geophys. Res. **98**, 18923–18935 (1994)
Newman, M.J., Rood, R.T.: Implications of solar evolution for the Earth's early atmosphere. Science **198**, 1035 (1977)
Pallé, E.: Possible satellite perspective effects on the reported correlations between solar activity and clouds. Geophys. Res. Lett. **32**, L03802 (2001)
Ramaswamy, V., et al.: Radiative forcing of climate change. In: Houghton, J.T., et al. (eds.) Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, pp. 349–416. Cambridge University Press, Cambridge (2001)
Reid, G.C.: Influence of solar variability on global sea surface temperatures. Nature **329**, 142–143 (1987)
Rossow, W.B., Cairns, B.: Monitoring changes of clouds. J. Clim. **31**, 305–347 (1995)
Salby, M., Callaghan, P.: Evidence of the solar cycle in the general circulation of the stratosphere. J. Clim. **17**, 34–46 (2004)
Schneider, D.: Living in sunny times. Am. Sci. **93**, 22–24 (2005)
Sharma, M.: Variations in solar magnetic activity during the last 200000 years: is there a Sun-climate connection? Earth Planet. Sci. Lett. **199**, 459–472 (2002)
Siegenthaler, U., Heimann, M., Oeschger, H.: ^{14}C variations caused by changes in the global carbon cycle. Radiocarbon **22**, 177–191 (1980)
Snowball, I., Sandgren, P.: Geomagnetic field variations in northern Sweden during the Holocene quantified from varved lake sediments and their implications for cosmogenic nuclide production rates. Holocene **12**, 517–530 (2002)
Sofia, S., Fox, P.: Solar variability and climate. Clim. Change **27**, 249–257 (1994)
Sofia, S., Li, L.H.: Solar variability and climate. J. Geophys. Res. **106**, 12969–12974 (2001)
Solanki, S.K., Unruh, Y.C.: A model of the wavelength dependence of solar irradiance variations. Astron. Astrophys. **329**, 747–753 (1998)
Solanki, S.K., Fligge, M.: A reconstruction of total solar irradiance since 1700. Geophys. Res. Lett. **26**, 2465–2468 (1999)
Solanki, S.K, Krivova, N.A.: Can solar variability explain global warming since 1970. J. Geophys. Res. **108**, 1200–1207 (2003)
Solanki, S.K., Usoskin, I.G., Kromer, B., Schüssler, M., Beer, J.: Unusual activity of the Sun during recent decades compared to the previous 11,000 years. Nature **431**, 1084–1087 (2004)
Staubwasser, M., Sirocko, F., Grootes, P.M., Erlenkeuser, H.: South Asian monsoon climate change and radiocarbon in the Arabian Sea during early and middle Holocene. Paleoceanography **17**, 1063 (2002)
St-Onge, G., Stoner, J.S., Hillaire-Marcel, C.: Holocene Paleomagnetic records from the St. Lawrence Estuary, eastern Canada: Centennial-to millennial-scale geomagnetic modulation of cosmogenic isotopes. Earth Planet. Sci. Lett. **209**, 113–130 (2003)
Stuiver, M., Polach, H.A.: Discussion: reporting of ^{14}C data. Radiocarbon **19**, 355–363 (1977)
Stuiver, M., Quay, P.D.: Changes in atmospheric carbon-14 attributed to a variable sun. Science **207**, 11–19 (1980)
Svensmark, H., Friis-Christensen, E.: Variation of cosmic ray flux and global cloud coverage -a missing link in solar-climate relationships. J. Atmos. Sol. Terr. Phys. **59**, 1225–1232 (1997)
Svensmark, H.: Influence of cosmic rays on earth's climate. Phys. Rev. Lett. **81**, 5027–5030 (1998)

Thamban, M., Kawahata, H., Rao, V.P.: Indian summer monsoon variability during the Holocene as recorded in sediments of the Arabian Sea: Timing and implications. J. Oceanogr. **63**, 1009–1020 (2007)

Tiwari, M., Ramesh, R., Somayajulu, B.L.K., Jull, A.J.T., Burr, G.S.: Solar control of Southwest Monsoon (SWM) on centennial time scales. Curr. Sci. **89**, 1583–1588 (2005)

von Rad, U., Schaaf, M., Michels, K. H., Schulz, H., Berger, W. H., Sirocko, F.: A 5000-yr record of climate change in varved sediments from the oxygen minimum zone off Pakistan, Northeastern Arabian Sea. Quat. Res. **51**, 39–53 (1999)

Wagner, G., Laj, C., Beer, J., Kissel, C., Muscheler, R., Masarik, J., Synal, H.-A.: Reconstruction of the paleoaccumulation rate of central Greenland during the last 75 kyr using the cosmogenic radionuclides ^{36}Cl and ^{10}Be and geomagnetic field intensity data. Earth Planet. Sci. Lett. **193**, 515–521 (2001)

