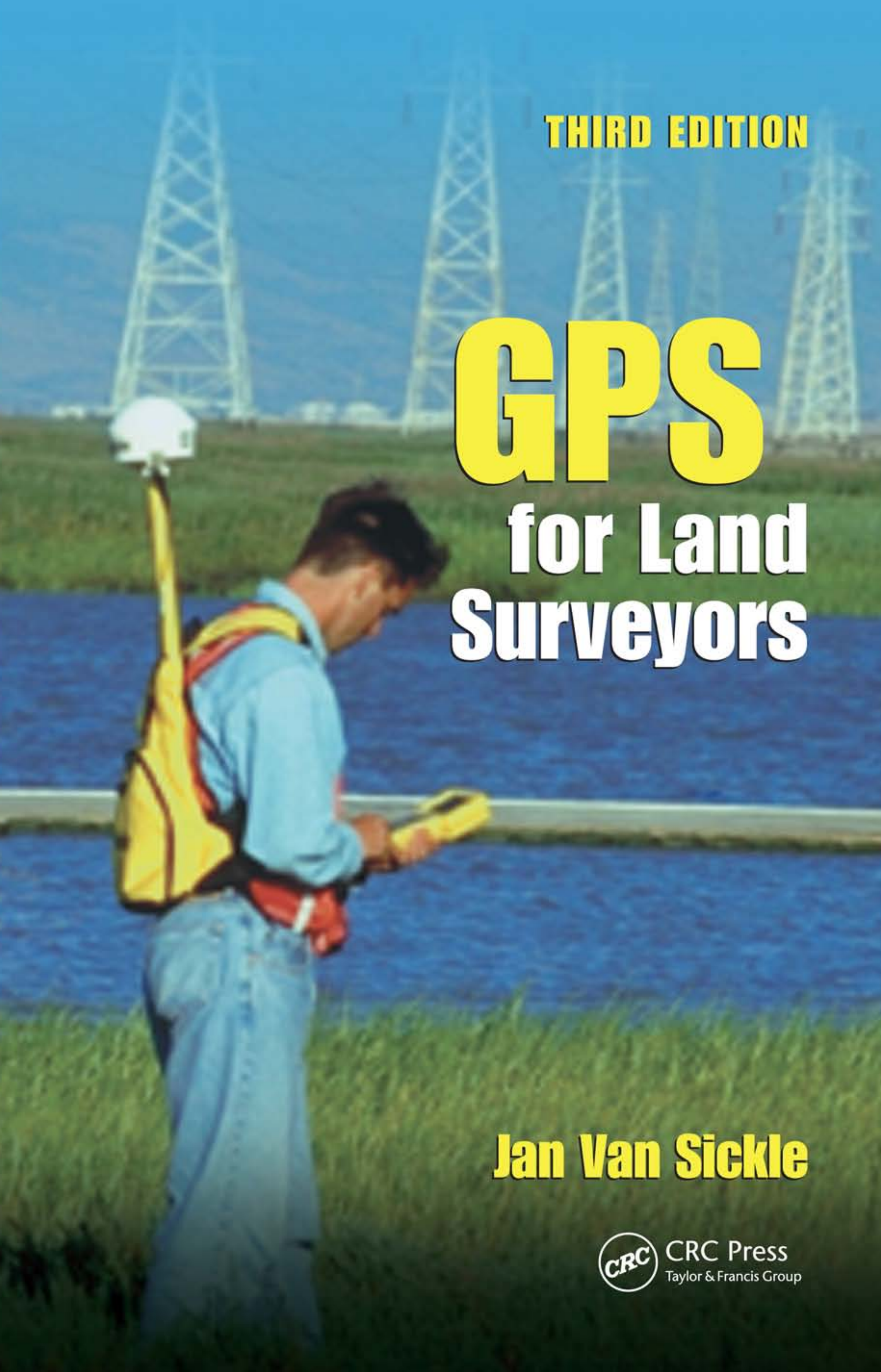


THIRD EDITION

GPS for Land Surveyors

Jan Van Sickle

CRC Press
Taylor & Francis Group

THIRD EDITION

GPS
for Land Surveyors

THIRD EDITION

GPS
for Land Surveyors

Jan Van Sickle

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2008 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20110725

International Standard Book Number-13: 978-0-203-30522-5 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface to the Third Edition xix
About the Author xxi

Chapter 1 The GPS Signal 1

Global Positioning System (GPS) Signal Structure 1
  GPS and Trilateration 1
    A Passive System 1
    Time 1
    Control 2
  The Navigation Code 2
    Wavelength 3
    Codes 3
    GPS Time 4
    Satellite Clocks 5
    GPS Week 5
    Julian Date 6
    The Broadcast Ephemeris 6
    Atmospheric Correction 7
    Antispoofing and Coordinated Universal Time (UTC) 7
    The Almanac 7
    Satellite Health 7
  The P and Coarse/Acquisition (C/A) Codes 8
    Pseudorandom Noise (PRN) 8
    P Code 8
    C/A Code 8
    Standard Positioning Service (SPS) and Precise Positioning Service (PPS) 8
  The Production of a Modulated Carrier Wave 9
    EDM Ranging 9
    GPS Ranging 9
    Oscillators 11
  A Chain of Electromagnetic Energy 11
    Phase Shift 12
    The Cycle Ambiguity Problem 14
Two Observables 15
    Encoding by Phase Modulation 15
Pseudoranging 17
    Propagation Delay 17
    Code Correlation 17

Autocorrelation 18
Correlation Peak 19
Lock and the Time Shift 21
Imperfect Oscillators 21
A Pseudorange Equation 21
The One-Percent Rule of Thumb 23
Carrier Phase Ranging 24
Carrier Phase Comparisons 24
Phase Difference 24
Beat 25
The Doppler Effect 26
GPS and the Doppler Effect 26
A Carrier Phase Approximation 26
An Illustration of the Cycle Ambiguity Problem 29
Exercises 30
Answers and Explanations 32

Chapter 2 Biases and Solutions 35

The Error Budget 35
A Look at the Biases in the Observation Equations 35
The Biases 35
User Equivalent Range Error 35
The Satellite Clock Bias, dt 35
Relativistic Effects on the Satellite Clock 36
Satellite Clock Drift 36
The Ionospheric Effect, d_{ion} 37
Group and Phase Delay 37
Total Electron Content 37
Ionosphere and the Sun 38
Ionospheric Gradients 38
Satellite Elevation and Ionospheric Effect 38
The Magnitude of the Ionospheric Delay 39
The Ionosphere Affects Codes and the Carrier Differently 39
Different Frequencies Are Affected Differently 39
Broadcast Correction 40
The Receiver Clock Bias, dT 40
Typical Receiver Clocks 40
The Orbital Bias, d_ρ 40
Forces Acting on the Satellites 40
Tracking Facilities 41
The Tropospheric Effect, d_{trop} 42
Multipath 43
Limiting the Effect of Multipath 45
Antenna Design and Multipath 45
Receiver Noise 45

Differencing 46
  Classifications of Positioning Solutions 46
    Kinematic GPS 46
    Static GPS 47
    Hybrid Techniques in GPS 47
    Relative and Point Positioning 47
    The Navigation Solution 47
    Four Unknowns 47
    Four Satellites and Four Equations 48
  Relative Positioning 49
    Correlation 49
    Baselines 49
    Networks 49
    Differencing 49
  Single Difference 49
    Elimination of the Satellite Clock Errors 49
    Other Errors Remain 51
  Double Difference 51
    Elimination of the Receiver Clock Errors 51
    No Clock Errors at All 51
  Triple Difference 51
    Integer Cycle Ambiguity 51
  Summary 56
    Typical Techniques 56
    Pseudorange and Carrier Phase 56
    Kinematic GPS 56
    Other Techniques 56
    Differencing 56
    Biases 57
Exercises 57
Answers and Explanations 60

Chapter 3 The Framework 63

Technological Forerunners 63
  Consolidation 63
  Terrestrial Radio Positioning 63
    Radar 63
    Distance by Timing 63
    Shoran Surveying 64
    Hiran Surveying 64
    Sputnik 64
    Satellite Advantages 64
  Optical Systems 65
    Triangulation with Photographs 65

        - Laser Ranging ... 65
        - Optical Drawbacks ... 66
    - Extraterrestrial Radio Positioning ... 66
        - Satellite Tracking ... 66
        - Prime Minitrack ... 66
        - Very Long Baseline Interferometry ... 66
- Very Long Baseline Interferometry ... 67
- Transit ... 68
        - The Doppler Shift ... 68
        - TRANSIT Shortcomings ... 68
        - TRANSIT and GPS ... 69
        - Linking Datums ... 69
- Navstar GPS ... 69
        - Orbits and Clocks ... 69
        - Increased Accuracy ... 69
        - Military Application ... 69
        - Secure, Passive, and Global ... 70
        - Expense and Frequency Allocation ... 70
        - Large Capacity Signal ... 70
        - The Satellite Constellation ... 71
        - The Perfect System? ... 71
    - GPS in Civilian Surveying ... 71
        - Federal Specifications ... 72
        - Interferometry ... 72
        - Civil Applications of GPS ... 72
- GPS Segment Organization ... 73
    - The Space Segment ... 73
- GPS Constellation ... 73
    - Orbital Period ... 73
        - Design ... 74
        - Dilution of Precision ... 74
    - Bad Dilution of Precision ... 74
    - Good Dilution of Precision ... 76
        - Outages ... 76
        - Satellite Positions in Mission Planning ... 76
        - Satellite Names ... 78
        - Block I ... 79
        - Block II and Block IIA ... 79
        - Block IIR ... 79
        - Block IIF ... 79
    - Block II GPS Satellite ... 80
        - Signal Deterioration ... 80
        - GPS Satellites ... 80
- The Control Segment ... 81
        - Master Control Station ... 81
        - Other Stations ... 81

        Kalman Filtering 82
        An Analogy 82
        Constant Tracking 82
        Postcomputed Ephemerides 82
        GPS for Orbit Calculation 83
    The User Segment 83
        Constantly Increasing Application of GPS 83
Exercises 84
Answers and Explanations 86

Chapter 4 Receivers and Methods 91

Common Features of GPS Receivers 91
    A Block Diagram of a Code Correlation Receiver 91
        Receivers for GPS Surveying 91
    The Antenna 92
        Bandwidth 93
        Nearly Hemispheric Coverage 93
        Antenna Orientation 94
        Height of Instrument 94
    The Radio Frequency (RF) Section 94
        Channels 95
        Multiplexing and Sequencing 96
        Tracking Loops 96
        Pseudoranging 96
        Carrier Phase Measurement 97
        Carrier Tracking Loop 97
        Doppler Shift 97
    Typical GPS Doppler Shift 97
        The Typical Change in the Doppler Shift 98
        Continuously Integrated Doppler 98
        Integer Ambiguity 99
        Signal Squaring 99
    The Microprocessor 100
        Differential Positioning 100
        Differential GPS 101
        Positional Accuracies 101
        Kinematic and Real-Time Kinematic 101
    The Control and Display Unit (CDU) 101
        Typical Displays 101
    The Storage 102
        Downloading 102
    The Power 102
        Battery Power 102
Choosing a GPS Receiver 103
    Trends in Receiver Development 104

        High Accuracy Early 104
        Another Direction 104
        More Convenience 104
        Multichannel and Code-Correlating 105
        Dual-Frequency 105
        Adding Codeless Capability 105
        Adding P Code Tracking 105
        Typical GPS Surveying Receiver Characteristics 106
        GPS Receiver Costs 106
Some GPS Surveying Methods 106
    Static 106
        Prerequisites for Static GPS 106
        Productivity 108
        Session Length 108
        Large Amounts of Data 108
        Resolution of the Cycle Ambiguity 108
        Preprogrammed Observations 108
        Observation Settings 109
        Data Interval 109
        Compatible Receivers 109
        Receiver Capabilities and Baseline Length 109
    Differential GPS 110
    Kinematic 110
        Reference Receiver and Rover Receivers 110
        Leapfrog Kinematic 110
        Kinematic Positional Accuracy 110
        Maintaining Lock 111
        Reconnaissance 111
        Applications 111
        Most Carrier Phase Receivers Capable of Kinematic GPS 111
        Wideband 111
        Practical Considerations in Kinematic GPS 112
    Pseudokinematic 112
        Double Occupation 112
        No Need for Continuous Lock 112
        Best Used in Easy Access Situations 113
        Mostly Radial Surveys 113
    Rapid-Static 113
        Wide Laning 113
    On-the-Fly 114
        Accurate Initial Positions 114
        Photogrammetry without Ground Control 114
    Real-Time Kinematic (RTK) 114
Exercises 115
Answers and Explanations 117

Chapter 5 Coordinates 121

A Few Pertinent Ideas About Geodetic Datums for GPS 121
  Plane Surveying 121
    Development of State Plane Coordinate Systems 121
    GPS Surveyors and Geodesy 122
  Some Geodetic Coordinate Systems 122
    Three-Dimensional (3-D) Cartesian Coordinates 122
    Polar Motion 122
    The Conventional Terrestrial System (CTS) in GPS 123
    Latitude and Longitude 123
  Elements of a Geodetic Datum 124
    The Deflection of the Vertical 124
  Geocentric, Geodetic, and Astronomic Latitude 125
    Datums 127
    Development of the Ellipsoidal Model 127
  Biaxial Ellipsoidal Model of the Earth 128
    The Role of an Ellipsoid in a Datum 128
  Regional Ellipsoids 130
    Measurement Technology and Datum Selection 130
  Position Derived from GPS 131
    The Development of a Geocentric Model 131
  The Geoid 132
    An Equipotential Surface 133
    Geoidal Undulation 133
  The Modern Geocentric Datum 133
    World Geodetic System 1984 (WGS84) 134
  North American Datum 1983 135
    NAD27 135
    The Development of the North American Datum 1983 (NAD83) 135
    The International Terrestrial Reference System (ITRS) 136
    ITRF00, WGS84, and NAD83 136
    The Management of NAD83 137
    Transformations from NAD27 to NAD83 137
    Densification and Improvement of NAD83 138
    High Accuracy Reference Networks 138
    Height Modernization and Base Networks 139
    Continuously Operating Reference Stations 140
State Plane Coordinates 140
  NAD83 Positions and Plane Coordinates 140
  Map Projection 141
  Distortion 143
    Decreasing Distortion 144
  Secant and Cylindrical Projections 144
    Secant Projections 144
    Choices 145

State Plane Coordinate System (SPCS) Map Projections 147
SPCS27 to SPCS83 148
Changes in Zones 149
State Plane Coordinates Scale and Distance 150
Geodetic Lengths to Grid Lengths 150
Universal Transverse Mercator Coordinates 158
Universal Transverse Mercator (UTM) Zones of the World 158
Heights 162
Ellipsoidal Heights 162
Orthometric Heights 163
Spirit Leveling 163
Evolution of a Vertical Datum 164
Sea Level 164
Diurnal Tide 165
A Different Approach 166
The Zero Point 167
The Geoid 167
Geoid Models 169
Exercises 171
Answers and Explanations 173

Chapter 6 GPS Surveying Techniques 177

Static GPS Surveying 177
Planning 177
New Standards 177
New Design Criteria 177
The Lay of the Land 178
Maps 178
National Geodetic Survey (NGS) Control 178
NGS Control Data Sheets 178
Survey Order and Class 181
Coordinates 181
The Station Mark 182
Significance of the Information 182
Control from Continuously Operating Networks 182
NGS Continuously Operating Reference Stations 183
NGS CORS Reference Points 183
NGS CORS Precise Orbits 184
International Global Navigation Satellite System (GNSS) Service (IGS) 184
Project Design 184
Horizontal Control 184
Station Location 185
Vertical Control 185

- Preparation 186
  - Plotting Project Points 186
  - Evaluating Access 187
  - Planning Offsets 188
  - Planning Azimuth Marks 188
  - Obtaining Permissions 188
- Some GPS Survey Design Facts 189
  - Software Assistance 189
  - Position Dilution of Precision (PDOP) 189
  - Polar Plot 190
  - An Example 190
  - Choosing the Window 192
  - Ionospheric Delay 192
  - An Example 192
  - Naming the Variables 194
- Drawing the Baselines 194
  - Horizontal Control 194
  - Julian Day in Naming Sessions 195
  - Independent Lines 197
  - Redundancy 198
  - Federal Geodetic Control Committee (FGCC) Standardsfor Redundancy 199
  - Forming Loops 199
  - Finding the Number of Sessions 199
  - Ties to the Vertical Control 202

Real-Time Kinematic (RTK) and Differential GPS (DGPS) 202
- The General Idea 202
- Radial GPS 203
  - The Correction Signal 204

DGPS 204
- Real-Time 204
- Local and Wide Area DPGS 205
  - Maritime DGPS 206
  - Wide Area Augmentation System (WAAS) 206
- Latency 206
- Identical Constellation 207
- Geographic Information Systems (GIS) Applications for DGPS 207

RTK 208
- Fixing the Integer Ambiguity in RTK 208
- Radio License 210
- Cell Phones 211
- Typical RTK 211
- The Vertical Component in RTK 212
- Some Practical RTK Suggestions 212
  - Typical Satellite Constellations 212
  - Dual-Frequency Receiver 213

    - Setting up a Base Station ... 213
    - After the RTK Survey ... 213
  - Comparing RTK and DGPS ... 213
    - Multipath in RTK and DGPS ... 213
    - Initialization ... 214
    - Base Station ... 214
    - Radio Technical Commission for Maritime (RTCM) Services Version 3 ... 215
- Real-Time Network Services ... 215
  - Precise Point Positioning ... 217
  - Summary ... 218
- Exercises ... 219
- Answers and Explanations ... 223

Chapter 7 Observing and Processing ... 229

- Static GPS Control Observations ... 229
  - Equipment ... 229
    - Conventional Equipment ... 229
    - Safety Equipment ... 230
    - Communications ... 230
    - GPS Equipment ... 231
    - Auxiliary Equipment ... 232
    - Papers ... 232
  - Station Data Sheet ... 232
    - Station Name ... 234
    - Rubbings ... 235
    - Photographs ... 235
    - Quad Sheet Name ... 235
    - To-Reach Descriptions ... 235
    - Flagging and Describing the Monument ... 236
  - Visibility Diagrams ... 236
    - An Example ... 236
    - Drawing Obstructions ... 237
    - Working around Obstructions ... 238
    - Approximate Station Coordinates ... 238
    - Multipath ... 240
    - Point Offsets ... 240
    - Look for Multipath ... 240
  - Monumentation ... 241
  - Logistics ... 241
    - Scheduling ... 241
  - Observation ... 242
    - Arrival ... 242
    - Setup ... 242
    - Height of Instrument ... 243

        Observation Logs 243
        Weather 245
    Daily Progress Evaluation 245
Real-Time Kinematic (RTK) and Differential GPS (DGPS) Observations 245
        Point Offsets 246
        Dynamic Lines 247
        Planning 247
    A Few RTK Procedures 248
    Site Calibration 248
Processing 249
    Postprocessing GPS Static Control Surveys 249
        Correlation of Biases 250
    Quantity of Data 250
        Organization Is Essential 250
        File Naming Conventions 250
    Downloading 250
        Making Room 251
    Control 251
    The First Position 251
    Triple Difference 252
        Components of a Triple Difference 252
    Double Difference 252
        The Integer Ambiguity 252
        The Float Solution 253
        The Fixed Solution 253
    Cycle Slip Detection and Repair 253
        Cycle Slip Causes 254
        Repairing Cycle Slips 254
    Fixing the Integer Ambiguity and Obtaining Vector Solutions 255
    Least-Squares Adjustment 256
    Using a Processing Service 257
Exercises 257
Answers and Explanations 261

Chapter 8 GPS Modernization and Global Navigation Satellite System (GNSS) 267

GPS Modernization 267
    Block I, Block II, and Block IIR Satellites 267
    Power Spectral Density Diagrams 268
    L1 Signal 269
    L2 Signal 269
    New Signals 269
    The M-Code 270
    L2C 271
        Civil-Moderate (CM) and Civil-Long (CL) 272

        CNAV 272
        Multiplexing 273
        Phase-Locked Loop 273
    Practical Advantages 273
L5 274
    The L5 Carrier 274
    The Block IIF Satellites 276
    Summary of Coarse Acquisition (C/A), L2C, and L5 276
    New Signal Availability 276
        Availability 276
        Ionospheric Bias 278
        Correlation Protection 278
    Another Civil Signal—L1C 278
Global Navigation Satellite System (GNSS) 279
GLONASS 280
    GLONASS Uragan M and K 280
        GLONASS Constellation 281
    GLONASS Signals 282
        Code Division Multiple Access (CDMA) and Frequency Division Multiple Access (FDMA) 282
        Changes to FDMA 284
    GLONASS Time 284
    GALILEO 284
    GIOVE-A and GIOVE-B 285
        GALILEO Signals and Services 285
    Interoperability 286
    GALILEO Signals 287
        Frequency Coincidence 287
BEIDOU/COMPASS 288
    Frequencies 288
    The Quasi-Zenith Satellite System (QZSS) 289
The Future 289
    More Satellites 289
        Accessibility 290
        Flexibility 290
        Reliability 290
        Faster Positioning 290
        Faster Initialization 290
        GPS Accuracy 290
        GNSS Accuracy—Faster 291
        Simplification 292
Interoperability 292
    GPS–GALILEO–GLONASS Constellations 292
        Inconsistency 292

        Consistency 292
        Robust Solutions 292
    Summary 293
Exercises 294
Answers and Explanations 296

References 299

Glossary 301

Index 325

Preface to the Third Edition

Global Positioning System (GPS) itself is now a part of a larger international context—the Global Navigation Satellite System, *GNSS*. The changes in GPS and GNSS hardware, software, and procedures are accelerating. Keeping up is a real challenge. My hope in offering this third edition of *GPS for Land Surveyors* is to contribute some small assistance in that process.