Wang, Y.M., Lean, J.L., Sheeley, N.R.: Modeling the sun's magnetic field and irradiance since 1713. Astrophys. J. **625**, 522–538 (2005a)

Wang, Y., Cheng, H., Edwards, R.L., He, Y., Kong, X., An, Z., Wu, J., Kelly, M.J., Dykoski, C.A., Li, X.: The Holocene Asian monsoon: Links to solar changes and North Atlantic climate. Science **308**, 854–857 (2005b)

Willson, R.C.: Total solar irradiance trend during solar cycles 21 and 22. Science **277**, 1963–1965 (1997)

Willson, R.C., Mordvinov, A.V.: Secular total solar irradiance trend during solar cycles 21–23. Geophys. Res. Lett. **30**, 1199 (2003)

Woods, T.N., et al.: Validation of the UARS solar ultraviolet irradiances: comparison with the ATLAS 1 and 2 measurements. J. Geophys. Res. **101**, 9541–9569 (1996)

Yadava, M.G., Ramesh, R.: Significant longer-term periodicities in the proxy record of the Indian monsoon rainfall. New Astron. **12**, 544–555 (2007)

The Planetary X-ray Emission

Anil Bhardwaj

Abstract With the advent of Chandra and XMM-Newton X-ray observatories, several new solar system bodies have been discovered to be an X-ray source at energies below 2 keV: leading to significant advances in the field of planetary X-ray emission and astronomy during the last one decade. Apart from the Sun, the known X-ray emitters now include Venus, Earth, Mars, Jupiter, and Saturn, planetary satellites (Moon, Io, Europa, and Ganymede), all active comets, the Io plasma torus, the rings of Saturn, the exospheres of Earth, Venus and Mars, and the heliosphere. Progress in modeling X-ray emission and laboratory studies of X-ray production have also greatly contributed in enhancing our understanding of X-rays from the solar system bodies. This chapter provides an overview of the planetary X-rays, their production source processes and important characteristics.

1 Introduction

Soon after the first detection of solar X-rays in 1949, the first planetary X-rays detected were terrestrial X-rays in the 1950s. In 1962, the foremost attempt to detect X-rays from the Moon failed, but it resulted in the birth of the field of X-ray astronomy through the discovery of the first extra-solar source, the Scorpius X-1. Thus, X-ray astronomy is one of the youngest branches of astronomy.

Generally, X-ray emission is thought to originate from hot collisional plasma such as the million-degree gas residing in the solar and stellar coronae, where X-rays are mainly produced by thermal bremsstrahlung and by excitation of highly stripped ions. However, in the solar system X-rays have been observed from rather cold planetary and cometary atmospheres and surfaces, thus requiring an external source to power the production. This makes the field of planetary X-rays an interesting

A. Bhardwaj
Space Physics Laboratory, Vikram Sarabhai Space Centre, Trivandrum 695022, India
e-mail: Anil_Bhardwaj@vssc.gov.in

N. Gopalswamy et al. (eds.), *Heliophysical Processes*, Astrophysics and Space Science Proceedings, DOI 10.1007/978-3-642-11341-3_16,

discipline, where X-rays are produced from a wide variety of phenomena and under a broad range of conditions (cf. Bhardwaj et al. 2007; Bhardwaj and Lisse 2007).

Launch of the first X-ray satellite UHURU in 1970 marked the beginning of the satellite-based X-ray astronomy. Apollo 15 and 16 missions, in early 1970s, studied fluorescently scattered X-rays from the Moon. In 1979, the Einstein observatory discovered, after a long search (see Bhardwaj and Gladstone 2000 for history), X-rays from Jupiter (Metzger et al. 1983). Till 1990, the three objects known to emit X-rays were Earth, Moon and Jupiter. In 1996, Rontgensatellit (ROSAT) made an important contribution to the field of planetary X-rays by discovering X-ray emission from a comet (Lisse et al. 1996). This discovery revolutionized the field of solar system X-rays and highlighted the importance of the solar wind charge exchange (SWCX) mechanism (Cravens 1997, 2002) in the production of X-rays in the solar system.

The field of planetary X-ray astronomy has grown rapidly over the past two decades and today it is a very dynamic and at the forefront of new research. This has been possible due to observations made by satellite-based orbiting observatories – starting with Einstein, followed by ROSAT, and in the new millennium, by the more sophisticated X-ray observatories, viz., Chandra and XMM-Newton, and the legacy is continued by Suzaku (Astro-E2) and Swift. This has resulted in several new solar system objects being now known as X-ray emitters at energies generally below 2 keV (cf. Table 1). Apart from the Sun, the known X-ray emitters (cf. Fig. 1) now include planets – Venus, Earth, Mars, Jupiter, and Saturn; planetary satellites – the Moon, Io, Europa, and Ganymede; all active comets, the Io plasma torus, the rings of Saturn, the coronae (exospheres) of Earth, Mars and Venus, and the heliosphere.

This chapter discusses the production mechanisms and characteristics of the X-ray emission from the solar system bodies. Readers are referred to recent reviews for more detailed and different perspectives (e.g., Bhardwaj and Gladstone 2000; Cravens 2002; Lisse et al. 2004; Krasnopolsky et al. 2004; Bhardwaj 2006; Bhardwaj et al. 2009; Bhardwaj and Lisse 2007; Bhardwaj et al. 2007; Dennerl 2009). This chapter is largely based on the review of Bhardwaj et al. (2007).

Table 1 Classes of solar system objects detected in X-rays

Class	Object(s)
Star	Sun
Planets	Venus, Earth, Mars, Jupiter, Saturn
Satellites	Moon, Io, Europa, Ganymede, [Titan][a]
Minor bodies	Comets, Asteroids
Planetary coronae (exospheres)	Geocorona, Martian and Venusian Exospheres
Extended objects (around planets)	Io Plasma Torus, Rings of Saturn
Large cavity	Heliosphere

[a]X-rays from Titan have not been observed, but in a rare celestial event captured by the Chandra X-ray Observatory on January 5, 2003, Titan passed in front of the Crab Nebula. The X-ray shadow cast by Titan allowed astronomers to make the first X-ray measurement of the extent of its atmosphere. Adopted from Bhardwaj et al. (2007)

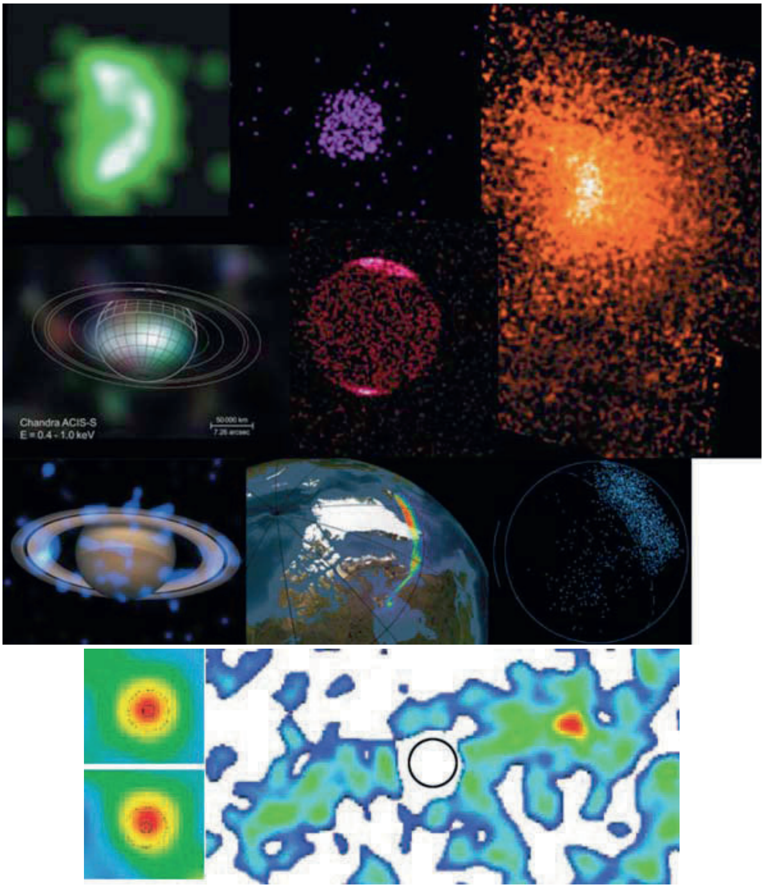


Fig. 1 Chandra montage of solar system X-ray sources. Clockwise, from *upper left*: Chandra images of Venus, Mars, comet C/Ikeya–Zhang 2001, Jupiter, and Saturn. *Middle panel*: *left*, Saturn rings; *middle*, Earth; and *right*, Moon. *Bottom panel*: *left top*, Io; *left bottom* Europa; and *right*, Io plasma torus

2 Processes of Planetary X-ray Production

Several physical mechanisms are responsible for the production of X-ray emission from planets. These will be briefly described here; reviews covering some aspects of this topic can also be consulted (Krasnopolsky et al. 2004; Lisse et al. 2004; Cravens 2002; Bhardwaj et al. 2007). The major mechanisms that produce X-rays in solar system bodies include: (1) collisional excitation of neutral species and ions by charged particle impacts (particularly electrons) followed by line emission, (2) electron collisions with neutrals and ions producing continuum bremsstrahlung emission, (3) solar photon scattering from neutrals in planetary atmospheres – both elastic scattering and K-shell fluorescent scattering, (4) charge exchange of solar