This book has been written to find a middle ground. It is intended to be neither simplistic nor overly technical, but an introduction to the concepts needed to understand and use GPS and GNSS, not a presentation of the latest research in the area. An effort has been made to explain the progression of the ideas at the foundation of satellite positioning and delve into some of the particulars.

This is a practical book, a guide to some of the techniques used in GPS and GNSS, from their design through observation, processing, real-time kinematic (RTK), and real-time networks. Today, some of the aspects of satellite navigation are familiar and some are not. This book is about making them all familiar.

About the Author

Jan Van Sickle, P.L.S., has more than 40 years experience as a land surveyor. He was privileged to supervise surveys with the first commercially available GPS receivers in the early 1980s and has worked with GPS since. He received his M.Eng. from the University of Colorado in 2006. He is the author of *GPS for Land Surveyors*, *Basic GIS Coordinates*, and *1001 Solved Surveying Fundamentals Problems*. The latter is quoted in the feature Surveying Solutions each month in *POB* magazine. He taught the advanced surveying program at the Denver Institute of Technology and GPS seminars throughout the United States. He is currently a senior lecturer at Pennsylvania State University. Van Sickle is a licensed surveyor in California, Colorado, Oregon, North Dakota, and Texas.

1 The GPS Signal

GLOBAL POSITIONING SYSTEM (GPS) SIGNAL STRUCTURE

GPS AND TRILATERATION

GPS can be compared to trilateration. Both techniques rely exclusively on the measurement of distances to fix positions. One of the differences between them, however, is that the distances, called ranges in GPS, are not measured to control points on the surface of the earth. Instead, they are measured to satellites orbiting in nearly circular orbits at a nominal altitude of about 20,183 km above the earth.

A Passive System

The ranges are measured with signals that are broadcast from the GPS satellites to the GPS receivers in the microwave part of the electromagnetic spectrum; this is sometimes called a passive system. GPS is passive in the sense that only the satellites transmit signals; the users simply receive them. As a result, there is no limit to the number of GPS receivers that may simultaneously monitor the GPS signals. Just as millions of television sets may be tuned to the same channel without disrupting the broadcast, millions of GPS receivers may monitor the satellite's signals without danger of overburdening the system. This is a distinct advantage, but as a result, GPS signals must carry a great deal of information. A GPS receiver must be able to gather all the information it needs to determine its own position from the signals it collects from the satellites.

Time

Time measurement is essential to GPS surveying in several ways. For example, the determination of ranges, like distance measurement in a modern trilateration survey, is done electronically. In both cases, distance is a function of the speed of light, an electromagnetic signal of stable frequency and elapsed time. In a trilateration survey, frequencies generated within an electronic distance measuring device, EDM, can be used to determine the elapsed travel time of its signal because the signal bounces off a reflector and returns to where it started. But the signals from a GPS satellite do not return to the satellite; they travel one way, to the receiver. The satellite can mark the moment the signal departs and the receiver can mark the moment it arrives. And since the measurement of the ranges in GPS depends on the measurement of the time it takes a GPS signal to make the trip, the elapsed time must be determine by decoding the GPS signal itself.

Control

Both GPS surveys and trilateration surveys begin from control points. In GPS the control points are the satellites themselves; therefore, knowledge of the satellite's position is critical. Measurement of a distance to a control point without knowledge of that control point's position would be useless. It is not enough that the GPS signals provide a receiver with information to measure the range between itself and the satellite. That same signal must also communicate the position of the satellite, at that very instant. The situation is complicated somewhat by the fact that the satellite is always moving with respect to the receiver at an along track speed of approximately 4 km per second. In a GPS survey, as in a trilateration survey, the signals must travel through the atmosphere. In a trilateration survey, compensation for the atmospheric effects on the EDM signal, estimated from local observations, can be applied at the signal's source. This is not possible in GPS. The GPS signals begin in the virtual vacuum of space, but then, after hitting the earth's atmosphere, they travel through much more of the atmosphere than EDM signals.

Therefore, the GPS signals must give the receiver some information about needed atmospheric corrections.

It takes more than one measured distance to determine a new position in a trilateration survey or in a GPS survey. Each of the several distances used to define one new point must be measured to a different control station. For trilateration, three distances are adequate for each new point. For a GPS survey, the minimum requirement is a measured range to each of at least four GPS satellites.

Just as it is vital that every one of the three distances in a trilateration is correctly paired with the correct control station, the GPS receiver must be able match each of the signals it tracks with the satellite of its origin. Therefore, the GPS signals themselves must also carry a kind of satellite identification. To be on the safe side, the signal should also tell the receiver where to find all of the other satellites as well.

To sum up, a GPS signal must somehow communicate to its receiver: (1) what time it is on the satellite, (2) the instantaneous position of a moving satellite, (3) some information about necessary atmospheric corrections, and (4) some sort of satellite identification system to tell the receiver where it came from and where the receiver may find the other satellites.

THE NAVIGATION CODE

How does a GPS satellite communicate all that information to a receiver? It uses codes. Codes are carried to GPS receivers by carrier waves. A carrier wave has at least one characteristic such as phase, amplitude, or frequency that may be changed or *modulated*, to carry information. For example, the information, music, or speech received from an AM radio station is placed on the carrier wave by amplitude modulation, and the information on the signal from an FM radio station is there because of frequency modulation. The two GPS carriers are electromagnetic waves too. They come from a part of the L-band. The L-band is a designation that includes the ultrahigh frequencies from approximately 390 megahertz (MHz) to 1550 MHz. Actually, the so-called L1 GPS frequency is a bit higher than the strict L-band definition, as you will see.

Wavelength

A wavelength with duration of 1 second, known as 1 cycle per second, is said to have a frequency of 1 hertz (Hz) in the International System of Units (SI). A frequency of 1 Hz is rather low. The lowest sound human ears can detect has a frequency of about 25 Hz. The highest is about 15,000 hertz, or 15 kilohertz (KHz). Most of the modulated carriers used in EDM and all those in GPS instruments have frequencies that are measured in units of a million cycles per second, or megahertz, MHz. At the moment two fundamental frequencies assigned to GPS are called L1 at 1575.42 MHz and L2 at 1227.60 MHz, and another carrier is coming. It is known as L5 at 1176.45 MHz.

Codes

GPS codes are binary, zeroes and ones, the language of computers. Currently, the three basic codes in GPS are the precise code, or P code, the coarse/acquisition (C/A) code, and the Navigation code. There are a few related secondary codes, and new codes will implemented in the future modernization of GPS. Among these new codes will be the M code and the L2C code.

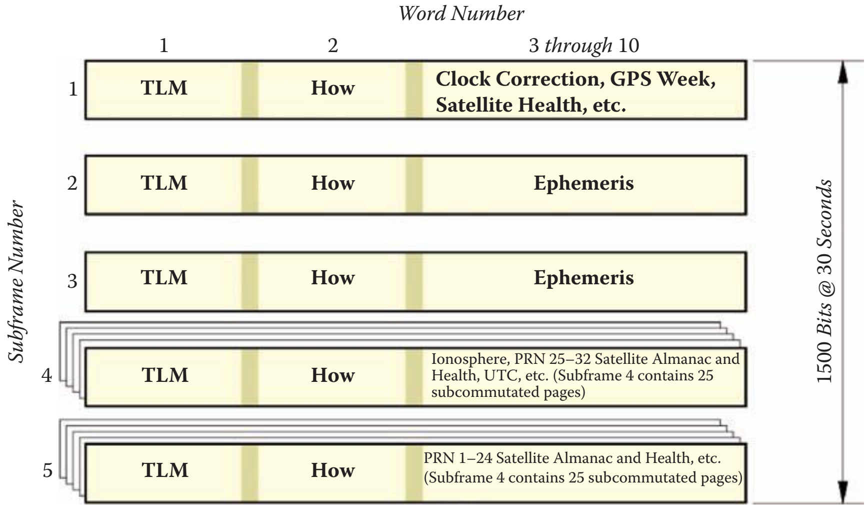
The Navigation code, or message, is the vehicle for telling the GPS receivers some of the most important things they need to know. Here are some of the parameters of its design.

The Navigation code has a low frequency, 50 Hz. It is modulated onto the GPS carriers. It communicates a stream of data called the GPS message, or Navigation message (Figure 1.1). The entire Navigation message, the Master Frame, contains 25 frames. Each frame is 1500 bits long and is divided into five subframes. Each subframe contains 10 words and each word is comprised of 30 bits. Therefore, the entire Navigation message contains 37,500 bits and at a rate of 50 bits-per-second takes 12½ minutes to broadcast and to receive.

Frame to frame, the information contained in subframes 1, 2, and 3 does not change; however, the information in subframes 4 and 5 does. These two subframes 4 and 5 are comprised of 25 subcommutated pages. That means that each of these pages of subframe 4 and 5 contain different information and a receiver must collect a total all 25 frames before it has all of the data. The receiver gets the first of the 25 pages on the first time through the five subframes. It gets the second of the 25 pages on the second time through the subframes and so on.

The first two words in every subframe are the telemetry (TLM) word and the handover word (HOW). The first 8 bits in the telemetry word contain synchronization information. They are utilized by the receiver to integrate itself with the Navigation message and decode it. The handover word contains the truncated Z-count. The Z-count is one of the primary units for GPS time and can be considered a counter that is incremented at the beginning of each consecutive subframe. It provides the receiver with the instant at the leading edge of the next data subframe.

Unfortunately, the accuracy of some aspects of the information included in the Navigation message deteriorates with time. Translated into the rate of change in the three-dimensional position of a GPS receiver, it is about 4 cm per minute. Therefore,



Each subframe has ten 30*-bit Words,* 6 *Seconds. The entire Navigation Message. The Master Frame, contains* 37,500 *bits—*12½ *min to broadcast and receive at* 50 *bps.*

FIGURE 1.1 Navigation Message

mechanisms are in place to prevent the message from getting too old. The message is renewed once or twice each day by government uploading facilities around the world or by the satellites themselves as is the case of the Block IIR satellites. Here the word block is a designation used to classify the generations of GPS satellites and the R stands for replenishment, more about that later. These installations, along with their tracking and computing counterparts, are known collectively as the Control Segment (see Chapter 3).

GPS Time

Examples of time sensitive information are found in subframe 1 and subframe 4 of the Navigation message. In fact, subframe 4 contains information that helps a receiver relate UTC to GPS Time. Using a standard time scale called GPS Time, the message contains information needed by the receiver to correlate its clock with that of the clock of the satellite. But the constantly changing time relationships in GPS can only be partially defined in these subframes. It takes more than a portion of the Navigation message to define those relationships. In fact, the rate of the GPS Time standard is defined outside the system.

The **rate** of GPS Time is kept within 1 microsecond, and usually less than 25 nanoseconds, of the **rate** of the worldwide time scale, which is called Coordinated Universal Time (UTC) as determined by the U.S. Naval Observatory (USNO). The rate of UTC, which is steered by timing laboratories around the world, is more stable than the rotation of the earth itself. This causes a discrepancy between UTC and the earth's actual motion. This difference is kept within 0.9 seconds by the periodic introduction of leap seconds in UTC. But since GPS is not earthbound, leap seconds are not used in GPS Time. GPS Time was identical to UTC on midnight, January 6, 1980. Since then many leap seconds have been added to UTC but none have been added to GPS Time. This complicates the relationship between UTC and GPS Time. Even though their rates are virtually identical, the numbers expressing a particular instant in GPS Time are different by some seconds from the numbers expressing the same instant in UTC. For example, GPS time was 14 seconds ahead of UTC on January 1, 2006.

Information in subframe 4 includes the relationship between GPS time and UTC, and it also notes future scheduled leap seconds. In this area subframe 4 can accommodate 8 bits, 255 leap seconds, which should suffice until about 2330. Subframe 1 contains some information on the health of the satellite, an estimate of the signal dispersion due to the ionosphere, and the GPS week number, more about that in "Satellite Health," later in this section. But most importantly, subframe 1 contains coefficients that are used to estimate the difference between the satellite clock and GPS Time.

Satellite Clocks

Each GPS satellite carries its own onboard clocks, also known as time standards, in the form of very stable and accurate atomic clocks regulated by the vibration frequencies of the atoms of two elements. Two of the onboard clocks are regulated by cesium and two are regulated by rubidium, except on the Block IIR satellites, where all the clocks are rubidium. There has been some consideration of equipping the upcoming Block IIF satellites with hydrogen maser clocks which will have even higher precision than the rubidium clocks.

Since the clocks in any one satellite are completely independent from those in any other, they are allowed to drift up to one millisecond from the strictly controlled GPS Time standard. Instead of constantly tweaking the satellite's onboard clocks to keep them all in lockstep with each other and with GPS Time, their individual drifts are carefully monitored by the government tracking stations of the Control Segment. These stations record each satellite clock's deviation from GPS Time, and the drift is eventually uploaded into subframe 1 of each satellite's Navigation message, where it is known as the broadcast clock correction.

A GPS receiver may relate the satellite's clock to GPS Time with the correction given in the broadcast clock correction of its Navigation message. This is obviously only part of the solution to the problem of directly relating the receiver's own clock to the satellite's clock. The receiver will need to rely on other aspects of the GPS signal for a complete time correlation.

The drift of each satellite's clock is not constant. Nor can the broadcast clock correction be updated frequently enough to completely define the drift. Therefore, one of the 10 words included in subframe 1 provides a definition of the reliability of the broadcast clock correction. This is called IODC, or Issue of Data Clock.

GPS Week

As mentioned earlier, subframe 1 contains information on the GPS week. It is worthwhile to mention that in GPS weeks are counted consecutively. The first GPS week, GPS week 0, began at 00 hour UTC on January 6, 1980 and ended on January 12, 1980. It was followed by GPS week 1, GPS week 2, and so on. However, about 19 years later, at the end of GPS week 1023—August 15 through August 21, 1999—it was necessary to start the numbering again at 0.

This necessity accrued from the fact that the following week, August 22 through August 28, 1999, would have been GPS week 1024. That was a problem because the capacity of the GPS week field in the Navigation message was only 10-bits, and the largest count a 10-bit field can accommodate is 2^{10} or 1024. Therefore, near midnight on August 21, 1999, actually 23:59:47 UTC to accommodate leap seconds, the GPS week consecutive numbering began again at GPS week 0. To alleviate this problem in the future, the modernized Navigation messages has a 13-bit field for the GPS week count. This means that the GPS week will not need to roll over again for a long time.

Julian Date

Here is a little more concerning dates. Most practitioners of GPS use the term *Julian date* to mean the day counted consecutively from January 1 of the current year (even though Julian dates, taken literally, should be counted from January 1, 4713 B.C.). For convenience most GPS receivers will also display dates in the more commonly used Gregorian format.

Another example of time-sensitive information is found in subframes 2 and 3 of the Navigation Message. They contain information about the position of the satellite, with respect to time. This is called the satellite's ephemeris.

The Broadcast Ephemeris

The satellite's ephemeris provides the information necessary to communicate its position relative to the earth. The ephemeris is given in a right ascension (RA) system of coordinates. There are six orbital elements; among them are the semimajor axis of the orbit, a, and the eccentricity of the orbit, e. These two elements define the shape and the size of the orbit. However, the orientation of the orbital plane in space is defined by other things, specifically the right ascension of its ascending node, Ω, and the inclination of its plane, i. These parameters along with the argument of the perigee, ω, and the description of the position of the satellite on the orbit, known as the true anomaly, provides all the information the user's computer needs to calculate earth-centered, earth-fixed, World Geodetic System 1984, GPS Week 1150 (WGS84 [G1150]) coordinates of the satellite at any moment. The broadcast ephemeris, however, is far from perfect.

The broadcast ephemeris is expressed in parameters that appear Keplerian. Its elements are named for the seventeenth century German astronomer Johann Kepler. But in this case, they are the result of least-squares, curve-fitting analysis of the satellite's actual orbit. Therefore, like the broadcast clock correction, accuracy of the broadcast ephemeris deteriorates with time. As a result, one of the most important parts of this portion of the Navigation message is called IODE. IODE is an acronym that stands for Issue of Data Ephemeris, and it appears in both subframes 2 and 3.

Atmospheric Correction

Subframe 4 addresses atmospheric correction. As with subframe 1, the data offer only a partial solution to a problem. The Control Segment's monitoring stations find the apparent delay of a GPS signal caused by its trip through the ionosphere through an analysis of the different propagation rates of the two frequencies broadcast by all GPS satellites, L1 and L2. These two frequencies and the effects of the atmosphere on the GPS signal will be discussed later. For now, it is sufficient to say that a single-frequency receiver depends on the ionospheric correction in subframe 4 to help remove part of the error introduced by the atmosphere.

Antispoofing and Coordinated Universal Time (UTC)

Subframe 4 contains a flag that tells the receiver when a security system, known as antispoofing, also known as simply AS, has been activated by the government ground control stations. Since December of 1993, the P code on all Block satellites has been encrypted to become the more secure Y code. Antispoofing in this context refers to a countermeasure against the malicious broadcast of a signal masquerading as the GPS signal.

The Almanac

Subframes 4 and 5 tell the receiver where to find all the GPS satellites. This information is sometimes called the almanac. While subframe 4 contains the almanac data for satellites with pseudorandom noise (PRN) numbers from 25 through 32, subframe 5 contains almanac data for satellites with PRN numbers from 1 through 24. More about PRN numbers in a moment. Nevertheless here, these ephemerides are not complete. Once the receiver finds its first satellite, it can look at these truncated ephemerides (the almanac) of that satellite's Navigation message to figure the position of more satellites to track. But since the navigation message repeats every 12.5 min, the process of downloading the files typically takes from 12.5 to 25 min. Then to collect any particular satellite's entire ephemeris, a receiver must acquire that particular satellite's signal and collect its specific Navigation message.

Satellite Health

Subframe 1 contains information about the health of the satellite the receiver is tracking when it receives the Navigation message and allows it to determine if the satellite is operating within normal parameters. Subframes 4 and 5 include health data all of the satellites.

GPS satellites are vulnerable to a wide variety of breakdowns, particularly clock trouble. That is one reason some of them carry as many as four clocks. Health data are also periodically uploaded by the ground control. These subframes inform users of any satellite malfunctions before they try to use a particular signal.

As mentioned before, each of these five subframes begins with the same two words: the telemetry word (TLM) and the handover word (HOW). Unlike nearly everything else in the Navigation message, these two words are generated by the satellite itself.

The TLM indicates the status of uploading from the Control Segment while it is in progress. It also has a constant unchanging 8-bit preamble of 10001011, and a string helps the receiver reliably find the beginning of each subframe.

The HOW provides the receiver information on the time of the GPS week and the number of the subframe, among other things. For example, the HOW's Z count tells the receiver exactly where the satellite stands in the generation of positioning codes. In fact, the handover word actually helps the receiver go from tracking the C/A code to tracking the P code, the primary GPS positioning codes.

THE P AND COARSE/ACQUISITION (C/A) CODES

Like the Navigation message, the P and C/A codes are designed to carry information from the GPS satellites to the receivers. They, too, are impressed on the L1 and L2 carrier waves by modulation. However, unlike the Navigation message, the P and C/A codes are not vehicles for broadcasting information that has been uploaded by the Control Segment. They carry the raw data from which GPS receivers derive their time and distance measurements.

Pseudorandom Noise (PRN)

The P and C/A codes are complicated; so complicated, in fact, that they appear to be nothing but noise at first. And even though they are known as pseudorandom noise, or PRN codes, actually, these codes have been carefully designed. They have to be. They must be capable of repetition and replication.