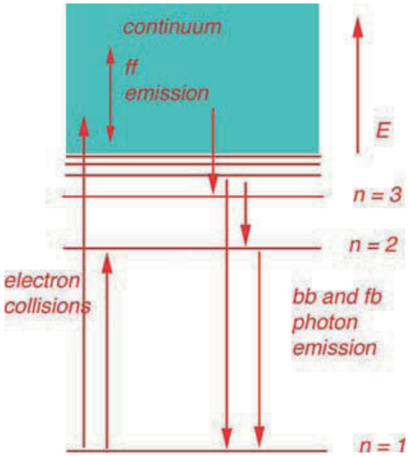


Fig. 2 Schematic of energy levels for a hydrogen-like atom. Freefree (ff), boundfree (fb), and boundbound (bb) transitions are indicated

wind ions (energies of about 1 keV/amu) with neutrals, followed by X-ray emission, and (5) X-ray production from the charge exchange of energetic (energies of about a MeV/amu) heavy ions of planetary magnetospheric origin with neutrals or by direct excitation of ions in collisions with neutrals.

X-ray emission can be understood by considering a highly-excited atom or ion. Figure 2 is a schematic showing the electron energy levels for a hydrogenic atomic species. The energies for a species with nuclear charge, Z, and just one electron are given by the Bohr energy expression: $E_n = -Z^2\,13.6\,\text{eV}/n^2$, where n is the principal quantum number. The ground-state energy for atomic hydrogen ($Z = 1$) is -13.6 eV (this is also the ionization potential), whereas for O^{7+} ions ($Z = 8$), the ground-state energy ($n = 1$ level) is 64 times this, or -870 eV. The energy levels for a multi-electron atom/ion are not so easily described but the X-ray emission processes are the same in their essential features. The types of transitions include boundbound (denoted bb), boundfree (bf), and freefree (ff). For example, collisions with fast electrons can excite an atom from the ground-state to an excited state. This would then be followed by the emission of one, or more, photons in the form of line emission. For example, the $n = 2$ to $n = 1$ transition in H produces Lyman alpha photons with energies of 10.2 eV ($\lambda = 121.6$ nm – ultraviolet radiation), whereas the same transition for O^{7+} ions produces 653 eV photons ($\lambda = 1.9$ nm – soft X-ray radiation).

3 Electron Collision: Line and Bremsstrahlung Emission

Energetic electrons impacting neutrals and ions can produce X-rays. Fast electrons can collide with the target species and excite them to either bound states or to the continuum (resulting in ionization). The former can lead to line emission, including X-ray emission as occurs at the solar corona and in Earth's auroral regions (i.e., characteristic inner-shell line emissions of nitrogen (Kα at 0.393 keV), oxygen (Kα at 0.524 keV) and argon (Kα at 2.958 keV, Kβ at 3.191 keV)). However, freefree collisions will produce continuum X-ray emission (i.e., the bremsstrahlung process). According to the thick-target bremsstrahlung theory the energy of the X-ray photon produced in this process has any energy up to the energy of the incoming particle, but the probability distribution increases exponentially at lower energies. The X-ray bremsstrahlung production efficiency depends on the square of the deceleration and thus is proportional to $1/m^2$ where m is the mass of the precipitating particle. This implies that electrons are 10^6 times more efficient than protons at producing X-ray bremsstrahlung. The production efficiency is a non-linear function of energy, with increasing efficiency for increasing incident energy. For example, for a 200 keV electron the probability of producing an X-ray photon at any energy below 200 keV is 0.5%, while the probability for a 20 keV electron to produce an X-ray photon below 20 keV is only 0.0057%.

Bremsstrahlung is an important process for the formation of the X-ray continuum of the Sun, and it also explains the X-ray emission in the terrestrial aurora, and the Jovian auroral X-rays at energies >2 keV (Singhal et al. 1992; Branduardi-Raymont et al. 2007). For the auroral cases, the fast electrons responsible are produced externally to the planetary atmosphere in the magnetosphere and then precipitate along magnetic field lines into the atmosphere. In a "thermal plasma", such as the solar corona, in which the electron energy distribution is Maxwellian, "thermal bremsstrahlung" radiation is produced. Thermal bremsstrahlung is important in the Sun and it might also be important for explaining some of the emission from the hot plasma in the Io Plasma Torus (Elsner et al. 2002).

4 Solar Photon Scattering and Fluorescence from Planetary Atmospheres and Surfaces

X-rays can be both absorbed and elastically scattered (both incoherently and coherently) by atoms or molecules in an atmosphere. In particular, solar X-rays can be scattered from planetary atmospheres, which act, in effect, as diffuse mirrors. In the soft X-ray part of the spectrum, scattering cross sections are much smaller than absorption cross sections, and only a small fraction of incident solar X-rays will reflect from the target atmosphere. However, planetary X-rays produced by this process have been observed. On Earth, Jupiter and Saturn scattered X-rays have been observed both during major solar flares as well as during non-flaring conditions (cf. Bhardwaj et al. 2007; Bhardwaj 2006).

The absorption of X-rays, usually beyond the K-shell edge, can also result in X-ray emission. In the K-shell fluorescence process, ionization from the K-shell leaves a vacancy and an X-ray photon is emitted when a valence electron makes a transition to fill this vacancy. However, the excess energy is usually taken up by the emission of an Auger electron rather than a photon. For oxygen, the photon yield is only about 0.2% in the soft X-ray part of the spectrum. K-shell fluorescence from carbon (found in atmospheric methane for Jupiter or in carbon dioxide for Venus or Mars) makes a minor contribution to the disk emission of the outer planets, but is the dominant X-ray source at Venus and Mars, for which carbon dioxide is the major neutral species (Dennerl 2009; Dennerl et al. 2006; Bhardwaj et al. 2007). L-shell fluorescence can also occur, but the radiative yields are much lower than for the K shell. Theoretical calculations have demonstrated that solar scattering and K-shell fluorescence are not important as X-ray sources for comets (Krasnopolsky et al. 2004). The reason is that to obtain a unit optical depth for X-ray absorption requires a neutral column density of $\approx 10^{20}\,\mathrm{cm}^{-2}$ (the inverse of the total cross section) that is easy to obtain in a planetary atmosphere, but not in the more tenuous cometary atmospheres.

The fluorescence mechanism can also operate when solar X-ray photons are absorbed by solid surfaces, such as the surface of the Moon, the surfaces of the Galilean satellites or ring particles at Saturn. The ultimate energy source for this emission derives from the solar photons (and, hence, the solar corona) rather than the relatively low-temperature surfaces themselves.

5 Charge Exchange of Highly Ionized Heavy Solar Wind Ions

Another X-ray production mechanism important for many solar system environments is the solar wind charge exchange (SWCX) mechanism (cf. Cravens 1997, 2002; Krasnopolsky et al. 2004; Lisse et al. 2004; Bhardwaj et al. 2007). X-rays are generated by ions left in excited states after charge transfer collisions between highly ionized solar wind ions and target neutrals. Such solar wind ions are found as minor, heavy ions that account for about 0.1% of the solar wind and exist in highly charged states such as: O^{7+}, O^{6+}, C^{6+}, C^{5+}, N^{6+}, Ne^{8+}, Si^{9+}, and Fe^{12+}. The source of the solar wind is the million degree solar corona, which is the ultimate source of power for the X-ray emission in the SWCX mechanism. This mechanism was first proposed to explain the surprising ROSAT observations of soft X-ray and extreme ultraviolet emission from comet Hyakutake in 1996 (Cravens 1997). Since then, the SWCX mechanism has also been shown to operate in the heliosphere, in the terrestrial magnetosheath (geocoronal emission), and at Mars and Venus exospheres (cf. Bhardwaj et al. 2007; Dennerl 2009).

6 Charge Exchange and Direct Collisional Excitation of Very Energetic Heavy Ions

X-rays have been observed from the high-latitude auroral regions of Jupiter. The harder (>2 keV) X-rays are largely produced by electron bremsstrahlung, but most of the observed X-ray power is in softer X-rays with energies less than 1 keV (e.g., Bhardwaj and Gladstone 2000; Bhardwaj et al. 2007; Branduardi-Raymont et al. 2007). The explanation for this X-ray emission is the precipitation of energetic ($\sim$1 MeV/nucleon) heavy ions into the Jovian polar atmosphere.

For this mechanism, ambient ions (such as O^+ and S^+) in Jupiter's outer magnetosphere are accelerated to high energies and precipitate into the upper atmosphere. Collisions of relatively low-charge ion species with a neutral atmosphere (e.g., H_2 or H in Jupiter's atmosphere) will not produce X-rays because the transitions are in the ultraviolet part of the spectrum rather than the X-ray part. However, at energies of a few hundred keV/nucleon or more (e.g., Kharchenko et al. 2008), the ions can undergo electron removal (or stripping) collisions and become more highly-charged. The cross section for this process remains high for ion energies down to as low as a few hundred keV per nucleon. X-rays are then emitted from these highly charged ions primarily by charge transfer collisions. Direct excitation of heavy ion transitions in collisions might also be possible.