P Code

For example, the P code generated at a rate of 10.23 million bits per second is available on both L1 and L2. Each satellite is assigned a portion of the very long P code and it repeats its portion of the P code every 7 days. The entire code is renewed every 37 weeks. All GPS satellites broadcast their codes on the same two frequencies, L1 and L2 for now, though a new frequency L5 will be implemented, more about that in Chapter 8. Still, a GPS receiver must somehow distinguish one satellite's transmission from another. One method used to facilitate this satellite identification is the assignment of one particular week of the 37-week-long P code to each satellite. For example, space vehicle 14 (SV 14) is so named because it broadcasts the fourteenth week of the P code.

C/A Code

The C/A code is generated at a rate of 1.023 million bits per second, 10 times slower than the P code. Here satellite identification is quite straightforward. Not only does each GPS satellite broadcast its own completely unique 1023 bit C/A code, it repeats its C/A code every millisecond. The C/A code is broadcast on L1 only.

Standard Positioning Service (SPS) and Precise Positioning Service (PPS)

The C/A code is the vehicle for the SPS, which is used for most civilian surveying applications. It will soon be joined by a new civilian signal on L2 known as L2C.

The P code on the other hand provides the same service for PPS. The idea of SPS and PPS was developed by the Department of Defense. SPS was designed to provide a minimum level of positioning capability considered consistent with national security, ±100m, 95% of the time, when intentionally degraded through Selective Availability (SA).

The intentional dithering of the satellite clocks by the Department of Defense called Selective Availability was instituted in 1989. The accuracy of the C/A point positioning was too good! As mentioned above, the accuracy was supposed to be ±100 meters horizontally, 95% of the time with a vertical accuracy of about ±175 meters. But, in fact, it turned out that the C/A-code point positioning gave civilians access to accuracy of about ±20 meters to ±40 meters. That was not according to plan, so the satellite clocks' accuracy was degraded on the C/A code. The good news is that the intentional error source is gone now.

Selective Availability was switched off on May 2, 2000 by presidential order. The intentional degradation of the satellite clocks is a thing of the past. Actually, Selective Availability never did hinder the surveying applications of GPS, more about that in Chapter 3. But do not think that satellite clock errors do not contribute error to GPS positioning anymore—they do.

PPS is designed for higher positioning accuracy and was originally available only to users authorized by the Department of Defense; that has changed somewhat, more about that later in this chapter.

The Production of a Modulated Carrier Wave

Since all the codes mentioned come to a GPS receiver on a modulated carrier, it is important to understand how a modulated carrier is generated. The signal created by an EDM is a good example of a modulated carrier.

EDM Ranging

As mentioned earlier, an EDM only needs one frequency standard because its electromagnetic wave travels to a retroprism and is reflected back to its origination. The EDM is both the transmitter and the receiver of the signal. Therefore, in general terms, the instrument can take half the time elapsed between the moment of transmission and the moment of reception, multiply by the speed of light, and find the distance between itself and the retroprism (Distance = Elapsed Time x Rate).

Illustrated in Figure 1.2, the fundamental elements of the calculation of the distance measured by an EDM, ρ, are the time elapsed between transmission and reception of the signal, Δt, and the speed of light, c.

Distance = ρ
Elapsed Time = Δt
Rate = c

GPS Ranging

However, the one-way ranging used in GPS is more complicated. It requires the use of two clocks. The broadcast signals from the satellites are collected by the

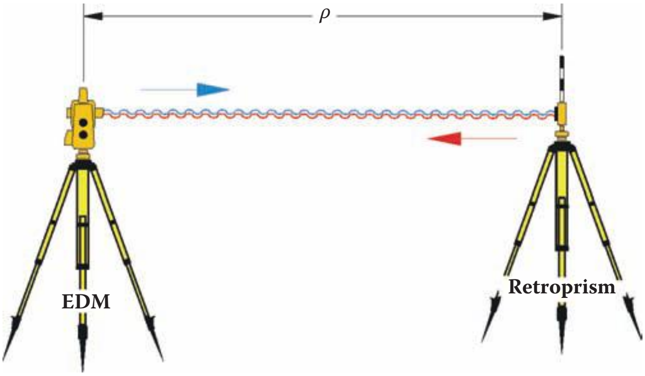


FIGURE 1.2 Two-Way Ranging

receiver, not reflected. Nevertheless, in general terms, the full time elapsed between the instant a GPS signal leaves a satellite and arrives at a receiver, multiplied by the speed of light, is the distance between them.

Unlike the wave generated by an EDM, a GPS signal cannot be analyzed at its point of origin. The measurement of the elapsed time between the signal's transmission by the satellite and its arrival at the receiver requires two clocks, one in the satellite and one in the receiver. This complication is compounded because to correctly represent the distance between them, these two clocks would need to be perfectly synchronized with one another. Since such perfect synchronization is physically impossible, the problem is addressed mathematically.

In Figure 1.3, the basis of the calculation of a range measured from a GPS receiver to the satellite, ρ, is the multiplication of the time elapsed between a signal's transmission and reception, Δt, by the speed of light, c.

A discrepancy of 1 microsecond, 1 millionth of a second, from perfect synchronization, between the clock aboard the GPS satellite and the clock in the receiver can create a range error of 300 meters, far beyond the acceptable limits for nearly all surveying work.

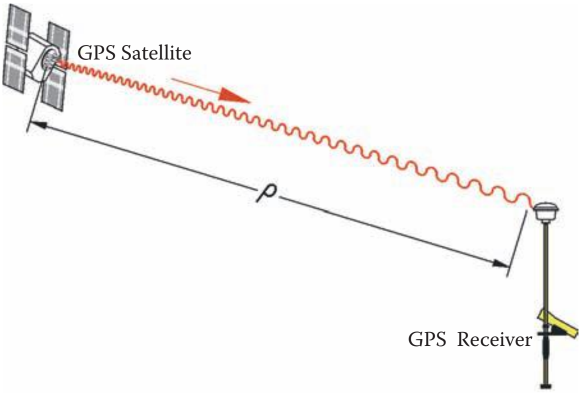


FIGURE 1.3 One-Way Ranging

Oscillators

The time measurement devices used in both EDM and GPS measurements are clocks only in the most general sense. They are more correctly called oscillators, or frequency standards. In other words, rather than producing a steady series of audible ticks they keep time by chopping a continuous beam of electromagnetic energy at extremely regular intervals. The result is a steady series of wavelengths and the foundation of the modulated carrier.

For example, the action of a shutter in a movie projector is analogous to the modulation of a coherent beam by the oscillator in an EDM. Consider the visible beam of light of a specific frequency passing through a movie projector. It is interrupted by the shutter, half of a metal disk rotating at a constant rate that alternately blocks and uncovers the light. In other words, the shutter chops the continuous beam into equal segments. Each length begins with the shutter closed and the light beam entirely blocked. As the shutter rotates open, the light beam is gradually uncovered. It increases to its maximum intensity, and then decreases again as the shutter gradually closes. The light is not simply turned on and off; it gradually increases and decreases. In this analogy the light beam is the carrier, and it has a wavelength much shorter than the wavelength of the modulation of that carrier produced by the shutter. This modulation can be illustrated by a sine wave.

The wavelength begins when the light is blocked by the shutter. The first minimum is called a 0° phase angle. The first maximum is called the 90° phase angle and occurs when the shutter is entirely open. It returns to minimum at the 180° phase angle when the shutter closes again. But the wavelength is not yet complete. It continues through a second shutter opening, 270°, and closing, 360°. The 360° phase angle marks the end of one wavelength and the beginning of the next one. The time and distance between every other minimum, that is from the 0° to the 360° phase angles, is a wavelength and is usually symbolized by the Greek letter lambda, λ.

As long as the rate of an oscillator's operation is very stable, both the length and elapsed time between the beginning and end of every wavelength of the modulation will be the same.

A Chain of Electromagnetic Energy

GPS oscillators are sometimes called clocks because the frequency of a modulated carrier, measured in hertz, can indicate the elapsed time between the beginning and end of a wavelength, which is a useful bit of information for finding the distance covered by a wavelength. The length is approximately:

$$\lambda = \frac{c_a}{f}$$

where

λ = the length of each complete wavelength in meters;
c_a = the speed of light corrected for atmospheric effects;
f = the frequency in hertz.

For example, if an EDM transmits a modulated carrier with a frequency of 9.84 MHZ and the speed of light is approximately 300,000,000 meters per second (a more accurate value is 299,792,458 meters per second, but the approximation 300,000,000 meters per second will be used here for convenience), then:

$$\lambda = \frac{c_a}{f}$$

$$\lambda = \frac{300{,}000{,}000 mps}{9{,}840{,}000 Hz}$$

$$\lambda = 30.49m$$

the modulated wavelength would be about 30.49 meters long, or approximately 100 feet.

The modulated carrier transmitted from an EDM can be compared to a Gunter's chain constructed of electromagnetic energy instead of wire links. Each full link of this electromagnetic chain is a wavelength of a specific frequency. The measurement between an EDM and a reflector is doubled with this electronic chain because, after it extends from the EDM to the reflector, it bounces back to where it started. The entire trip represents twice the distance and is simply divided by 2. But like the surveyors who used the old Gunter's chain, one cannot depend that a particular measurement will end conveniently at the end of a complete link (or wavelength). A measurement is much more likely to end at some fractional part of a link (or wavelength). The question is, where?

Phase Shift

With the original Gunter's chain, the surveyor simply looked at the chain and estimated the fractional part of the last link that should be included in the measurement. Those links were tangible. Since the wavelengths of a modulated carrier are not, the EDM must find the fractional part of its measurement electronically. Therefore, it does a comparison. It compares the phase angle of the returning signal to that of a replica of the transmitted signal to determine the *phase shift*. That phase shift represents the fractional part of the measurement. This principle is used in distance measurement by both EDM and GPS systems.

How does it work? First, it is important to remember that points on a modulated carrier are defined by phase angles, such as 0°, 90°, 180°, and 270° (Figure 1.4). When two modulated carrier waves reach exactly the same phase angle at exactly the same time, they are said to be *in phase, coherent, or phase locked*. However, when two waves reach the same phase angle at different times, they are *out of phase* or *phase shifted*. For example, in Figure 1.5, the sine wave shown by the dashed line has returned to an EDM from a reflector. Compared with the sine wave shown by the solid line, it is out of phase by one-quarter of a wavelength. The distance between the EDM and the reflector, ρ, is then:

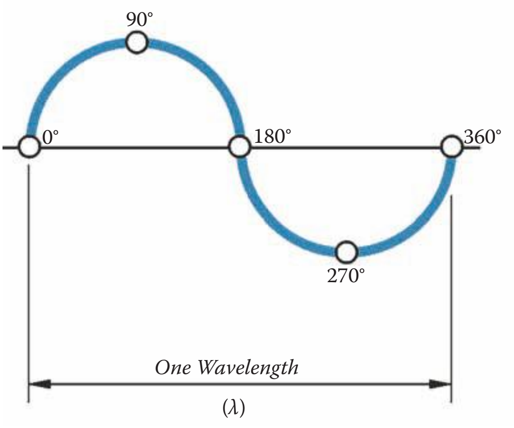


FIGURE 1.4 0° to 360° = One Wavelength

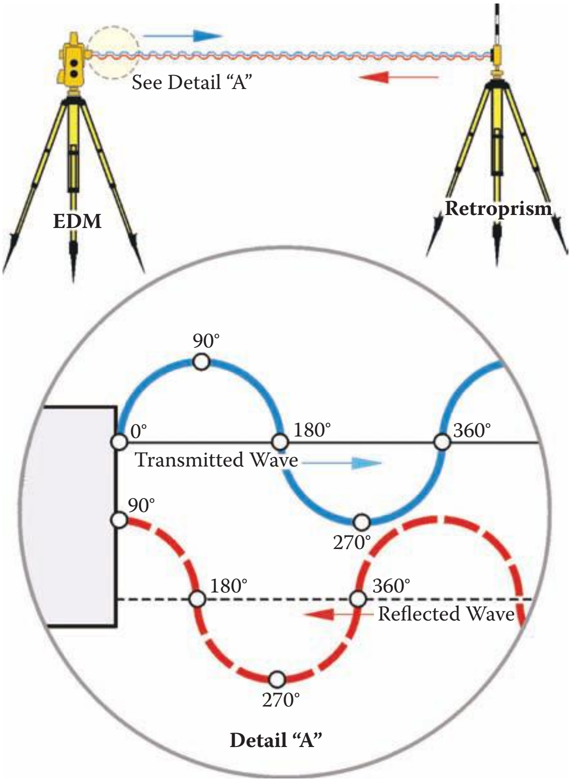


FIGURE 1.5 An EDM Measurement

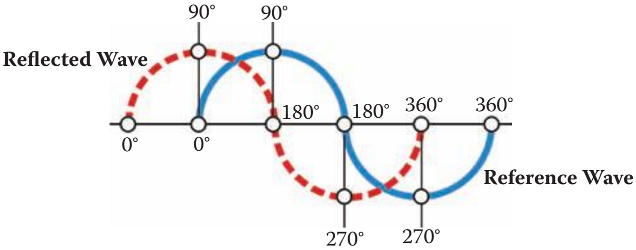


FIGURE 1.6 Reference and Reflected Waves

$$\rho = \frac{\left(N\lambda + d\right)}{2}$$

where

N = the number of full wavelengths the modulated carrier has completed
d = the fractional part of a wavelength at the end that completes the doubled distance.

In this example, d is three-quarters of a wavelength because it lacks its last quarter. But how would the EDM know that?

It knows because at the same time an external carrier wave is sent to the reflector, the EDM keeps an identical internal reference wave at home in its receiver circuits. In Figure 1.6, the external beam returned from the reflector is compared to the reference wave and the difference in phase between the two can be measured.

Both EDM and GPS ranging use the method represented in this illustration. In GPS, the measurement of the difference in the phase of the incoming signal and the phase of the internal oscillator in the receiver reveals the small distance at the end of a range. In GPS, the process is called *carrier phase ranging*. And as the name implies this measurement is actually done on the **carrier** itself.

The Cycle Ambiguity Problem

While this technique discloses the fractional part of a wavelength, a problem remains, determining the number of full wavelengths, *N*, of the EDM's modulated carrier between the transmitter and the receiver. This cycle ambiguity problem is solved in EDM measurements by modulating the carrier in ever longer wavelengths. For example, the meter and part of a meter aspect of a measured distance can be resolved by measuring the phase difference of a 10 meter wavelength. This procedure may be followed by the resolution of the tens of meters using a wavelength of 100 meters. The hundreds of meters can then be resolved with a wavelength of 1000 meters, and so on.

Such a method is convenient for the EDM's two-way ranging system, but impossible in the one-way ranging used in GPS measurements. GPS ranging must use an entirely different strategy for solving the cycle ambiguity problem because the satellites broadcast only two carriers, currently. Those carriers have constant wavelengths.

And they propagated in one direction, from the satellites to the receivers. Therefore, unlike an EDM measurement, the wavelengths of these carriers in GPS cannot be changed to resolve the cycle ambiguity problem, that is, the number of cycles between transmission and reception. Still, the carrier phase measurements remain an important *observable* in GPS ranging.

TWO OBSERVABLES

The word *observable* is used throughout GPS literature to indicate the signals whose measurement yields the range or distance between the satellite and the receiver. The word is used to draw a distinction between the thing being measured, the observable and the measurement, the observation.

In GPS there are two types of observables: the *pseudorange* and the carrier phase. The latter, also known as the *carrier beat phase*, is the basis of the techniques used for high-precision GPS surveys. On the other hand, the pseudorange can serve applications when virtually instantaneous point positions are required or relatively low accuracy will suffice.

These basic observables can also be combined in various ways to generate additional measurements that have certain advantages. It is in this latter context that pseudoranges are used in many GPS receivers as a preliminary step toward the final determination of position by a carrier phase measurement.

The foundation of pseudoranges is the correlation of code carried on a modulated carrier wave received from a GPS satellite with a replica of that same code generated in the receiver.

Most of the GPS receivers used for surveying applications are capable of code correlation. That is, they can determine pseudoranges from the C/A code or the P code. These same receivers are usually capable of determining ranges using the unmodulated carrier as well. However, first let us concentrate on the pseudorange.

Encoding by Phase Modulation

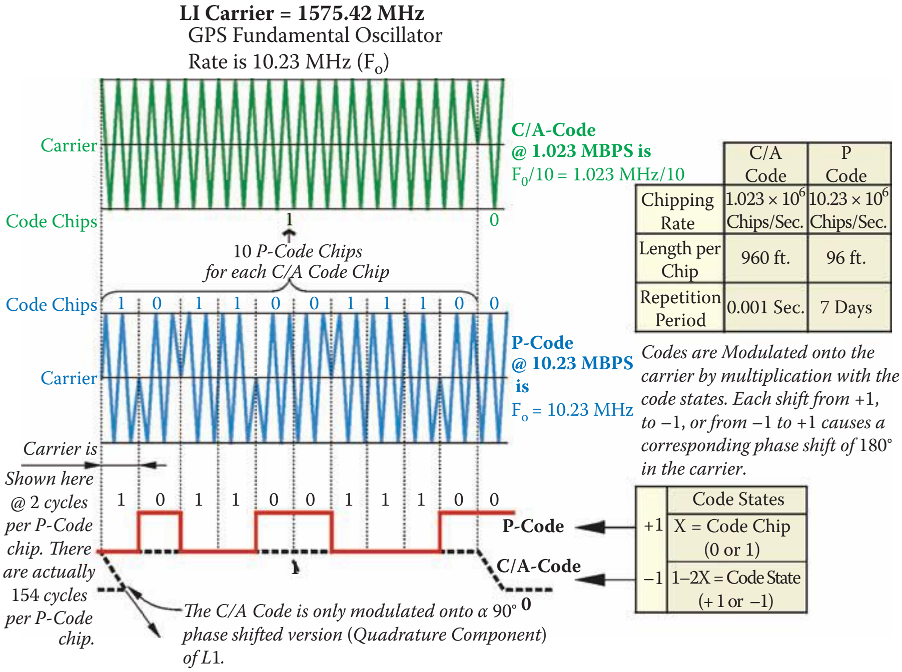
A carrier wave can be modulated in various ways. Radio stations use carrier waves that are AM, amplitude modulated or FM, frequency modulated. When your radio is tuned to 105 FM, you are not actually listening to 105 MHz despite the announcer's assurances, it is well above the range of human hearing. 105 MHz is just the frequency of the carrier wave that is being modulated. Those modulations occur at a much slower rate, making the speech and music intelligible.

The GPS carriers L1 and L2 could have been modulated in a variety of ways too, but the C/A and P codes impressed on the GPS carriers are the result of phase modulations. One consequence of this method of modulation is that the signal can occupy a broader bandwidth than it otherwise would. The GPS signal is said to have a *spread spectrum* because of its intentionally increased bandwidth. In other words the overall bandwidth of the GPS signal is much wider than the bandwidth of the information it is carrying. The C/A code signal is spread over a width of 2 MHz or so, and the P code signal is spread over a width about 20 MHz. This characteristic offers several advantages, including a better signal to noise ratio, more accurate ranging, less interference, and increased security. However, spreading the spectral density of the signal also reduces its power to a range from—160 *decibels per watt (dBw)* to—166 dBw. So low it is often described as tantamount to a 25-watt light bulb seen from 10,000 miles away. Clearly, it is somewhat difficult to receive the GPS signal under cover of foliage.

In any case the most commonly used spread spectrum modulation technique is known as *binary phase shift keying* (*BPSK*). This is the technique used to create the Navigation Message, the P code and the C/A code. All of them consist of 0s and 1s and this binary code is imprinted on the carrier wave without altering the amplitude, the frequency, or the wavelength of the carrier. Rather, the *binary biphase* modulation is accomplished by phase changes of 180º. When the value of the message is to change from zero or one, or from one to zero, the phase of the carrier wave is instantly reversed. It is flipped 180º. And each one of these flips occurs when the phase of the carrier is at the zero-crossing.

Each zero and one of the binary code is known as a *code chip*. Zero represents the *normal* state, and one represents the *mirror image* state. Please note in Figure 1.7 that each shift from zero to one, and from one to zero, is accompanied by a corresponding change in the phase of the carrier.

The rate of all of the components of GPS signals are multiples of the standard rate of the oscillators, which is 10.23 MHz. This rate is known as the *fundamental clock rate* and is symbolized F_o. For example, the GPS carriers are 154 times F_o, or 1575.42 MHz, and 120 times F_o, or 1227.60 MHZ, L1 and L2 respectively. The



	C/A Code	P Code
Chipping Rate	1.023×10^6 Chips/Sec.	10.23×10^6 Chips/Sec.
Length per Chip	960 ft.	96 ft.
Repetition Period	0.001 Sec.	7 Days

Codes are Modulated onto the carrier by multiplication with the code states. Each shift from +1, *to* −1, *or from* −1 *to* +1 *causes a corresponding phase shift of* 180° *in the carrier.*

	Code States
+1	X = Code Chip (0 or 1)
−1	1–2X = Code State (+1 or −1)

Carrier is Shown here @ 2 *cycles per P-Code chip. There are actually* 154 *cycles per P-Code chip.*

The C/A Code is only modulated onto a 90° *phase shifted version* (*Quadrature Component*) *of L*1.