7 Summary

At Jupiter and Earth, both auroral and non-auroral disk X-ray emissions have been observed. X-rays have been detected from Saturn's disk, but no convincing evidence of the X-ray aurora has been observed. The first soft (<2 keV) X-ray observation of Earth's aurora by Chandra shows that it is highly variable. The non-auroral X-ray emissions from atmospheres of Jupiter, Saturn, and Earth, and those from Mars' and Venus' atmospheres, and Moon's surface, are mainly produced due to scattering of solar X-rays. The spectral characteristics of the X-ray emission from comets, heliosphere, geocorona, and Martian and Venusian exosphere are quite similar: they all are produced by the SWCX mechanism, but they appear to be quite different from those of Jovian auroral X-rays.

Table 2 summaries our current knowledge of soft X-ray emission from solar system objects. All the solar system X-ray emission (except that from the Sun) is powered by one or more of the three sources: (1) solar X-ray photons, (2) highly ionized heavy ions of the solar wind, and (3) planetary magnetospheric energetic plasma (electrons and ions).

These sources produce X-ray emission on a variety of solar system bodies via mechanisms described in previous section. The first source produces X-rays via resonant and fluorescent scattering of incident solar X-rays from an atmosphere or the surface. The second source produces X-rays in charge exchange collisions with neutrals (SWCX). The third source generates X-rays via precipitation of the energetic plasma from planetary magnetosphere into the atmosphere or onto the surface. In the first case, the incident X-ray photons are basically recycled (except for a change in the wavelength). However, in the second and third cases new X-ray photons are generated. In the second source process, the production of X-rays reduces the charge state of the incident solar wind ions. In the third source process, the conversion of the energy of the precipitating particles into X-ray photons provides diagnostics of the energetic plasma species and their energy in the planetary magnetosphere.

Table 2 Summary of the characteristics of soft X-ray emission from solar system bodies[a]

Object	Emitting region	Power emitted[b]	Special characteristics	Production mechanism
Earth	Auroral atmosphere	10–30 MW	Correlated with magnetic storm and substorm activity	Bremsstrahlung from precipitating electrons, and characteristic line X-rays from atmospheric neutrals due to precipitating electron impact
	Non-auroral atmosphere	40 MW	Correlated with solar X-ray flux	Fluorescent scattering of solar X-rays by atmospheric gases
Moon	Dayside	0.07 MW	Correlated with solar X-rays	Scattering and fluorescence due to solar X-rays by the surface elements on dayside
	Geocoronal (nightside)		Nightside emissions are ∼1% of the dayside missions	SWCX with geocorona
Venus	Sunlit atmosphere	50 MW	Emissions from ∼120 to 140 km above the surface	Fluorescent scattering of solar X-rays by C and O atoms in the atmosphere
	Exosphere	1–5 MW	Emissions confined largely within 1.5 Venus radii	SWCX with Venusian exosphere
Mars	Sunlit atmosphere	1–4 MW	Emissions from ∼110 to 130 km above the surface	Fluorescent scattering of solar X-rays by C and O atoms in the upper atmosphere
	Exosphere	1–10 MW	emissions extend out to ∼8 Mars radii	SWCX with Martian corona
Jupiter	Auroral atmosphere	0.4–1 GW	Pulsating (∼20–60 min) X-ray hot spot in north and south polar regions	Energetic ion precipitation from magnetosphere and/or solar wind + electron bremsstrahlung
	Non-auroral atmosphere	0.5–2 GW	relatively uniform over disk	Mainly scattering of solar X-rays + possible ion precipitation from radiation belts
Saturn	Sunlit disk	0.1–0.4 GW	Varies with solar X-rays	Scattering of solar X-rays (+ Electron bremsstrahlung?)

(Continued)

Table 2 (Continued)

Comets	Sunward-side coma	0.2–1 GW	Intensity peaks in sunward direction $\sim 10^5$–10^6 km ahead of cometary nucleus	SWCX with cometary neutrals
Io Plasma Torus	Plasma torus	0.1 GW	Dawn-dusk asymmetry observed	Electron bremsstrahlung + ?
Io	Surface	2 MW	Emissions from upper few microns of the surface	Energetic Jovian magnetospheric ions impact on the surface
Europa	Surface	1.5 MW	Emissions from upper few microns of the surface	Energetic Jovian magnetospheric ions impact on the surface
Rings of Saturn	Surface	30–80 MW	Emissions confined to a narrow energy band around at 0.53 keV	Fluorescent X-ray emission from atomic oxygen in H_2O ice excited by incident solar X-rays + ?
Asteroid	Sunlit surface		Emission vary with solar X-ray flux	Fluorescent X-ray emission from elements in the surface excited by incident solar X-rays
Heliosphere	Entire heliosphere	10^{16} W	Emissions vary with solar wind variation	SWCX with heliospheric neutrals

SWCX = charge exchange of heavy highly ionized solar wind ions with neutrals
[a]Taken mainly from Bhardwaj et al. (2007) and Bhardwaj and Lisse (2007)
[b]The values quoted are those at the time of observation. X-rays from all bodies are expected to vary with time. For comparison the total X-ray luminosity from the Sun is 10^{20} W

At Earth auroral X-rays have been used to estimate the characteristics of the precipitating particles (mainly electrons) in the auroral regions (cf. Bhardwaj et al. 2007). At Jupiter auroral X-rays have provided constraints on the energetics of the precipitating particles (mainly ions) and their composition (cf. Kharchenko et al. 2008). At atmosphere-less bodies X-rays (mainly fluorescently scattered solar X-rays) have been used to derive elemental composition of the surface (e.g., on the Moon and asteroids).