FIGURE 1.7 Code Modulation of the L1 Carrier coming carrier L5 will also be a multiple of the fundamental clock rate, 115 times F_o or 1176.45 MHz.

The codes are also based on F_o. 10.23 *code chips* of the P code, zeros or ones, occur every microsecond. In other words, the *chipping rate* of the P code is 10.23 million bits per second (Mbps), exactly the same as F_o, 10.23 MHZ.

The chipping rate of the C/A code is 10 times slower than the P code, a tenth of F_o, 1.023 Mbps. Ten P code chips occur in the time it takes to generate one C/A code chip, allowing P code derived pseudoranges to be much more precise. This is one reason the C/A code is known as the *coarse/acquisition code*.

Even though both codes are broadcast on L1, they are distinguishable from one another by their transmission in *quadrature*. That means that the C/A code modulation on the L1 carrier is phase shifted 90° from the P code modulation on the same carrier.

PSEUDORANGING

Strictly speaking, a pseudorange observable is based on a time shift. This time shift can be symbolized by $d\tau$, d tau, and is the time elapsed between the instant a GPS signal leaves a satellite and the instant it arrives at a receiver. The concept can be illustrated by the process of setting a watch from a time signal heard over a telephone.

Propagation Delay

Imagine that a recorded voice said, "The time at the tone is 3 hours and 59 minutes." If a watch was set at the instant the tone was **heard**, the watch would be wrong. Supposing that the moment the tone was broadcast was indeed 3 hours and 59 minutes, the moment the tone is heard must be a bit later. It is later because of the time it took the tone to travel through the telephone lines from the point of broadcast to the point of reception. This elapsed time would be approximately equal to the length of the circuitry traveled by the tone divided by the speed of the electricity, which is the same as the speed of all electromagnetic energy, including light and radio signals. In fact, it is possible to imagine measuring the actual length of that circuitry by doing the division.

In GPS, that elapsed time is known as the propagation delay and it is used to measure length. The measurement is accomplished by a combination of codes. The idea is somewhat similar to the strategy used in EDMs. But where an EDM generates an internal replica of its *modulated carrier wave* to correlate with the signal it receives by reflection, a pseudorange is measured by a GPS receiver using a replica of the **code** that has been impressed on the modulated carrier wave. The GPS receiver generates this replica itself to compare with the code it receives from the satellite.

Code Correlation

To conceptualize the process, one can imagine two codes generated at precisely the same time and identical in every regard: one in the satellite and one in the receiver. The satellite sends its code to the receiver but on its arrival, the two codes do not line up. The codes are identical but they do not correlate until the replica code in the receiver is time shifted.

Then the receiver generated replica code is shifted relative to the received satellite code. It is this time shift that reveals the propagation delay, the time it took the signal to make the trip from the satellite to the receiver, $d\tau$. It is the same idea described above as the time it took the tone to travel through the telephone lines, except the GPS code is traveling through space and atmosphere. Once the time shift is accomplished, the two codes match perfectly and the time the satellite signal spends in transit has been measured, well, almost.

It would be wonderful if that time shift could simply be divided by the speed of light and yield the true distance between the satellite and the receiver at that instant, and it is close. But there are physical limitations on the process that prevent such a perfect relationship, more about that later in this chapter.

AUTOCORRELATION

As mentioned earlier, the almanac information from the Navigation message of the first satellite a GPS receiver acquires tells it which satellites can be expected to come into view. With this information the receiver can load up pieces of the C/A codes for each of those satellites. Then the receiver tries to line up the replica C/A codes with the signals it is actually receiving from the satellites. The time required for correlation to occur is influenced by the presence and quality of the information in the almanac.

Actually lining up the code from the satellite with the replica in the GPS receiver is called *autocorrelation*, and depends on the transformation of code chips into *code states*. The formula used to derive code states (+1 and -1) from code chips (0 and 1) is:

$$\text{code state} = 1 - 2x$$

where x is the code chip value. For example, a normal code state is +1, and corresponds to a code chip value of 0. A mirror code state is -1, and corresponds to a code chip value of 1.

The function of these code states can be illustrated by asking three questions: First, if a tracking loop of code states generated in a receiver does **not** match code states received from the satellite, how does the receiver know? In that case, for example, the sum of the products of each of the receiver's 10 code states, with each of the code states from the satellite, when divided by 10, *does not equal 1.* Second, what does the receiver do when the code states in the receiver do not match the code states received from the satellite? It shifts the frequency of its search a little bit from the center of the L1 1575.42 MHZ. This is done to accommodate the inevitable Doppler shift of the incoming signal since the satellite is always either moving toward or away from the receiver. The receiver also shifts its piece of code in time. These iterative small shifts in both time and frequency continue until the receiver code states do in fact match the signal from the satellite. Third, how does the receiver know when a tracking loop of replica code states *does* match code states from the satellite? In the case illustrated in Figure 1.8, the sum of the products of each code state of the receiver's replica 10, with each of the 10 from the satellite, divided by 10, *is exactly 1.*

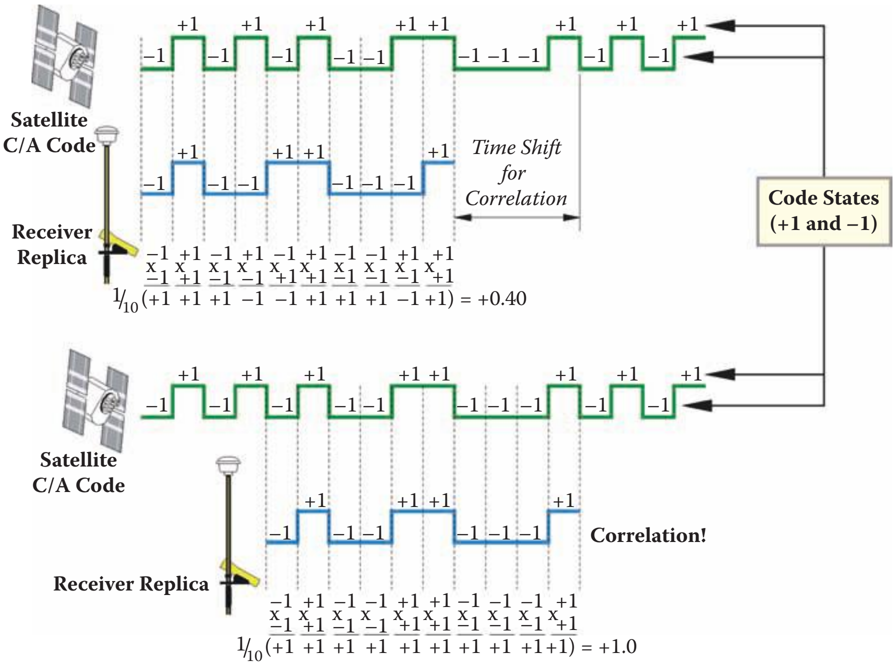


FIGURE 1.8 Code Correlation

The *autocorrelation function* is:

$$\frac{1}{N}\int_{0}^{T} X(t) * X(t-\tau)\,dt = \frac{1}{N}\sum_{i=1}^{N} X_i * X_{i-j}$$

In Figure 1.8, before the code from the satellite and the replica from the receiver are matched, the sum of the products of the code states is not 1. Following the correlation of the two codes the sum of the code states is exactly 1, and the receiver's replica code fits the code from the satellite like a key fits a lock.

Correlation Peak

As shown in the somewhat idealized triangular plot in Figure 1.9, the condition of maximum correlation has been achieved at 1. That is when the receiver generated replica matches the code being received from the GPS satellite. In practice, the correlation peak is actually a bit rounded rather than being so emphatically triangular, but in any case once the correlation peak has been reached it is then maintained by continual adjustment of the receiver generated code as described earlier.

It is usual to find three correlators used for tracking. One is at the *prompt (P), punctual*, or on-time position with the other two symmetrically located somewhat early (*E*) and late (*L*). The separation between the early, prompt, and late code phases

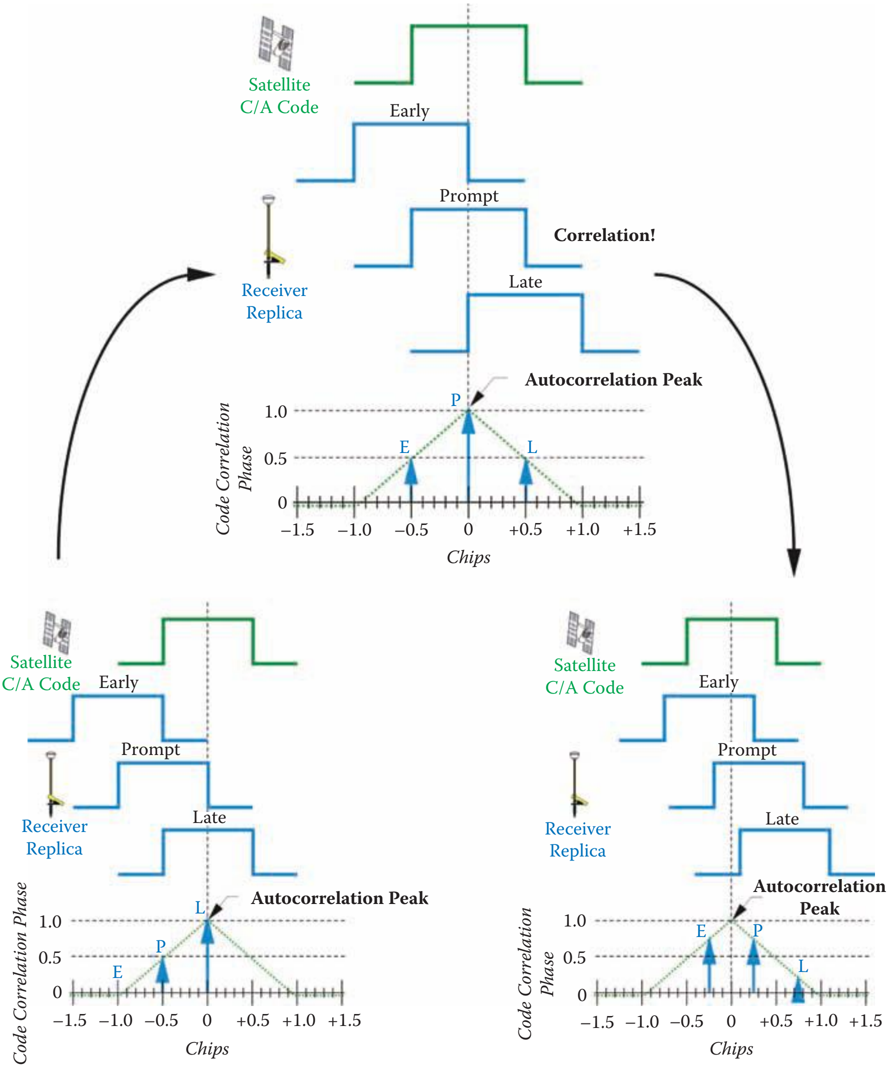


FIGURE 1.9 Correlation Peak

is always symmetrical. The separation is often 1 chip as illustrated here, but in a *narrow correlator* they may be separated by as little as a 1/10th of a chip. Whatever the size of the separation the symmetry is important to their purpose. That purpose is giving the receiver the information it needs to perform the shifts, the continual adjustments, alluded to above.

Because there is an equal distance between them when the receiver's replica code is aligned with the code from the satellite the early and late correlators will have exactly the same amplitude on the triangle as shown on the right side of in Figure 1.9. On the other hand, when the codes are not aligned, the early and late correlator amplitudes will be different. And, it is important to note that the symmetry ensures that the difference in their amplitudes will be unequal in proportion to the amount that the codes are out of phase with one another. Using this information the receiver can not only calculate the amount of the correlation error, but also whether the receiver's replica code is ahead (early) or behind (late) the incoming satellite code. Using this information it can correct the replica code generation to match the satellite's signal. Correlation typically takes about 1 to 5 milliseconds (ms) and does not exceed 20 ms. Please note that the same early, prompt, and late correlator approach is also utilized to align the receiver's replica carrier phase with the incoming satellite carrier phase.

LOCK AND THE TIME SHIFT

Once correlation of the two codes is achieved with a *delay lock loop* (*DLL*), it is maintained by a *correlation channel* within the GPS receiver, and the receiver is sometimes said to have *achieved lock* or to be *locked on* to the satellites. If the correlation is somehow interrupted later, the receiver is said to have *lost lock*. However, as long as the lock is present, the Navigation message is available to the receiver. Remember that one of its elements is the broadcast clock correction that relates the satellite's onboard clock to GPS time, and a limitation of the pseudorange process comes up.

Imperfect Oscillators

One reason the time shift, $d\tau$, found in autocorrelation cannot quite reveal the true range, ρ, of the satellite at a particular instant is the lack of perfect synchronization between the clock in the satellite and the clock in the receiver. Recall that the two compared codes are generated directly from the fundamental rate, F_o, of those clocks. And since these widely separated clocks, one on Earth and one in space, cannot be in perfect lockstep with one another, the codes they generate cannot be in perfect synch either. Therefore, a small part of the observed time shift, $d\tau$, must always be due to the disagreement between these two clocks. In other words, the time shift not only contains the signal's transit time from the satellite to the receiver, it contains clock errors too.

In fact, whenever satellite clocks and receiver clocks are checked against the carefully controlled GPS time they are found to be drifting a bit. Their oscillators are imperfect. It is not surprising that they are not quite as stable as the more than 150 atomic clocks around the world that are used to define the rate of GPS time. They are subject to the destabilizing effects of temperature, acceleration, radiation, and other inconsistencies. As a result, there are two clock offsets that bias every satellite to receiver pseudorange observable. That is one reason it is called a *pseudo*-range (see Figure 1.10).

A PSEUDORANGE EQUATION

Clock offsets are only one of the errors in pseudoranges. Their relationship can be illustrated by the following equation (Fotopoulos 2000):

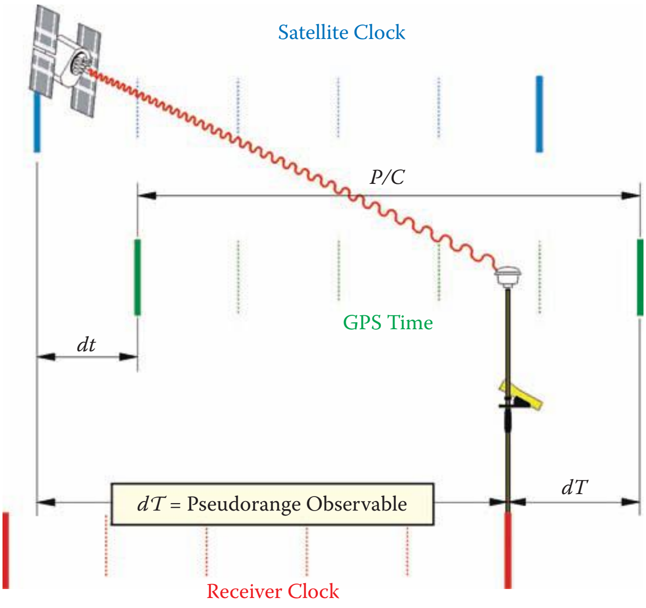


FIGURE 1.10 Pseudorange

$$p = \rho + d_\rho + c\left(dt - dT\right) + d_{ion} + d_{trop} + \varepsilon_{m_p} + \varepsilon_p \quad \text{(pseudorange)}$$

where

p = the pseudorange measurement
ρ = the true range
d_ρ = satellite orbital errors
c = the speed of light
dt = the satellite clock offset from GPS time
dT = the receiver clock offset from GPS time
d_{ion} = ionospheric delay
d_{trop} = tropospheric delay
ε_{mp} = multipath
ε_p = receiver noise

Please note that the pseudorange, p, and the true range, ρ, cannot be made equivalent, without consideration of clock offsets, atmospheric effects, and other biases that are inevitably present.

This discussion of time can make it easy to lose sight of the real objective, which is the position of the receiver. Obviously, if the coordinates of the satellite and the coordinates of the receiver were known perfectly, then it would be a simple matter to determine time shift $d\tau$ or find the true range ρ between them.

In fact, receivers placed at known coordinated positions can establish time so precisely they are used to monitor atomic clocks around the world. Several receivers simultaneously tracking the same satellites can achieve resolutions of 10 nanoseconds or better. Also, receivers placed at known positions can be used as base stations to establish the relative position of receivers at unknown stations, a fundamental principle of most GPS surveying.

It can be useful to imagine the true range term ρ, also known as the *geometric range*, actually includes the coordinates of both the satellite and the receiver. However, they are hidden within the measured value, the pseudorange, p, and all of the other terms on the right side of the equation. The objective then is to mathematically separate and quantify these biases so that the receiver coordinates can be revealed. Clearly, any deficiency in describing, or *modeling*, the biases will degrade the quality of the final determination of the receiver's position.

THE ONE-PERCENT RULE OF THUMB

Here is convenient approximation. The maximum resolution available in a pseudorange is about 1 percent of the chipping rate of the code used, whether it is the P code or the C/A code. In practice, positions derived from these codes are rather less reliable than described by this approximation, but it offers a basis to compare P code and C/A code pseudoranging.

A P code chip occurs every 0.0978 of a microsecond. Therefore, a P code based measurement can have a maximum precision of about 1 percent of about one-tenth of a microsecond, or 1 nanosecond. One nanosecond multiplied by the speed of light is approximately 30 centimeters, 1 percent of the length of a single P code chip.

The C/A code based pseudorange is 10 times less precise. Its chipping rate is 10 times slower. A C/A code pseudorange has a maximum resolution of about 3 meters, that is 1 percent of the length of a single C/A code chip.

This same 1 percent rule of thumb can illustrate the increased precision of the carrier phase observable over the pseudorange. First, the length of a single wavelength of each carrier is calculated using the same formula as was used previously.

$$\lambda = \frac{c_a}{f}$$

where

λ = the length of each complete wavelength in meters;
c_a = the speed of light corrected for atmospheric effects;
f = the frequency in hertz.

The L1-1575.42 MHz **carrier** transmitted by GPS satellites has a wavelength of approximately 19 cm.

$$\lambda = \frac{c_a}{f}$$

$$\lambda = \frac{300 * 10^6\, mps}{1575.42 * 10^6\, Hz}$$

$$\lambda = 0.19m$$

The L2-1227.60 MHz frequency **carrier** transmitted by GPS satellites has a wavelength of approximately 24 cm.

$$\lambda = \frac{c_a}{f}$$

$$\lambda = \frac{300 * 10^6 \, mps}{1227.60 * 10^6 \, Hz}$$

$$\lambda = 0.24m$$

Therefore, using either carrier, the carrier phase measurement resolved to 1% of the wavelength would be about 2 mm.

A 3 m ranging precision is not adequate for most land surveying applications, not to say all. However, the actual positional accuracy of a single C/A code receiver was about ±100 m with selective availability, SA, turned on and ±30 m with SA turned off unless differential GPS, *DGPS*, techniques are used, more about that in Chapter 6. But carrier phase observations are certainly the preferred method for the higher precision work most surveyors have come to expect from GPS.

CARRIER PHASE RANGING

Carrier Phase Comparisons

Carrier phase is the observable at the center of high accuracy surveying applications of GPS. It depends on the carrier waves themselves, the unmodulated L1 and L2, rather than their P and C/A codes. One bit of good news was always apparent from this strategy immediately; the user was immune from SA, selective availability. As mentioned earlier SA was the intentional degradation of the SPS, the standard positioning service available through the C/A code. Since carrier phase observations do not use codes, they were never affected by SA. Now that SA has been turned off, the point is moot, unless SA should be re-instituted.

Understanding carrier phase is perhaps a bit more difficult than the pseudorange, but the basis of the measurements has some similarities. For example, the foundation of a pseudorange measurement is the correlation of the codes received from a GPS satellite with replicas of those codes generated within the receiver. The foundation of the carrier phase measurement is the combination of the unmodulated carrier itself received from a GPS satellite with a replica of that carrier generated within the receiver.

Phase Difference

A few similarities can also be found between a carrier phase observation and a distance measurement by an EDM. As mentioned earlier an EDM sends a modulated carrier wave to the reflector, and generates an identical internal reference. When the external beam returns from the reflector, it is compared with the reference wave. The difference in phase between the two reveals the fractional part of the measurement, even though the number of complete cycles between the EDM and the reflector may not be immediately apparent until modulated carriers of longer wavelengths are used.

Likewise, it is the phase difference between the incoming signal and the internal reference that reveals the fractional part of the carrier phase measurement in GPS. The incoming signal is from a satellite rather than a reflector of course, but like an EDM measurement, the internal reference is derived from the receiver's oscillator and the number of complete cycles is not immediately known.

Beat

The carrier phase observable is sometimes called the *reconstructed carrier phase* or *carrier beat phase* observable. In this context, a *beat* is the pulsation resulting from the combination of two waves with different frequencies. An analogous situation occurs when two musical notes of different pitch are sounded at the same time. Their two frequencies combine and create a third note, called the beat. Musicians can tune their instruments by listening for the beat that occurs when two pitches differ slightly. This third pulsation may have a frequency equal to the difference or the sum of the two original frequencies.

The beat phenomenon is by no means unique to musical notes; it can occur when any pair of oscillations with different frequencies are combined. In GPS, a beat is created when a carrier generated in a GPS receiver and a carrier received from a satellite are combined (see Figure 1.11). At first, that might not seem sensible.

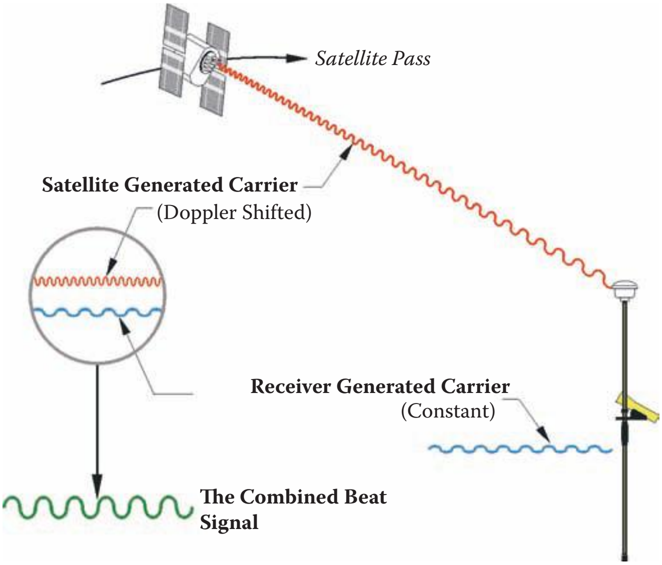


FIGURE 1.11 The Carrier Beat Phase

How could a beat be created by combining two absolutely identical unmodulated carriers? There should be no difference in frequency between an L1 carrier generated in a satellite and an L1 carrier generated in a receiver. They both should have a frequency of 1575.42 MHz. If there is no difference in the frequencies, how can there be a beat? But there is a slight difference between the two carriers. Something happens to the frequency of the carrier on its trip from a GPS satellite to a receiver. Its frequency changes. The phenomenon is described as the Doppler effect.

The Doppler Effect

Sound provides a model for the explanation of the phenomenon. An increase in the frequency of a sound is indicated by a rising pitch; a lower pitch is the result of a decrease in the frequency. A stationary observer listening to the blasting horn on a passing train will note that as the train gets closer, the pitch rises, and as the train travels away, the pitch falls. Furthermore, the change in the sound, clear to the observer standing beside the track, is not heard by the engineer driving the train. He hears only one constant, steady pitch. The relative motion of the train with respect to the observer causes the apparent variation in the frequency of the sound of the horn.

In 1842, Christian Doppler described this frequency shift now named for him. He used the analogy of a ship on an ocean with equally spaced waves. In his allegory, when the ship is stationary the waves strike it steadily, one each second. But if the ship sails into the waves, they break across its bow more frequently. If the ship then turns around and sails with the waves, they strike less frequently across its stern. The waves themselves have not changed; their frequency is constant. But to the observer on the ship their frequency seems to depend on his motion.

GPS and the Doppler Effect

From the observer's point of view, whether it is the source, the observer or both that are moving, the frequency increases while they move together and decreases while they move apart. Therefore, if a GPS satellite is moving toward an observer, its carrier wave comes into the receiver with a higher frequency than it had at the satellite. If a GPS satellite is moving away from the observer, its carrier wave comes into the receiver with a lower frequency than it had at the satellite. Since a GPS satellite is virtually always moving with respect to the observer, any signal received from a GPS satellite is Doppler-shifted.

A Carrier Phase Approximation

The carrier phase observation in cycles is often symbolized by *phi*, φ, in GPS literature. Other conventions include the use of superscripts to indicate satellite designations and the use of subscripts to define receivers. For example, in the following equation $\varphi^s{}_r$ is used to symbolize the carrier phase observation between satellite *s* and receiver *r*. The difference that defines the carrier beat phase observation is (Wells et al. 1986):

$$\varphi = \varphi_r^s = \varphi^s\left(t\right) - \varphi_r\left(T\right)$$

where $\varphi^s(t)$ is the phase of the carrier broadcast from the satellite s at time t. Please note that the frequency of this carrier is the same, nominally constant frequency, that is generated by the receiver's oscillator. $\varphi_r(T)$ is its phase when it reaches the receiver r at time T.

A description of the use of the carrier phase observable to measure range can start with the same basis as the calculation of the pseudorange, travel time. The time elapsed between the moment the signal is broadcast, t, and the moment it is received, T, multiplied by the speed of light, c, will yield the range between the satellite and receiver, ρ:

$$(T-t)c \approx \rho$$

Even using this simplified equation it is possible to get an approximate idea of the relation between time and range for some assumed, nominal values. These values could not be the basis of any actual carrier phase observation because, among other reasons, it is not possible for the receiver to know when a particular carrier wave left the satellite. But for the purpose of illustration, suppose a carrier left a satellite at 00 hours 00 minutes 00.000 seconds, and arrived at the receiver 67 milliseconds later.

$$(T-t)c \approx \rho$$

$$(00h:00m:00.067s-00h:00m:00.000s)\,300{,}000\,km/s \approx \rho$$

$$(0.067)\,300{,}000\,km/s \approx \rho$$

$$20{,}100\,km \approx \rho$$

This estimate indicates that if the carrier broadcast from the satellite reaches the receiver 67 milliseconds later, the range between them is approximately 20,100 km.

Carrying this example a bit farther, the wavelength, λ, of the L1 carrier can be calculated by dividing the speed of light, c, by the L1 frequency, f:

$$\lambda = \frac{c}{f}$$

$$\lambda = \frac{300{,}000{,}000\,m/s}{1{,}575{,}420{,}000\,Hz}$$

$$\lambda = 0.1904254104\,m$$

Dividing the approximated range, ρ, by the calculated L1 carrier wavelength, λ, yields a rough estimation of the carrier phase in cycles, φ:

$$\frac{\rho}{\lambda} \approx \varphi$$

$$\frac{20{,}100{,}000m}{0.1904254104m} \approx \varphi$$

$$105{,}553{,}140 \ \mathit{cycles} \approx \varphi$$

The 20,100 km range implies that the L1 carrier would cycle through approximately 105,553,140 wavelengths on its trip from the satellite to the receiver.

Of course, these relationships are much simplified. However, they can be made fundamentally correct by recognizing that ranging with the carrier phase observable is subject to all of the same biases and errors as the pseudorange. For example, terms such as the receiver clock offset may be incorporated, again symbolized by *dT* as it was in the pseudorange equation. The imperfect satellite clock can be included; its error is symbolized by *dt*. The tropospheric delay d_{trop}, the ionospheric delay d_{ion} and multipath-receiver noise ε_{φ} are also added to the range measurement. The ionospheric delay will be negative here, more about that later. With these changes the simplified travel time equation can be made a bit more realistic.

$$\left[\left(T + dT\right) - \left(t + dt\right)\right]c \approx \rho - d_{ion} + d_{trop} + \varepsilon_{\varphi}$$

This, more realistic, equation can be rearranged to isolate the elapsed time *(T-t)* on one side, by dividing both sides by c, and then moving the clock errors to the right side (Wells et al. 1986).

$$\frac{\left[\left(T + dT\right) - \left(t + dt\right)\right]c}{c} = \frac{\rho - d_{ion} + d_{trop} + \varepsilon_{\varphi}}{c}$$

$$\left[\left(T + dT\right) - \left(t + dt\right)\right] = \frac{\rho - d_{ion} + d_{trop} + \varepsilon_{\varphi}}{c}$$

$$\left(T - t + dT - dt\right) = \frac{\rho - d_{ion} + d_{trop} + \varepsilon_{\varphi}}{c}$$

$$(T - t + dT - dt) + (dt - dT) = \frac{\rho - d_{ion} + d_{trop} + \varepsilon_{\varphi}}{c} + \left(dt - dT\right)$$

$$T - t = dt - dT + \frac{\rho - d_{ion} + d_{trop} + \varepsilon_{\varphi}}{c}$$

This expression now relates the travel time to the range. However, in fact, a carrier phase observation cannot rely on the travel time, for two reasons. First, in a carrier phase observation the receiver has no codes with which to tag any particular instant on the incoming continuous carrier wave. Second, since the receiver cannot distinguish one cycle of the carrier from any other, it has no way of knowing the initial phase of the signal when it left the satellite. In other words, the receiver cannot know the travel time and, therefore, it is hard to see how it can determine the number of complete cycles between the satellite and itself. This unknown quantity is called the *cycle ambiguity.*

Remember, the approximation of 105,553,140 wavelengths in this example is for the purpose of comparison and illustration only. In actual practice, a carrier phase observation must derive the range from a measurement of phase at the receiver, not from a known travel time of the signal.

The missing information is the number of complete phase cycles between the receiver and the satellite at the instant that the tracking began. The critical unknown integer, symbolized by *N*, is the cycle ambiguity, and it cannot be directly measured by the receiver. The receiver can count the complete phase cycles it receives from the moment it starts tracking until the moment it stops. It can also monitor the fractional phase cycles, but the cycle ambiguity *N* is unknown.

An Illustration of the Cycle Ambiguity Problem

The situation is somewhat analogous to an unofficial technique used by some nineteenth century contract surveyors on the Great Plains. The procedure can be used as a rough illustration of the cycle ambiguity problem in GPS.

It was known as the *buggy wheel* method of chaining. Some of the lines of the public land system that crossed open prairies were originally surveyed by loading a wagon with stones or stakes and tying a cloth to a spoke of the wheel. One man drove the team, another kept the wagon on-line with a compass, and a third counted the revolutions of the flagged wheel to measure the distance. When there had been enough turns of the improvised odometer to measure half a mile, they set a stone or stake to mark the corner and then rolled on, counting their way to the next corner.

A GPS receiver is like the man assigned to count the turns of the wheel. He is supposed to begin his count from the moment the crew leaves the newly set corner, but instead, suppose he jumps into the wagon, gets comfortable, and takes an unscheduled nap. When he wakes up, the wagon is on the move. Trying to make up for his laxness, he immediately begins counting. But at that moment the wheel is at a half turn, a fractional part of a cycle. He counts the subsequent half turn and then, back on the job, he intently counts each and every full revolution as they come around. His tally grows as the cycles accumulate, but he is in trouble and he knows it. He cannot tell how far the wagon has traveled; he was asleep for the first part of the trip. He has no way of knowing how far they had come before he woke up and started counting. He is like a GPS receiver that cannot know how far it is from the satellite when it starts counting phase cycles. They can tell it nothing about how many cycles stood between itself and the satellite when the receiver was locked on and began tracking. Those unknown cycles are the cycle ambiguity.

The 360° cycles in the carrier phase observable are wavelengths λ, not revolutions of a wheel. Therefore, the cycle ambiguity included in the complete carrier phase equation is an integer number of wavelengths, symbolized by λN (Fotopoulos 2000). So, the complete carrier phase observable equation can now be stated as:

$$\phi = \rho + d_\rho + c(dt - dT) + \lambda N - d_{ion} + d_{trop} + \varepsilon_{m_\phi} + \varepsilon_\phi \text{ (carrier phase)}$$

where

p = the pseudorange measurement
ρ = the true range
d_ρ = satellite orbital errors
c = the speed of light
dt = the satellite clock offset from GPS time
dT = the receiver clock offset from GPS time
λ = the carrier wavelength
N = the integer ambiguity in cycles
d_{ion} = ionospheric delay
d_{trop} = tropospheric delay
$\varepsilon_{m\varphi}$ = multipath
ε_φ = receiver noise

EXERCISES

1. What is the function of the information in subframe 5 of the Navigation message?

   (a) Once a receiver is locked onto a satellite it helps the receiver to determine the position of the satellite that is transmitting the Navigation message.
   (b) Once a receiver is locked onto a satellite it helps the receiver to correct the part of the delay of the signal caused by the ionosphere.
   (c) Once a receiver is locked onto a satellite it helps the receiver to correct the received satellite time to GPS time.
   (d) Once a receiver is locked onto a satellite it helps the receiver acquire the signals of the other satellites.

2. Which height most correctly expresses a nominal altitude of GPS satellites above the earth?

   (a) 20,000 miles
   (b) 35,420 km
   (c) 20,183,000 m
   (d) 108,000 nautical miles

3. Which of the following statements about the clocks in GPS satellites is not correct?

   (a) The signal for each satellite is independent from the other satellites and is generated from its own onboard clock.
   (b) The clocks in GPS satellites may also be called oscillators or frequency standards.
   (c) Every GPS satellite is launched with very stable atomic clocks onboard.
   (d) The clocks in any one satellite are allowed to drift up to one nanosecond from GPS time before they are tweaked by the Control Segment.

4. The Global Positioning System is known as a passive system. What does that mean?

   (a) The ranges are measured with signals in the microwave part of the electromagnetic spectrum.
   (b) Only the satellites transmit signals; the users receive them.
   (c) A GPS receiver must be able to gather all the information it needs to determine its own position from the signals it bounces off the satellites.
   (d) The signals from a GPS receiver return to the satellite.

5. Which comparison of EDM and GPS processes is correct?

   (a) EDM and GPS signals are both reflected back to their sources.
   (b) EDM measurements require atmospheric correction, GPS ranges do not.
   (c) EDM's and GPS satellites both transmit modulated carriers.
   (d) Phase differencing is used in EDM measurement, but not in GPS.

6. What information below is critical to defining the relationship between GPS time and UTC?

   (a) multipath
   (b) the broadcast ephemeris
   (c) an antispoofing flag
   (d) leap seconds

7. The P code and the C/A code are created by shifts from 0 to 1 or 1 to 0 known as code chips. There is a corresponding change of 180° in the GPS carrier waves. What changes?

   (a) the fundamental clock rate
   (b) the frequency
   (c) the phase
   (d) the amplitude

8. The P code and the C/A code are both broadcast on L1. How can a GPS receiver distinguish between them on that carrier?

   (a) The chipping rate for the C/A code is ten times slower than the chipping rate for the P code.
   (b) They are broadcast in binary biphase modulation.
   (c) The fundamental clock rate is 10.23 MHz.
   (d) They are broadcast in quadrature.

9. Which of the following is the most correct description of the Doppler effect?

   (a) The distortion of electromagnetic waves due to the density of charged particles in the Earth's upper atmosphere.
   (b) The systematic changes that occur when a moving object or light beam pass through a gravitational field.
   (c) The systematic changes that occur when a moving object approaches the speed of light.
   (d) The shift in frequency of acoustic or electromagnetic radiation emitted by a source moving relative to an observer.

10. When Selective Availability (SA) was switched off how did it affect surveying applications of GPS?

   (a) The accuracy of most surveying applications of GPS doubled.
   (b) The P code was not encrypted.
   (c) The cycle ambiguity problem was eliminated.
   (d) There was no significant change.

ANSWERS AND EXPLANATIONS

1. Answer is (d)

Explanation: Subframe 5 contains the ephemerides of up to 24 satellites. The purpose of subframe 5 is to help the receiver acquire the signals of the other satellites. The receiver must first lock onto a satellite to have access to the Navigation message, of course, but once the Navigation message can be read the positions of all of the other satellites can be computed.

2. Answer is (c)

Explanation: The GPS constellation is still evolving, but the orbital configuration is fairly well settled. Each GPS satellite's orbit is nearly circular and has a nominal altitude of 20,183 km, or 20,183,000 meters.

The orbit is approximately 12,500 statute miles or 10,900 nautical miles above the surface of the planet. GPS satellites are higher than the usual orbit assigned to most other satellites including the space shuttle. Their high altitude allows each GPS satellite to be viewed simultaneously from a large portion of the earth at any given moment. However, GPS satellites are well below the height, 35,420 km, required for the sort of geosynchronous orbit used for communications satellites.

3. Answer is (d)

Explanation: Government tracking facilities monitor the drift of each satellite clock's deviation from GPS time. The drift is allowed to reach a maximum of one millisecond before it is adjusted. Until the difference reaches that level, it is contained in the broadcast clock correction in subframe 1 of the Navigation message.

4. Answer is (b)

Explanation: Some of the forerunners of GPS were navigational systems that involved transmissions from the users, but not GPS. The military designed the system to exclude anything that would reveal the location of the GPS receiver. If GPS had been built as a two-way system it would have been much more complicated especially as the number of users grew. Therefore, it is a passive system, meaning that the satellites transmit and the users receive.

5. Answer is (c)

Explanation: An unmodulated carrier carries no information and no code. While GPS receivers can and do make phase measurements on the unmodulated carrier waves, they also make use of the code information available on the modulated carrier transmitted by GPS satellites to determine pseudoranges. GPS uses a one-way system. The modulated carrier with code information travels from the satellite to the receiver where it is correlated with a reference. This one-way method requires a frequency standard in both the satellite and the receiver.

An EDM's measurement is also based on a modulated carrier. However, the EDM uses a two-way system. The modulated carrier is transmitted to a retroprism. It is reflected and returns to the EDM. The EDM can then determine the phase delay by comparing the returned modulated wave with a reference in the instrument. This two-way method requires only one frequency standard in the EDM since the modulated carrier is reflected.

6. Answer is (d)

Explanation: GPS time is calculated using the atomic clock at the Master Control Station (MCS) at the Schriever Air Force Base, formerly known as Falcon Air Force Station near Colorado Springs, Colorado. GPS time is kept within 1 millisecond of UTC. However, UTC is adjusted for leap seconds, and GPS time is not. Conversion of GPS time to UTC requires knowledge of the leap seconds applied to UTC since January 1980. This information is available from the United States Naval Observatory, USNO, time announcements.

7. Answer is (c)

Explanation: Phase modulations are used to mark the divisions between the code chips. Whenever the C/A code or the P code switches from a binary 1 to a binary 0 or vice versa, its L1 or L2 carriers has a sharp mirror-image shift in phase. About every millionth of a second the phase of the C/A code carrier can shift. But the P code carrier can have a phase shift about every ten millionth of a second.

8. Answer is (d)

Explanation: Even though both codes are broadcast on L1, they are distinguishable from one another by their transmission in quadrature. That means that the C/A code modulation on the L1 carrier is phase shifted 90° from the P code modulation on the same carrier.

9. Answer is (d)

Explanation: If the source of radiation is moving relative to an observer, there is a difference between the frequency perceived by the observer and the frequency of the radiation at its source. The shift is to higher frequencies when the source moves toward the observer and to lower frequencies when it moves away. This effect has been named for its discoverer, C.J. Doppler.

10. Answer is (d)

Explanation: When Selective Availability (SA) was switched off on May 2, 2000 by presidential order, the effect on surveying applications of GPS was negligible.

Selective Availability (SA) means that the GPS signals were transmitted from the satellites with intentional clock errors added. The stability of the atomic clocks onboard was deliberately degraded and thereby the Navigation message was degraded too. While Selective Availability was removed by Precise Positioning Service, P code, users with decryption techniques, civilian users of the Standard Positioning Service C/A code were not able to remove these errors.

Selective availability was on since the first launch of the Block II satellites in 1989. It was turned off briefly to allow coalition forces in the Persian Gulf to use civilian GPS receivers in 1990, but was turned on again immediately.

For the code phase receiver owner limited to pseudorange positioning, selective availability was a problem. A single pseudorange receiver could only achieve positional accuracy of about ±100 meters 95% of the time. This sort of single point positioning is not a usual surveying application of GPS. But with a second code phase receiver as a base station on a known position and using differential data processing techniques, submeter accuracy was possible even with SA turned on. The carrier phase receivers used in most surveying applications have always been, for all practical purposes, immune from the effects of SA.

2 Biases and Solutions

THE ERROR BUDGET

A Look at the Biases in the Observation Equations

The management of errors is indispensable for finding the true geometric range ρ from either a pseudorange, or carrier phase observation.

$$p = \rho + d_{\rho} + c\left(dt - dT\right) + d_{ion} + d_{trop} + \varepsilon_{m_p} + \varepsilon_{p} \quad \text{(pseudorange)}$$

$$\phi = \rho + d_{\rho} + c(dt - dT) + \lambda N - d_{ion} + d_{trop} + \varepsilon_{m_{\phi}} + \varepsilon_{\phi} \quad \text{(carrier phase)}$$

Both equations include environmental and physical limitations called *range biases*.