The solar wind charge-exchange process operates at almost all bodies in the solar system having neutrals that can interact with solar wind ions. Due to its high cross section this process is very efficient at bodies with a tenuous atmosphere or exosphere. What happens in individual cases then depends on the heliospheric distance and on the local conditions. For example, even though Venus is closer to the Sun, the Venusian exospheric X-ray emission is weaker than that of Mars because of the smaller extent of the Venusian exosphere. This process is also expected to be operating in the exospheres of Jupiter and Saturn, but the corresponding fluxes would be too weak to detect and distinguish from the background X-ray emission.

Mainly driven by SWCX, cometary X-rays provide an observable link between the solar corona, where the solar wind originates, and the solar wind where the comet resides. Thus, comets can be used as probes to measure the solar wind throughout the heliosphere. This will be especially useful in monitoring the solar wind in places that are hard to reach with spacecraft – such as over the solar poles, at large distances above and below the ecliptic plane, and at heliocentric distances greater than a few AU. For example, $\sim 1/3$ of the observed soft X-ray emission from comets is found in the 530–700 eV oxygen O^{+7} and O^{+6} lines; observing photons of this energy will allow studies of the oxygen ion charge ratio of the solar wind, which is predicted to vary significantly between the slow and fast solar wind. Further, the solar wind velocity might be inferred from cometary X-ray spectra and images (cf. Bhardwaj et al. 2007).

The Martian and Venusian exospheric X-rays provide another possibility to infer charge state ratios and other solar wind properties near the planet. Also, the study of exospheric X-rays is important for Mars, particularly because charge exchange interaction between atmospheric constituents and solar wind ions is considered as a vital non-thermal escape mechanism that may be responsible for a significant loss of the Martian atmosphere. Thus, X-ray observation of Mars provides a novel method for studying exospheric processes on a global scale, and may lead to a better understanding of the state of the Martian atmosphere and its evolution.

Another interesting fact that has emerged from studying planetary X-rays is that the upper atmospheres of planets (Mars, Venus, Jupiter, and Saturn) act as "diffuse mirrors" that backscatter incident solar X-rays. Thus, X-rays from planets might be used as potential remote-sensing tools to monitor X-ray flaring on portions of the hemisphere of the Sun facing away from near-Earth space weather satellites. Also, in the longer run, measured X-ray intensities from the planetary disk regions could be used to constrain both the absolute intensity and the spectrum of solar X-rays. Unfortunately, none of the currently planetary orbiting satellite missions, Mars Express (Mars), Venus Express (Venus), Cassini (Saturn), have any kind of X-ray instrument onboard.

Although, we now have a broad understanding of the X-ray production mechanisms at most solar system bodies, there are several interesting questions that need to be answered (see Bhardwaj et al. 2007 for details), which need longer systematic observations from established X-ray observatories, viz., Chandra, XMM-Newton, and from those more recently commissioned, viz., Suzaku (earlier known as Astro-E2), and Swift. Next generation observatories (like the proposed Constellation-X and XEUS), that will have larger collecting areas and a better sensitivity, will also help significantly to advance the study of solar system X-rays.

With a fleet of satellites in (or going to be in) the lunar orbit (e.g., Selene (Kauyga), Chang'e, Chandrayaan-1, LRO), the Moon is the focus of X-ray studies mainly for determining elemental composition and its variations at higher spatial resolution. Similarly, X-rays from Mercury will be studied for the first time by the XRS (X-ray Spectrometer) aboard NASA's Messenger spacecraft during its flyby of the planet in 2008–2009. The Messenger XRS will provide information on the elemental composition of Mercury's surface by observing the Kα lines of the elements present, lines that are induced by solar X-rays as well as by high-energy electron precipitation. Such X-ray measurements will continue after the insertion of Messenger in the Hermean orbit in 2011, and are likely to be followed up by investigations with the Bepi-Colombo mission to Mercury. These types of observations provide a map of the composition at airless bodies, which reveal critical information concerning their origin and evolution.