The Biases

Among these biases are certain atmospheric errors; two such errors are the ionospheric effect, d_{ion}, and the tropospheric effect, d_{trop}. The tropospheric effect may be somewhat familiar to EDM users, even if the ionospheric effect is not. Other biases, clock errors symbolized by *(dt – dT)* and receiver noise, ε_p and ε_{φ}, multipath, ε_{mp} and $\varepsilon_{m\varphi}$, and orbital errors, d_{ρ}, are unique to satellite surveying methods. The objective here is to separate and quantify each of these biases.

User Equivalent Range Error

When each bias is expressed as a range itself, each quantity is known as a *user equivalent range error, UERE*. This expression, often used in GPS literature, is a convenient way to clarify the individual contributions of each bias to the overall biased range measurement (Figure 2.1).

It is worthwhile to point out that the of some of the biases such as those attributable to the atmosphere d_{ion}, d_{trop} and satellite orbits d_{ρ} vary directly with the increase and decrease in the length of the baselines between GPS receivers. Others such as those due to receiver noise, ε_p and ε_{φ}, multipath, ε_{mp} and $\varepsilon_{m\varphi}$, do not.

The Satellite Clock Bias, *dt*

One of the largest errors can be attributed to the satellite clock bias. It can be quite large especially if the broadcast clock correction is not used by the receiver to bring the time signal acquired from a satellite's onboard clock in line with GPS time.

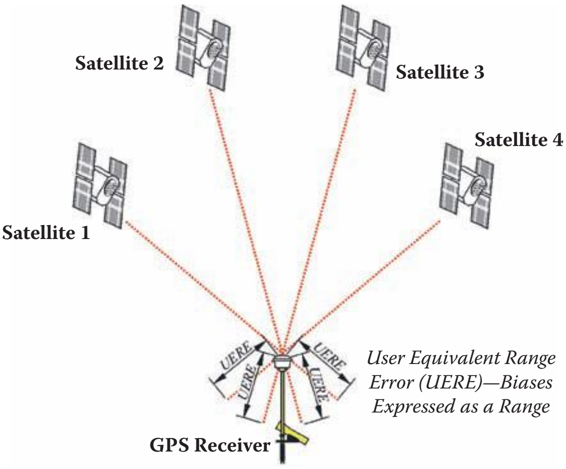


FIGURE 2.1 User Equivalent Range Error

Relativistic Effects on the Satellite Clock

Albert Einstein's special and general theories of relativity predicted that a clock in orbit around the earth would appear to run faster than a clock on its surface. And they do indeed: due to their greater speed and the weaker gravity around them, the clocks in the GPS satellites do appear to run faster than the clocks in GPS receivers. There are actually two parts to the effect.

Concerning the first part, time dilation is taken into account before the satellites' clocks are sent into orbit. To ensure the clocks will actually achieve the correct fundamental frequency of 10.23 MHz in space, their frequency is set a bit slow before launch to 10.22999999545 MHz.

The second part is attributable to the eccentricity of the orbit of GPS satellites. With an eccentricity of 0.02 this effect can be as much as 45.8 nanoseconds. Fortunately, the offset is eliminated by a calculation in the GPS receiver itself, thereby avoiding a ranging error of about 14 meters.

Both relativistic effects on the satellite clocks can be accurately computed and are removed from the system.

Satellite Clock Drift

Clock drift is another matter. As discussed in Chapter 1, the onboard satellite clocks are independent of one another. The rates of these rubidium, cesium, and hydrogen maser oscillators are more stable if they are not disturbed by frequent tweaking and adjustment is kept to a minimum. While GPS time itself is designed to be kept within one microsecond, *1 μsec or one-millionth of a second,* of UTC, excepting leap seconds, the satellite clocks can be allowed to drift up to a millisecond, *1 msec or one-thousandth of a second,* from GPS time.

There are three kinds of time are involved here. The first is UTC per the U.S. Naval Observatory (USNO). The second is GPS time. The third is the time determined by each independent GPS satellite.

Their relationship is as follows. The Master Control Station (MCS) at Schriever (formerly Falcon) Air Force Base near Colorado Springs, Colorado gathers the GPS satellites' data from monitoring stations around the world. After processing, this information is uploaded back to each satellite to become the broadcast ephemeris, broadcast clock correction and so forth. The actual specification for GPS time demands that it be within one microsecond of UTC as determined by USNO, without consideration of leap seconds. Leap seconds are used to keep UTC correlated with the actual rotation of the earth, but they are ignored in GPS time. In GPS time it is as if no leap seconds have occurred at all in UTC since 24:00:00, January 5, 1980. And in practice, GPS time is much closer than the microsecond specification, it is usually within about 25 nanoseconds, *25 nsec or 25 billionths of a second*, of UTC, minus leap seconds.

By constantly monitoring the satellites clock error, *dt,* the Control Segment gathers data for its uploads of the broadcast clock corrections. You will recall that clock corrections are part of the Navigation message.

THE IONOSPHERIC EFFECT, d_{ion}

Another of the largest errors in GPS positioning is attributable to the atmosphere. The relatively unhindered travel of the GPS signal through the virtual vacuum of space changes as it passes through the earth's atmosphere. Through both refraction and diffraction, the atmosphere alters the apparent speed and, to a lesser extent, the direction of the signal.

Group and Phase Delay

The ionosphere is the first part of the atmosphere that the GPS the signal encounters. It extends from about 50 km to 1000 km above the earth's surface. The ionosphere has layers sometimes known as the mesosphere and thermosphere that are themselves composed of C, D, E, and F regions (Figure 2.2). The regions C and D only exist in the daylight hours, the period when the regions E and F are further divided into F1 and F2. Farthest out is the area known as the exosphere. All of these divisions are based on electron density, which decreases as one gets further from the earth. Here they will all be combined under the general term, the ionosphere. Traveling through this part of the atmosphere the most troublesome effects on the GPS signal are known as the *group delay* and the *phase delay*. They both alter the measured range. The magnitude of these delays is determined by the density and stratification of the ionosphere at the moment the signal passes through it.

Total Electron Content

As mentioned the electron density of the ionosphere changes with the number and dispersion of free electrons released when gas molecules are ionized by the sun's ultraviolet radiation. This density is often described as *total electron content* or *TEC,* a measure of the number of free electrons in a column through the ionosphere with a cross-sectional area of 1 square meter.

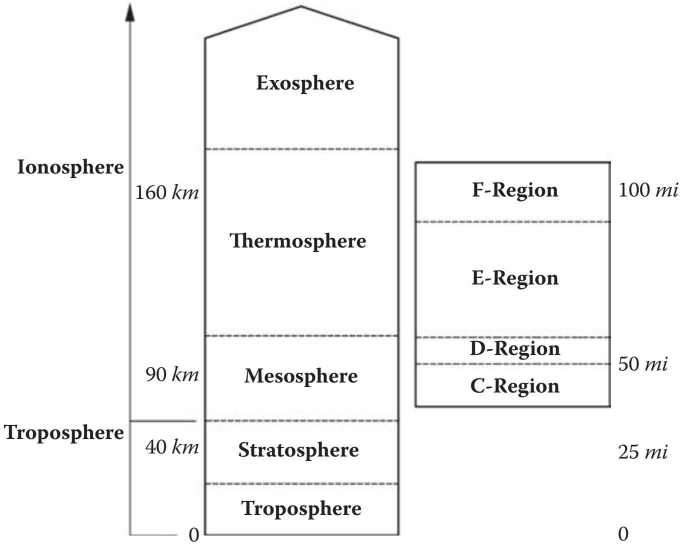


FIGURE 2.2 Atmospheric Model

Ionosphere and the Sun

The ionospheric delay changes slowly as it goes through a daily cycle. It is usually least between midnight and early morning and most around local noon or a little after. During the daylight hours in the midlatitudes the ionospheric delay may grow to be as much as five times greater than it was at night, but the rate of that growth is seldom more than 8 cm per minute. It is also nearly four times greater in November, when the earth is nearing its *perihelion*, its closest approach to the sun, than it is in July near the earth's *aphelion,* its farthest point from the sun. The effects of the ionosphere on the GPS signal usually reaches its peak in March, about the time of the vernal equinox.

Ionospheric Gradients

The ionosphere is not homogeneous. It changes from layer to layer within a particular area. Its behavior in one region of the earth is liable to be unlike its behavior in another. For example, ionospheric disturbances can be particularly harsh in the polar regions. But the highest TEC values and the widest variations in the horizontal gradients occur in the band of about 60° of *geomagnetic latitude.* That band lies 30° north and 30° south of the earth's magnetic equator.

Satellite Elevation and Ionospheric Effect

The severity of the ionospheric effect varies with the amount of time the GPS signal spends traveling through the layer. A signal originating from a satellite near the observer's horizon must pass through a larger amount of the ionosphere to reach the receiver than does a signal from a satellite near the observer's zenith. The longer the signal is in the ionosphere, the greater the ionosphere's effect, and the greater the impact of horizontal gradients within the layer.

The Magnitude of the Ionospheric Delay

The error introduced by the ionosphere can be very small, but it may be large when the satellite is near the observer's horizon, the vernal equinox is near, and sunspot activity is at its maximum. It varies with magnetic activity, location, time of day, and even the direction of observation.

The Ionosphere Affects Codes and the Carrier Differently

Fortunately, the ionosphere has a property that can be used to minimize its effect on GPS signals. It is *dispersive,* which means that the apparent time delay contributed by the ionosphere depends on the frequency of the signal. One result of this dispersive property is that during the signal's trip through the ionosphere, the codes, the modulations on the carrier wave, are affected differently than the carrier wave itself.

All the modulations on the carrier wave, the P code, the C/A code and the Navigation message, appear to be slowed. They are affected by the group delay. But the carrier wave itself appears to speed up in the ionosphere. It is affected by the phase delay.

It may seem odd to call an increase in speed a delay, but, governed by the same properties of electron content as the group delay, phase delay just increases negatively. Please note that the algebraic sign of d_{ion} is negative in the carrier phase equation and positive in the pseudorange equation.

Different Frequencies Are Affected Differently

Another consequence of the *dispersive* nature of the ionosphere is that the apparent time delay for a higher frequency carrier wave is less than it is for a lower frequency wave. That means that L1, 1575.42 MHz, is not affected as much as L2, 1227.60 MHz.

This fact provides one of the greatest advantages of a dual-frequency receiver over the single-frequency receivers because the 28% separation between the L1 and L2 frequencies of 347.82 MHz is large enough to facilitate estimation of the ionospheric group delay. Therefore, by tracking both carriers, a dual-frequency receiver has the facility of modeling and removing not all but a significant portion of the ionospheric bias.

The frequency dependence of the ionospheric effect is described by the following expression (Klobuchar 1983 in Brunner and Welch, 1993):

$$v = \frac{40.3}{cf^2} * TEC$$

where

v = the ionospheric delay
c = the speed of light in meters per second
f = the frequency of the signal in Hz
TEC = the quantity of free electrons per cubic meter

As the formula illustrates, the time delay is inversely proportional to the square of the frequency. In other words, the higher the frequency, the less the delay. Hence the dual-frequency receiver's capability to discriminate the effect on L1 from that on L2. Such a dual-frequency model can be used to reduce the ionospheric bias. Still, it is far from perfect and cannot ensure the effect's elimination.

Broadcast Correction

As mentioned in Chapter 1, an ionospheric correction is also available to the single frequency receiver in subframe 4 of the Navigation message. However, this broadcast correction should not be expected to remove more than about three-quarters of the ionospheric effect. It is important to note as mentioned earlier that where there is a short baseline between GPS receivers the effect of the ionosphere can be small, but as the baseline grows so does the effect.

THE RECEIVER CLOCK BIAS, dT

The third largest error can be caused by the receiver clock, which is its oscillator. Both a receiver's measurement of phase differences and its generation of replica codes depend on the reliability of this internal frequency standard.

Typical Receiver Clocks

GPS receivers are usually equipped with quartz crystal clocks, which are relatively inexpensive and compact. They have low power requirements and long life spans. For these types of clocks, the frequency is generated by the piezoelectric effect in an oven-controlled quartz crystal disk, a device sometimes symbolized by *OCXO*. Their reliability ranges from a minimum of about 1 part in 10^8 to a maximum of about 1 part in 10^{10}, a drift of about 0.1 nanosecond in 1 second. Even at that, quartz clocks are not as stable as the atomic standards in the GPS satellites and are more sensitive to temperature changes, shock, and vibration. Some receiver designs augment their frequency standards by also having the capability to accept external timing from cesium or rubidium oscillators.

THE ORBITAL BIAS, d_ρ

Orbital bias has the potential to be the fourth largest error. It is addressed in the broadcast ephemeris.

Forces Acting on the Satellites

The orbital motion of GPS satellites is not only a result of the earth's gravitational attraction; there are several other forces that act on the satellite. The primary disturbing forces are the non-spherical nature of the earth's gravity, the attractions of the sun and the moon, and solar radiation pressure. The best model of these forces is the actual motion of the satellites themselves and the government facilities distributed around the world, known collectively as the *Control Segment* or the *Operational Control System (OCS)*, track them for that reason, among others.

Tracking Facilities

The GPS system requires maintenance. Orbital and clock adjustments and other data uploads are necessary to keep the constellation from degrading. Every GPS satellite is tracked by monitoring stations and the orbital tracking data gathered are then passed on to the Master Control Station. There, new ephemerides are computed. This tabulation of the anticipated locations of the satellites with respect to time is then transferred to four uploading stations, where it is transmitted back to the satellites themselves.

The Master Control Station, *MCS,* is located at Schriever (formerly Falcon) Air Force Base in Colorado Springs, Colorado. The station is manned by the 2nd Space Operations Squadron. It computes updates for the Navigation message, generally, and the broadcast ephemeris, in particular, based on tracking information collected from monitoring stations around the world. All the U.S. Air Force installations have antennas for communicating with the GPS satellites except the station in Hawaii and the MCS. Other than the MCS, the monitoring stations at Diego Garcia, Kwajalein, Ascension, and Hawaii track up to 11 satellites simultaneously and send the data back to the MCS for processing in real-time. The station at Cape Canaveral is also used to check satellites before launch.

In the past this system had some limitations. For example, it was possible for a particular satellite to go unmonitored for up to two hours each day. It was known that the calculation of the ephemeris, the precise orbits, of the constellation would be improved by a wider geographical distribution than was available from the six OCS stations. And should one of the six OCS stations go down the effectiveness of the Control Segment could be considerably hampered. Therefore improvements known as the *Legacy Accuracy Improvement Initiative, L-AII,* were undertaken. (See Figure 2.3.)

During this initiative from August 18 to September 7 of 2005, six *National Geospatial Intelligence Agency, NGA,* stations were added to the Control Segment. This augmented the information forwarded to the MCS from the original six OCS stations with data from Washington, D.C., England, Argentina, Ecuador, Bahrain, and Australia. With this 12-station network in place every satellite in the GPS constellation is now monitored almost continuously from at least two stations. That is presuming that the satellite is tracked when it reaches at least 5° above the horizon. This ratio will increase to every satellite being monitored from three stations when five more NGA sites are added in the future. The new NGA stations will also improve the geographical diversity of the Control Segment, which improves the separation of clock from ephemeris errors as well as error isolation in general. In other words, redundant observations of a satellite anomaly from multiple stations provides increased certainty to the diagnosis and solution of problems.

Testing has shown that the augmented Control Segment and subsequent improved modeling has improved the accuracy of clock corrections and ephemerides in the Navigation Message substantially. These improvements may contribute to an increase in the accuracy of real-time GPS of 15% or more.

Another upgrade known as the *Architecture Evolution Plan, AEP* is also planned, which will include new functionality to support the coming Block IIF satellites and new signals.

FIGURE 2.3 Control Segment

GPS satellites are stabilized by *wheels,* a series of gyroscopic devices. Left alone these devices would continually accelerate until they were useless. Fortunately, the excess momentum is periodically dumped on instructions from the Control Segment.

The Tropospheric Effect, d_{trop}

The fifth largest UERE can be attributed to the effect of the troposphere.

Troposphere

The troposphere is that part of the atmosphere closest to the earth. In fact, it extends from the surface to about 9 km over the poles and 16 km over the equator, but in this work it will be combined with the tropopause and the stratosphere, as it is in much of GPS literature. Therefore, the following discussion of the tropospheric effect will include the layers of the earth's atmosphere up to about 50 km above the surface.

Tropospheric effect is independent of frequency. The tropospheric delay appears to add a slight distance to the range the receiver measures between itself and the satellite. The troposphere and the ionosphere are by no means alike in their effect on the satellite's signal. The troposphere is nondispersive for frequencies below 30 GHz. In other words, the refraction of a GPS satellite's signal is not related to its frequency in the troposphere.

The troposphere is part of the electrically neutral layer of the earth's atmosphere, meaning it is neither ionized nor dispersive. Therefore, both L1 and L2 are equally refracted. Like the ionosphere, the density of the troposphere also governs the severity of its effect on the GPS signal. For example, when a satellite is close to the horizon, the delay of the signal caused by the troposphere is maximized. The tropospheric delay of the signal from a satellite at zenith, directly above the receiver, is minimized.

Satellite elevation and tropospheric effect. The situation is analogous to atmospheric refraction in astronomic observations; the effect increases as the energy passes through more of the atmosphere. The difference in GPS is that it is the delay, not the angular deviation, caused by the changing density of the atmosphere that is of primary interest. The GPS signal that travels the shortest path through the troposphere will be the least delayed by it. So, even though the delay at an elevation angle of 90° at sea level will only be about 2.4 meters, it can increase to about 9.3 meters at 75° and up to 20 meters at 10°. There is less tropospheric delay at higher altitudes.

Modeling. Modeling the troposphere is one technique used to reduce the bias in GPS data processing, and it can be up to 95% effective. However, the residual 5% can be quite difficult to remove. For example, surface measurements of temperature and humidity are not strong indicators of conditions on the path between the receiver and the satellite. But instruments that can provide some idea of the conditions along the line between the satellite and the receiver are somewhat more helpful in modeling the tropospheric effect. Several models have been developed, including the Saastamoinen model and the Hopfield model. They perform well at reasonably high elevation angles. However, it is advisable to limit GPS observations to those signals above 15% or so to ameliorate the effects of atmospheric delay.

The dry and wet components of refraction. Refraction in the troposphere has a dry component and a wet component. The dry component which contributes most of the delay, perhaps 80% to 90%, is closely correlated to the atmospheric pressure. The dry component can be more easily estimated than the wet component. It is fortunate that the dry component contributes the larger portion of range error in the troposphere since the high cost of water vapor radiometers and radiosondes generally restricts their use to only the most high-precision GPS work.

Receiver spacing and the atmospheric biases. There are other practical consequences of the atmospheric biases. For example, the character of the atmosphere is never homogeneous and the importance of atmospheric modeling increases as the distance between GPS receivers grows. Consider a signal traveling from one satellite to two receivers that are close together. That signal would be subjected to very similar atmospheric effects, and, therefore, atmospheric bias modeling would be less important to the accuracy of the measurement of the relative distance between them. But a signal traveling from the same satellite to two receivers that are far apart may pass through levels of atmosphere quite different from one another. In that case, atmospheric bias modeling would be more important.

MULTIPATH

The range delay known as *multipath* is symbolized by ε_{mp} in the pseudorange equation and $\varepsilon_{m\varphi}$ in the carrier phase equation. As the name implies it is the reception of the GPS signal via multiple paths rather than from a direct line of sight (see Figure 2.4). Multipath differs from both the apparent slowing of the signal through the ionosphere and troposphere and the discrepancies caused by clock offsets. The range delay in multipath is the result of the reflection of the GPS signal.