Several other solar system bodies, which include Uranus, Neptune, Titan, and the inner-icy satellites of Saturn, are also expected to be X-ray sources, but are probably too faint to be detected even with modern X-ray observatories. Like Jupiter and Saturn, Uranus and Neptune ought to be X-ray sources due to the scattering and fluorescence of solar X-rays and due to a particle precipitation in the polar regions. An attempt to detect X-rays from Uranus by Chandra on August 7, 2002 during a 30 ks exposure time has been unsuccessful. Titan could be an X-ray source due to scattered solar X-rays and a particle precipitation from Saturn's magnetosphere. At times when Titan is outside Saturn's magnetosphere, X-rays from Titan could be produced by the SWCX process. Similar to the Galilean satellites, the X-ray emission from the satellites of Saturn could be produced by the precipitation of magnetospheric plasma onto their surfaces.

References

Bhardwaj, A., Gladstone, G.R.: Auroral emissions of the giant planets. Rev. Geophys. **38**, 295–353 (2000)

Bhardwaj, A.: X-ray emission from Jupiter, Saturn, and Earth: a short review. In: Bhardwaj A., et al. (eds.) Advances in Geosciences, vol. 3, pp. 215–230. World Scientific, Singapore (2006)

Bhardwaj, A., Lisse, C.M.: X-rays in the solar system. In: McFadden, L.A., Weisman, P.R., Johnson, T.V. (eds.) Encyclopedia of the Solar System, 2nd edn., pp. 637–685. Academic Press, New York (2007)

Bhardwaj, A., et al.: X-rays from solar system objects. Planet. Space Sci. **55**, 1135–1189 (2007)

Bhardwaj, A., et al.: X-ray emission from planets and comets: relationship with solar x-rays and solar wind. In: Bhardwaj A., et al. (eds.) Advances in Geosciences, vol. 15, pp. 229–244. World Scientific, Singapore (2009)

Branduardi-Raymont, G., Bhardwaj, A., Elsner, R.F., Gladstone, G.R., Ramsay, G., Rodriguez, P., Soria, R., Waite, J.H. Jr, Cravens, T.E.: A study of Jupiter's aurorae with XMM-Newton. Astron. Astrophys. **463**, 761–774 (2007)

Cravens, T.E.: Comet Hyakutake x-ray source: charge transfer of solar wind heavy ions. Geophys. Res. Lett. **24**, 105–109 (1997)

Cravens, T.E.: X-ray emission from comets. Science. **296**, 1042–1046 (2002)

Dennerl, K., Lisse, C.M., Bhardwaj, A., Burwitz, V., Englhauser, J., Gunell, H., Holmström, M., Jansen, F., Kharchenko, V., Rodríguez-Pascual, P.M.: First observation of Mars with XMM-Newton: high resolution x-ray spectroscopy with RGS. Astron. Astrophys. **451**, 709–722 (2006)

Dennerl, K.: X-rays from non-magnetic planets. In: Bhardwaj, A., et al. (eds.) Advances in Geosciences, vol. 15, pp. 53–74. World Scientific, Singapore (2009)

Elsner, R.F., Gladstone, G.R., Waite, J.H. Jr, Crary, F.J., Howell, R.R., Johnson, R.E., Ford, P.G., Metzger, A.E., Hurley, K.C., Feigelson, E.D., Garmire, G.P., Bhardwaj, A., Grodent, D., Majeed, T., Tennant, A.F., Weisskopf, M.C.: Discovery of soft x-ray emission from Io, Europa, and the Io Plasma Torus. Astrophys. J. **572**, 1077–1082 (2002)

Kharchenko, V., Bhardwaj, A., Dalgarno, A., Schultz, D., Stancil, P.: Spectra of the north and south Jovian x-ray auroras. J. Geophys. Res. **113**, A08229 (2008)

Krasnopolsky, V.A., Greenwood, J.B., Stancil, P.C.: X-ray and extreme ultraviolet emission from comets. Space Sci. Rev. **113**, 271–374 (2004)

Lisse, C.M., et al.: Discovery of x-ray and extreme ultraviolet emission from comet C/Hyakutake 1996 B2. Science. **274**, 205–209 (1996)

Lisse, C.M., Cravens, T.E., Dennerl, K.: X-ray and extreme ultraviolet emission from comets. In: Festou, M.C., Keller, H.U., Weaver, H.A. (eds.) Comet II, pp. 631–643. University Arizona Press, Tucson (2004)

Metzger, A.E., Gilman, D.A., Luthey, J.L., Hurley, K.C., Schnopper, H.W., Seward, F.D., Sullivan, J.: The detection of x-rays from Jupiter. J. Geophys. Res. **88**, 7731 (1983)

Singhal, R.P., Chakravarty, S.C., Bhardwaj, A., Prasad, B.: Energetic electron precipitation in Jupiter's upper atmosphere. J. Geophys. Res. **97**, 18245–18256 (1992)