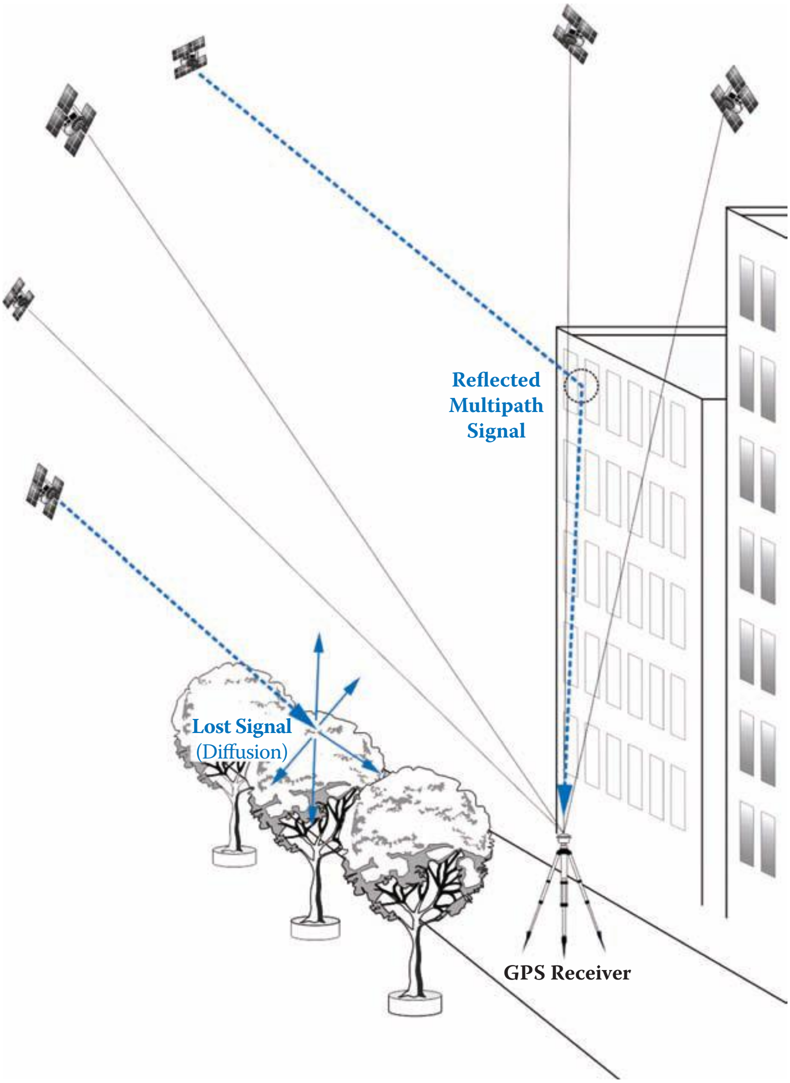


FIGURE 2.4 Multipath

The affect of multipath on pseudorange solutions is orders of magnitude larger than it is in carrier phase solutions. However, multipath in carrier phase is much harder to mitigate than multipath in pseudoranges. For example, the strategy of spacing the early, punctual and late correlators at 1/10th of a chip rather than the standard 1 chip is not nearly as effective in mitigating carrier phase multipath as it is in pseudorange solutions.

Multipath occurs when part of the signal from the satellite reaches the receiver after one or more reflections from the ground, a building, or another object. These reflected signals can interfere with the signal that reaches the receiver directly from the satellite and cause the correlation peak mentioned in Chapter 1 to become skewed.

Limiting the Effect of Multipath

The high frequency of the GPS codes tends to limit the field over which multipath can contaminate pseudorange observations. Once a receiver has achieved lock, that is, its replica code is correlated with the incoming signal from the satellite, signals outside the expected chip length can be rejected. Generally speaking multipath delays of less than one chip, those that are the result of a single reflection, are the most troublesome.

Fortunately, there are factors that distinguish reflected multipath signals from direct, line-of-sight, signals. For example, reflected signals at the frequencies used for L1 and L2 tend to be weaker and more diffuse than the directly received signals. Another difference involves the circular polarization of the GPS signal. The polarization is actually reversed when the signal is reflected, similar to the spin of a banked billiard ball. Reflected, multipath signals become *Left Hand Circular Polarized, LHCP,* whereas the signals received directly from the GPS satellites are *Right Hand Circular Polarized, RHCP.* But while the majority of multipath signals may be LHCP, it is possible for them to arrive at the received in-phase usually through an even number of multiple reflections. These characteristics allow some multipath signals to be identified and rejected at the receiver's antenna.

Antenna Design and Multipath

GPS antenna design can play a role in minimizing the effect of multipath. Ground planes, usually a metal sheet, are used with many antennas to reduce multipath interference by eliminating signals from low elevation angles. Generally, larger ground planes, multiple wavelengths in size, have a more stabilizing influence than smaller ground planes. However, such ground planes do not provide much protection from the propagation of waves along the ground plane itself. When a GPS signal's wave front arrives at the edge of an antenna's ground plane from below, it can induce a surface wave on the top of the plane that travels horizontally (see Figure 2.5).

Another way to mitigate this problem is the use of a choke ring antenna. Choke ring antennas, based on a design first introduced by the Jet Propulsion Laboratory (JPL), can reduce antenna gain at low elevations. This design contains a series of concentric circular troughs that are a bit more than a quarter of a wavelength deep. A choke ring antenna can prevent the formation of these surface waves.

But neither ground planes nor choke rings remove the effect of reflected signals from above the antenna very effectively. There are signal processing techniques that can reduce multipath, but when the GPS signal is reflected from ~20m or less from the antenna and is subject to a short delay, this approach is not as effective.

A widely used strategy is the 15° cutoff or *mask angle.* This technique calls for tracking satellites only after they are more than 15° above the receiver's horizon. Careful attention in placing the antenna away from reflective surfaces, such as nearby buildings, water or vehicles, is another way to minimize the occurrence of multipath.

RECEIVER NOISE

Multipath is an uncorrelated error, and as it is directly related to thermal noise, dynamic stress, and so on in the GPS receiver itself, receiver noise is also an

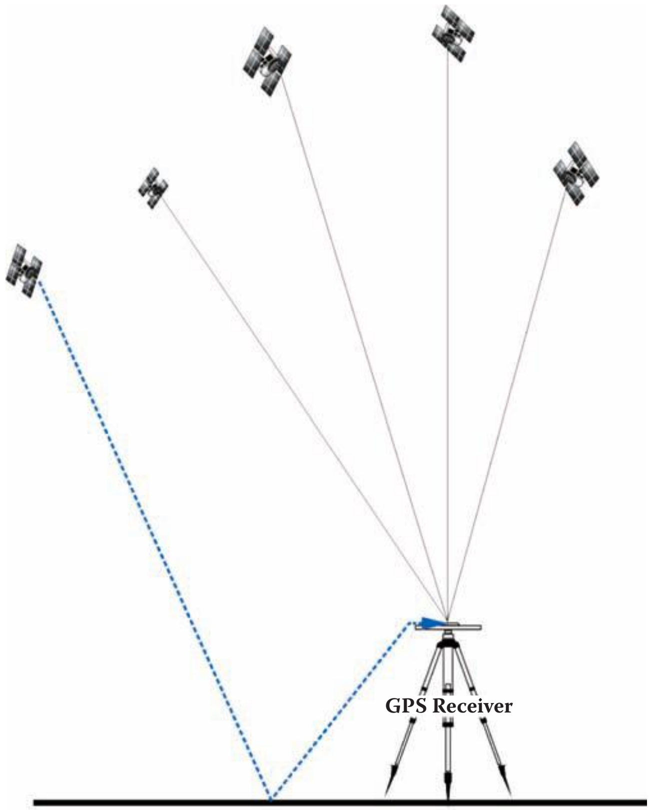


FIGURE 2.5 Multipath

uncorrelated error source. Fortunately it is a small bias symbolized by ε_p in the pseudorange equation and ε_φ in the carrier phase equation. Generally speaking the receiver noise error is about 1% of the wavelength of the signal involved. In other words in code solutions the size of the error is related to chip width. For example, the receiver noise error in a C/A code solution is about an order of magnitude more than a P code solution. And in carrier phase solutions the receiver noise error contributes millimeters to the overall error.

DIFFERENCING

CLASSIFICATIONS OF POSITIONING SOLUTIONS

Kinematic GPS

Speaking broadly, concerning carrier phase observations there are two applications of GPS in surveying and geodesy, kinematic and static positioning. Kinematic applications imply movement, one or more GPS receivers actually in motion during their observations. A moving GPS receiver on land, sea, or air is characteristic of kinematic GPS. Other characteristics of the application include results in real time and little redundancy. Hydrography, aerial mapping, gravimetric, and more and more land surveying projects are done using kinematic GPS. The kinematic GPS surveying used in land surveying today was originated by Dr. Benjamin Remondi in the 1980s.

Static GPS

Static applications use observations from receivers that are stationary for the duration of their measurement. Most static applications can afford higher redundancy and a bit more accuracy than kinematic GPS. The majority of GPS surveying control and geodetic work still relies on static applications.

Hybrid Techniques in GPS

There have been recent hybrids between kinematic and static applications. *Pseudokinematic* and *semikinematic* rely on short static observations and, in the first instance, revisits to positions once occupied; in the second, the alternate movement of multiple receivers.

Another hybrid is called *rapid static*. The receivers used in rapid static are stationary during their observations, but their observations are very short, and they rely on both code and carrier observations on both L1 and L2.

Relative and Point Positioning

Kinematic and static applications include two classifications *point,* or *autonomous positioning* and *relative positioning*. In other words, there are four possible combinations. They are: *static point positioning*, *static relative positioning*, *kinematic point positioning,* and *kinematic relative positioning.*

The term *differential* GPS, or *DGPS*, has come into common usage as well. Use of this acronym usually indicates a method of relative positioning where coded pseudorange measurements are used rather than carrier phase.

The Navigation Solution

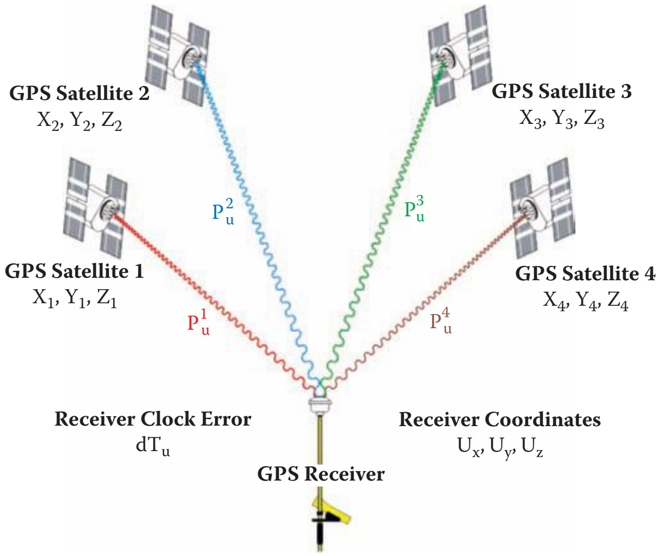
Another type of point positioning is known as *absolute positioning*, *single-point positioning,* or *the Navigation solution*. It is characterized by a single receiver measuring its range to a minimum of 4 satellites simultaneously (Figure 2.6).

The Navigation solution is in one sense the fulfillment of the original idea behind GPS. It relies on a coded pseudorange measurement and can be used for virtually instantaneous positioning.

In this method the positions of the satellites are available from the data in their broadcast ephemerides. The satellite clock offset and the ionospheric correction are also available from the Navigation messages of all 4 satellites.

Four Unknowns

Even if all these data are presumed to contain no errors, which they surely do, four unknowns remain: the position of the receiver in three Cartesian coordinates, u_x, u_y, and u_z , and the receiver's clock error dT_u. Three pseudoranges provide enough data



$$P_u^1 = \sqrt{(X_1-U_x)^2 + (Y_1-U_y)^2 + (Z_1-U_z)^2} + c(dT_u)$$

$$P_u^2 = \sqrt{(X_2-U_x)^2 + (Y_2-U_y)^2 + (Z_2-U_z)^2} + c(dT_u)$$

$$P_u^3 = \sqrt{(X_3-U_x)^2 + (Y_3-U_y)^2 + (Z_3-U_z)^2} + c(dT_u)$$

$$P_u^4 = \sqrt{(X_4-U_x)^2 + (Y_4-U_y)^2 + (Z_4-U_z)^2} + c(dT_u)$$

FIGURE 2.6 Navigation Solution

to solve for u_x, u_y, and u_z. And the fourth pseudorange provides the information for the solution of the receiver's clock offset.

The ability to achieve so much redundancy in the measurement of the dT, the receiver's clock error, is one reason the moderate stability of quartz crystal clock technology is entirely adequate as a receiver oscillator.

Four Satellites and Four Equations

A unique solution is found here because the number of unknowns is not greater than the number of observations. The receiver tracks 4 satellites simultaneously; therefore, these four equations can be solved simultaneously for every *epoch* of the observation. An epoch in GPS is a very short period of observation time, and is generally just a small part of a longer measurement. However, theoretically there is enough information in any single epoch of the Navigation solution to solve these equations. In fact, the trajectory of a receiver in a moving vehicle could be determined by this method. With 4 satellites available, resolution of a receiver's position and velocity are both available through the simultaneous solution of these four equations. These facts are the foundation of the kinematic application of GPS.

RELATIVE POSITIONING

Correlation

The availability of two or more GPS receivers makes relative positioning possible. Relative positioning can attain higher accuracy than point positioning because of the extensive correlation between observations taken to the same satellites at the same time from separate stations. Because the distance between such stations on the earth are short compared with the 20,000-km altitude of the GPS satellites, two receivers operating simultaneously, collecting signals from the same satellites, will record very similar errors.

Baselines

The vectors between such pairs of receivers are known as *baselines*. The simultaneity of observation, the resulting error correlation, and the carrier phase observable combine to yield typical baseline measurement accuracies of ± (1 cm + 2 ppm). It is possible for GPS measurements of baselines to be as accurate as 1 ppm or even 0.1 ppm. For example, a nine mile baseline correctly measured by GPS within ±0.05 ft. (1 ppm) to ±0.005 ft. (0.1 ppm).

Networks

Network and *multireceiver positioning* are obvious extensions of relative positioning. Both the creation of a closed network of points by combining individually observed baselines and the operation of three or more receivers simultaneously have advantages. For example, the baselines have redundant measurements and similar, if not identical, range errors. The processing methods in such an arrangement can nearly eliminate many of the biases introduced by imperfect clocks and the atmosphere. These processing strategies are based on computing the differences between simultaneous GPS carrier phase observations.

Differencing

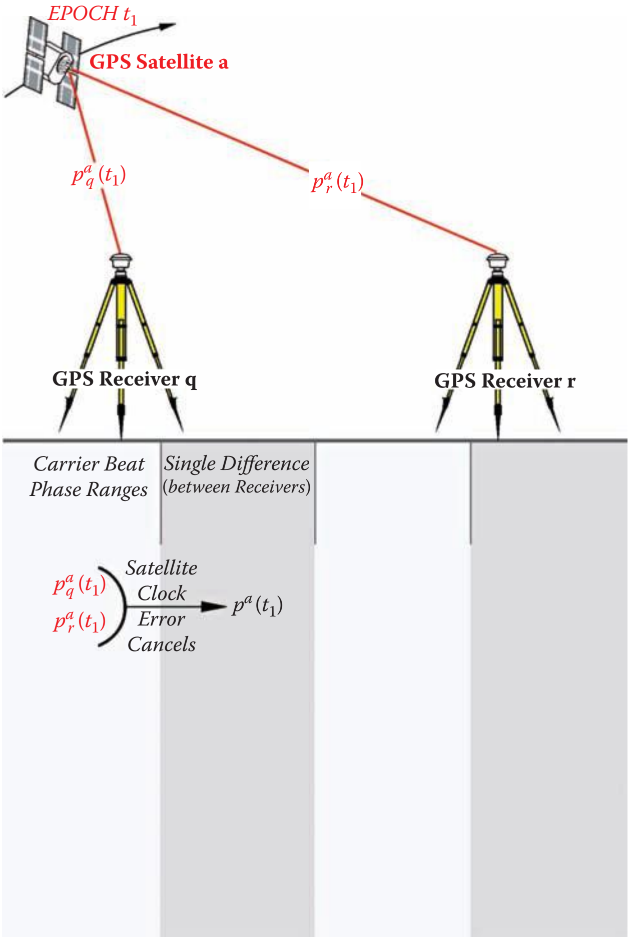
In GPS the word *differencing* has come to represent several types of simultaneous baseline solutions of combined measurements. The most frequently used are known as the *single difference*, *double difference,* and *triple difference*.

SINGLE DIFFERENCE

One of the foundations of differencing is the idea of the baseline as it is used in GPS. For example, a single difference, also known as a *between-receivers difference*, can refer to the difference in the simultaneous carrier phase measurements from one GPS satellite as measured by two different receivers (Figure 2.7).

Elimination of the Satellite Clock Errors

In this single-difference solution the satellite clock error is eliminated. The difference between dt at the first receiver and dt at the second receiver, Δdt, is zero. Since the two receivers are both observing the same satellite at the same time the satellite clock bias



Carrier Beat Phase Measurements from One GPS as Measured by Two Different Receivers

FIGURE 2.7 A between Receivers Single Difference is canceled. The atmospheric biases and the orbital errors recorded by the two receivers in this solution are nearly identical, so they too can also be virtually eliminated.

Other Errors Remain

Unfortunately, there are still two factors in the carrier beat phase observable that are not eliminated by single differencing. The difference between the integer cycle ambiguities at each receiver, ΔN, and the difference between the receiver clock errors, ΔdT, remain.

DOUBLE DIFFERENCE

Elimination of the Receiver Clock Errors

There is a GPS solution that will eliminate the receiver clock errors. It involves the addition of what might be called another kind of single difference, also known as a between-satellites difference. This term refers to the difference in the measurement of signals from 2 GPS satellites, as measured simultaneously at a single receiver (Figure 2.8).

The data available from the between-satellites difference allow the elimination of the receiver clock error. In this situation, there can be no difference in the clock since only one is involved. And the atmospheric effects on the 2 satellite signals are again nearly identical as they come into the lone receiver, so the effects of the ionospheric and tropospheric delays are virtually eliminated as well.

No Clock Errors at All

By using both the between-receivers difference and the between-satellites difference, a double difference is created. This combination is virtually free of receiver clock errors. Therefore, for all practical purposes, the double difference does not contain receiver clock errors or satellite clock errors (Figure 2.9).

However, the disadvantage with differencing is that the noise is increased by a factor of 2 with each difference operation. Nevertheless, the advantages far outweigh the disadvantages and double differencing is commonly used.

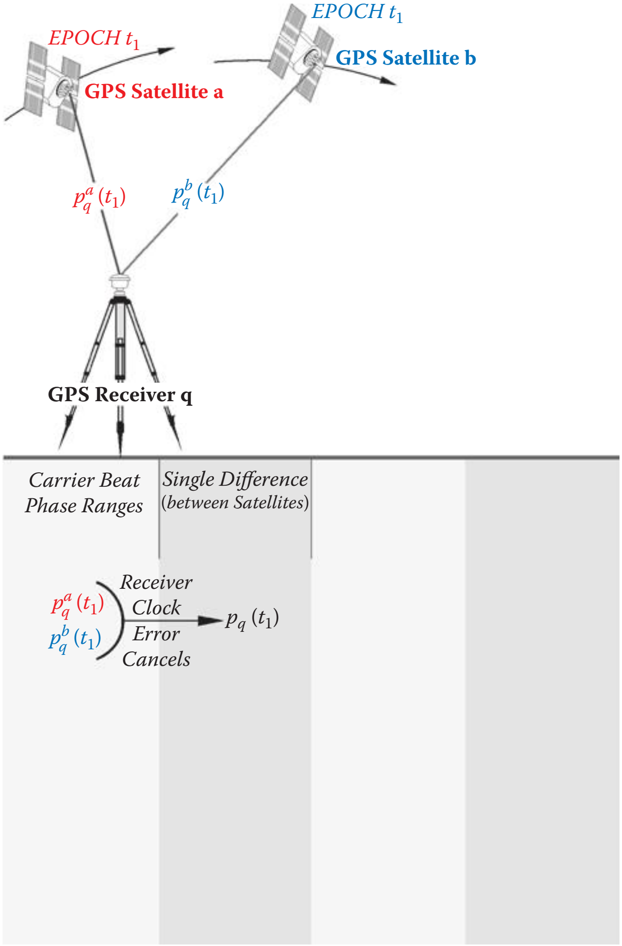
Still there is one stubborn factor in the carrier beat phase observable that is not eliminated: the integer cycle, or phase ambiguity, N.

TRIPLE DIFFERENCE

A third kind of differencing is created by combining two double differences. Each of the double differences involves 2 satellites and two receivers. The difference next derived is between two epochs. The triple difference is also known as the *receiver-satellite-time triple difference* (Figure 2.10), the difference of two double differences of two different epochs.

Integer Cycle Ambiguity

In the triple difference two receivers observe the same 2 satellites during two consecutive epochs. This solution can be used to quantify the integer cycle ambiguity,



Carrier Beat Phase Measurements from Two GPS Satellites, as Measured at a Single Receiver

FIGURE 2.8 A between Satellites Single Difference

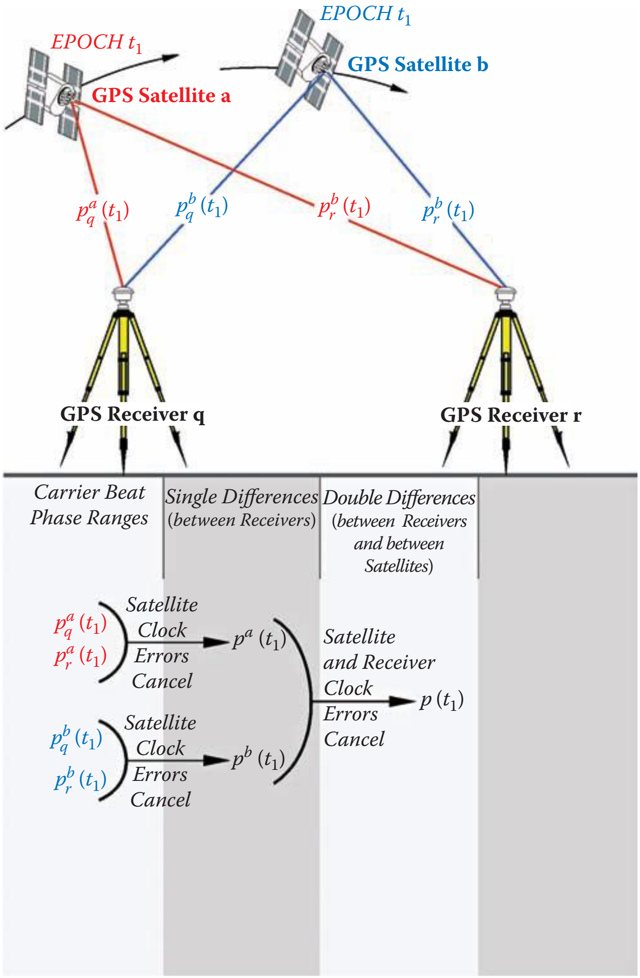


FIGURE 2.9 A Double Difference

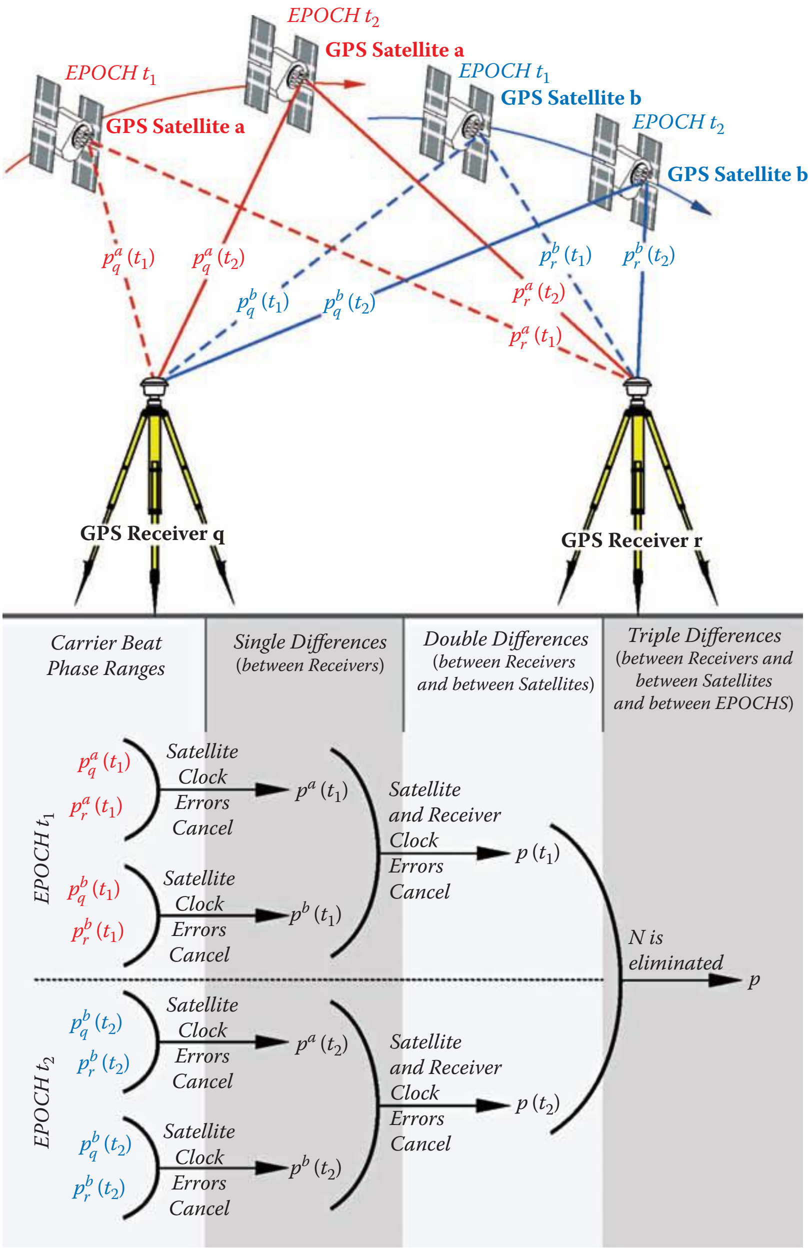

EPOCH t_2
GPS Satellite a
EPOCH t_1
GPS Satellite a
EPOCH t_1
GPS Satellite b
EPOCH t_2
GPS Satellite b
$p_q^a(t_1)$
$p_q^a(t_2)$
$p_q^b(t_1)$
$p_q^b(t_2)$
$p_r^b(t_1)$
$p_r^b(t_2)$
$p_r^a(t_2)$
$p_r^a(t_1)$
GPS Receiver q
GPS Receiver r
Carrier Beat Phase Ranges
Single Differences (between Receivers)
Double Differences (between Receivers and between Satellites)
Triple Differences (between Receivers and between Satellites and between EPOCHS)
EPOCH t_1
$p_q^a(t_1)$
$p_r^a(t_1)$
Satellite Clock Errors Cancel
$p^a(t_1)$
$p_q^b(t_1)$
$p_r^b(t_1)$
Satellite Clock Errors Cancel
$p^b(t_1)$
Satellite and Receiver Clock Errors Cancel
$p(t_1)$
N is eliminated
p
EPOCH t_2
$p_q^b(t_2)$
$p_r^b(t_2)$
Satellite Clock Errors Cancel
$p^a(t_2)$
$p_q^b(t_2)$
$p_r^b(t_2)$
Satellite Clock Errors Cancel
$p^b(t_2)$
Satellite and Receiver Clock Errors Cancel
$p(t_2)$

FIGURE 2.10

N, because if all is as it should be, N is constant over the two observed epochs. Therefore, the triple difference makes the detection and elimination of cycle slips relatively easy. Actually, a triple difference is not sufficiently accurate for short baselines. It is used to find the integer cycle ambiguity. Once the cycle ambiguity is determined it can be used with the double difference solution to calculate the actual carrier phase measurement.

Cycle slips. A cycle slip is a discontinuity in a receiver's continuous phase lock on a satellite's signal. The coded pseudorange measurement is immune from this difficulty, but the carrier beat phase is not.

Components of the carrier phase observable. From the moment a receiver locks onto a particular satellite, there are actually three components to the total carrier phase observable.

$$\varphi = \alpha + \beta + N$$

where

φ = total phase
α = fractional initial cycle
β = observed cycle count
N = cycle count at lock on

Fractional initial phase. First is the *fractional initial phase,* which occurs at the receiver at the first instant of the lock-on. The receiver starts tracking the incoming phase from the satellite. It cannot yet know how to achieve a perfect synchronization. Lacking this knowledge the receiver grabs onto the satellite's signal at some fractional part of a phase. It is interesting to note that this fractional part does not change for the duration of the observation and so is called the fractional initial phase. It is symbolized in the equation above by α.

The integer number of cycles. Second is the integer number of full cycles of phase that occur from the moment of the lock to the end of the observation. It is symbolized by β, the observed cycle count. This element is the receiver's consecutive counting of the change in full phase cycles, 1, 2, 3, 4 . . ., between the receiver and the satellite as the satellite flies over. Of the three terms, β is only number that changes—that is, if the observation proceeds correctly.

The cycle ambiguity. Third is the integer cycle ambiguity *N*. It represents the number of full phase cycles between the receiver and the satellite at the first instant of the receiver's lock-on. It is labeled as the cycle count at lock-on. N does not change from the moment of the lock onward, unless that lock is lost.

In other words, the total carrier phase observable consists of two values that do not change during the observation, the fractional phase α, and the integer cycle ambiguity N. Only the observed cycle count β changes, unless there is a cycle slip.

Lost lock. When lock is lost a cycle slip occurs. A power loss, an obstruction, a very low signal-to-noise ratio, or any other event that breaks the receiver's continuous reception of the satellite's signal causes a cycle slip. That is, the receiver loses its place in its count of the integer number of cycles β and, as a result, N is completely lost.

Finding and fixing cycle slips. There are several methods that may be used to regain a lost integer phase value, N. The triple difference is one of the better alternatives in this regard; as stated earlier the triple difference, does not depend on the initial integer ambiguity, because it is a constant in time. Therefore, when a large residual does appear in its component double differences it is very likely caused by a cycle slip. Even better, the obstructed signal can be singled-out by isolating all available satellite pairs until the problem is found. This utility in fixing cycle slips is the primary appeal of the triple difference. It can be used as a preprocessing step to weed out cycle slips and provide a first position for the receivers.

SUMMARY

Typical Techniques

Relative positioning by carrier beat phase measurement is the primary vehicle for high-accuracy GPS surveying. Simultaneous static observations, double differencing in postprocessing, and the subsequent construction of networks from GPS baselines are the hallmarks of geodetic and control work in the field. The strengths of these methods generally outweigh their weaknesses, particularly where there can be an unobstructed sky and relatively short baselines and where the length of observation sessions is not severely restricted.

Pseudorange and Carrier Phase

However, conditions are not always so ideal. Where obstructions threaten to produce cycle slips, coded pseudorange measurements may offer an important advantage over carrier beat phase. Pseudorange measurements also may be preferred where accuracy requirements are low and production demands are high.

Kinematic GPS

There can hardly be a question that kinematic GPS is the most productive of the several alternative methods, under the right circumstances. However, the necessity of maintaining lock on four or more satellites as the receiver is moved currently limits its application to very open areas.

Other Techniques

Rapid-static, pseudokinematic, and other hybrid methods are attempts to take advantage of some of the best aspects of static positioning, such as high accuracy and predictable production, while improving on its drawbacks. It is almost always desirable to increase production, by employing shortened observation sessions, provided it can be done while maintaining the required accuracy.

Differencing

Differencing is an ingenious approach to minimizing the effect of errors in carrier beat phase ranging. It is a technique that largely overcomes the impossibility of perfect time synchronization. Double differencing is the most widely used formulation. Double differencing still contains the initial integer ambiguities, of course. And the estimates of the ambiguities generated by the initial processing are usually not integers; in other words, some orbital errors and atmospheric errors remain. But with the knowledge that the ambiguities ought to be integers, during subsequent processing it is possible to force estimates for the ambiguities that are in fact integers. When the integers are so fixed, the results are known as a *fixed solution*, rather than a *float solution*. It is the double differenced carrier phase based fixed solution that makes the very high accuracy possible with GPS.

Biases

However, in this discussion of errors it is important to remember that multipath, cycle slips, incorrect instrument heights, and a score of other errors whose effects can be minimized or eliminated by good practice are simply not within the purview of differencing at all. The unavoidable biases that can be managed by differencing—including clock, atmospheric, and orbital errors—can have their effects drastically reduced by the proper selection of baselines, the optimal length of the observation sessions, and several other considerations included in the design of a GPS survey. But such decisions require an understanding of the sources of these biases and the conditions that govern their magnitudes. The adage of, “garbage in, garbage out,” is as true of GPS as any other surveying procedure. The management of errors cannot be relegated to mathematics alone.

EXERCISES

1. The earth’s atmosphere affects the signals from GPS satellites as they pass through. Which of these statements about the phenomenon is true?

   (a) The GPS signals are refracted by the ionosphere. Considering pseudoranges, the apparent length traveled by a GPS signal seems to be too short, and in the case of the carrier wave it seems to be too long.
   (b) The ionosphere is dispersive. Therefore, despite the number of charged particles the GPS signal encounters on its way from the satellite to the receiver, the time delay is directly proportional to the square of the transmission frequency. Lower frequencies are less affected by the ionosphere.
   (c) The ionosphere is a region of the atmosphere where free electrons do not affect radio wave propagation, due mostly to the effects of solar ultraviolet and x-ray radiation.
   (d) The TEC depends on the user’s location, observing direction, the magnetic activity, the sunspot cycle, the season, and the time of day, among other things. In the midlatitudes the ionospheric effect is usually least between midnight and early morning and most at local noon.

2. The clocks in GPS satellites are rubidium and cesium oscillators, whereas quartz crystal oscillators provide the frequency standard in most GPS receivers. Concerning the relationship between these standards, which of the following statements is **not** true?

   (a) Quartz clocks are not as stable as the atomic standards in the GPS satellites and are more sensitive to temperature changes, shock, and vibration.
   (b) Both GPS receivers and satellites rely on their oscillators to provide a stable reference so that other frequencies of the system can be generated from or compared with them.
   (c) The foundation of the oscillators in GPS receivers and satellites is the piezoelectric effect.
   (d) The oscillators in the GPS satellites are also known as atomic clocks.

3. What is an advantage available using a dual frequency GPS receiver that is **not** available using a single frequency GPS receiver?

   (a) A single frequency GPS receiver cannot collect enough data to perform single, double, or triple difference solutions.
   (b) A dual-frequency receiver affords an opportunity to track the P code, but a single frequency receiver cannot.
   (c) A dual frequency receiver has access to the Navigation code, but a single frequency receiver does not.
   (d) Over long baselines, a dual-frequency receiver has the facility of modeling and virtually removing the ionospheric bias, whereas a single frequency receiver cannot.

4. Which of the following statements is **not** correct concerning refraction of the GPS signal in the troposphere?

   (a) L1 and L2 carrier waves are refracted equally.
   (b) When a GPS satellite is near the horizon, its signal is most affected by the atmosphere.
   (c) The density of the troposphere governs the severity of its effect on a GPS signal.
   (d) The wet component of refraction in the troposphere contributes the larger portion of the range error.

5. In a between-receivers single difference across a short baseline, which of the following problems are **not** virtually eliminated?

   (a) satellite clock errors
   (b) atmospheric bias
   (c) integer cycle ambiguities
   (d) orbital errors

6. In a double-difference across a short baseline, which of the following problems are **not** virtually eliminated?

   (a) atmospheric bias
   (b) satellite clock errors
   (c) receiver clock errors
   (d) integer cycle ambiguity

7. Which of the following statements would **not** correctly complete the sentence beginning, “If cycle slips occurred in the observations…?”

   (a) “…they could have been caused by intermittent power to the receiver.”
   (b) “…they might be detected by triple differencing.”
   (c) “…they are equally troublesome in pseudorange and carrier phase measurements.”
   (d) “…it is best if they are repaired before double differencing is done.”

8. What is the correct value for the maximum time interval allowed between GPS time and a particular GPS satellite’s onboard clock?

   (a) one nanosecond
   (b) one millisecond
   (c) one microsecond
   (d) one femtosecond

9. Which of the following biases are **not** mitigated by using relative positioning GPS or DGPS?

   (a) the ionospheric effect
   (b) the tropospheric effect
   (c) multipath
   (d) satellite clock errors

10. Which of the following does the Control Segment **not** compile for upload to the GPS satellites?

    (a) almanac information
    (b) ephemeris information
    (c) satellite clock corrections
    (d) P and C/A codes

ANSWERS AND EXPLANATIONS

1. Answer is (d)

Explanation: The modulations on radio carrier waves encountering the free electrons in the ionosphere are retarded, but the carrier wave itself is advanced. The slowing is known as the ionospheric delay. And in the case of pseudoranges the apparent length of the path of the signal is stretched; it seems to be too long. However, when considering the carrier wave the path seems to be too short. It is interesting that the absolute value of these apparent changes is almost exactly equivalent.

The ionosphere is dispersive. Refraction in the ionosphere is a function of a wave's frequency. The higher the frequency, the less the refraction, and the shorter the delay in the case of modulations on the carrier wave. The ionospheric effect is proportional to the inverse of the frequency squared.

The magnitude of the ionospheric effect increases with electron density. Also known as the total electron content or TEC, the electron density varies in both time and space. However, in the midlatitudes the minimum most often occurs from midnight to early morning and the maximum near local noon.

2. Answer is (c)

Explanation: Rubidium and cesium atomic clocks are aboard the current constellation of GPS satellites. Their operation is based on the resonant transition frequency of the Rb-87 and CS-133 atoms, respectively. On the other hand, the quartz crystal oscillators in GPS receivers utilize the piezoelectric effect. Both GPS receivers and satellites rely on their oscillators to provide a stable reference so that other frequencies of the system can be generated from or compared with them.

3. Answer is (d)

Explanation: Sufficient data to calculate single, double, and triple differences can be available from both single and dual frequency receivers. The permission to track the P code is not restricted by single or dual frequency capability, but national security. GPS receivers have access to the Navigation code whether they track L1, L2, or both.

Using single frequency receivers ionospheric corrections can be computed from an ionospheric model. GPS satellites transmit coefficients for ionospheric corrections that most receivers' software use. The models assume a standard distribution of the total electron count; still, with this strategy about a quarter of the ionosphere's actual variance will be missed. But when receivers are close together, say 30 kilometers or less, for all practical purposes the ionospheric delay and carrier phase advance is the same for both receivers. Therefore, phase difference observations over short baselines are little affected by ionospheric bias with single frequency receivers.

However, that is not the case over long baselines, of say 100 kilometers. Dual frequency receivers can utilize the dispersive nature of the ionosphere to overcome the effects. The resulting time delay is inversely proportional to the square of the transmission frequency. That means that L1, 1575.42 MHz, is not affected as much as L2, 1227.60

MHz. By tracking both carriers, a dual-frequency receiver has the facility of modeling and virtually removing much of the ionospheric bias.

4. Answer is (d)

Explanation: The tropospheric effect is non-dispersive for frequencies under 30 GHz. Therefore, it affects both L1 and L2 equally. Refraction in the troposphere has a dry component and a wet component. The dry component is related to the atmospheric pressure and contributes about 90% of the effect. It is more easily modeled than the wet component. The GPS signal that travels the shortest path through the atmosphere will be the least affected by it. Therefore, the tropospheric delay is least at the zenith and most near the horizon. GPS receivers at the ends of short baselines collect signals that pass through substantially the same atmosphere and the tropospheric delay may not be troublesome. However, the atmosphere may be very different at the ends of long baselines.

5. Answer is (c)

Explanation: If two or more receivers collect carrier-phase observations from a constellation of satellites, observations that are subsequently loaded into a computer for processing, the first step is usually computation of single differences for each epoch. A between-receivers single difference is the difference in the simultaneous carrier phase measurements from one GPS satellite as measured by two different receivers during a single epoch. Single differences are virtually free of satellite clock errors. The atmospheric biases and the orbital errors recorded by the two receivers in this solution are nearly identical, so they too can be virtually eliminated. However, processing does not usually end at single differences in surveying applications because the difference between the integer cycle ambiguities at each receiver and the difference between the receiver clock errors remain in the solution.

6. Answer is (d)

Explanation: The differences of two single differences in the same epoch using two satellites is known as a double difference. The combination is virtually free of receiver clock errors, satellite clock errors, and over short baselines, orbital and atmospheric biases. However, the integer cycle ambiguities remain.

7. Answer is (c)

Explanation: A cycle slip is a discontinuity in a receiver's continuous phase lock on a satellites signal. When lock is lost, a cycle slip occurs. A power loss, an obstruction, a very low signal-to-noise ratio, or any other event that breaks the receiver's continuous reception of the satellite's signal causes a cycle slip. The coded pseudorange measurement is immune from this difficulty, but carrier phase measurements are not. It is best if cycle slips are removed before the double difference solution. When a large residual appears in the component double differences of a triple difference it is very likely caused by a cycle slip. This utility in finding and fixing cycle slips is the primary appeal of the triple difference. It can be used as a preprocessing step to weed out cycle slips and provide a first position for the receivers.

8. Answer is (b)

The GPS time system is composed of the Master Control Clock and the GPS satellite clocks and is measured from 24:00:00, January 5, 1980. But, while GPS time itself is kept within one microsecond of UTC, excepting leap seconds, the satellite clocks can be allowed to drift up to a millisecond from GPS time. The rates of these onboard rubidium and cesium oscillators are more stable if they are not disturbed by frequent tweaking and adjustment is kept to a minimum.

9. Answer is (c)

Explanation: Two or more GPS receivers make relative positioning and DGPS possible. These techniques can attain high accuracy of the extensive correlation between observations taken to the same satellites at the same time from separate stations. The distance between such stations on the earth are short compared with the 20,000-km altitude of the GPS satellites' two receivers operating simultaneously, collecting signals from the same satellites and through substantially the same atmosphere will record very similar errors of several categories. However, multipath is so dependent on the geometry of the particular location of a receiver it cannot be lessened in this way. Keeping the antenna from reflective surfaces, the use of a mask angle, or the use of a ground plane or choke ring antenna are methods used to reduce multipath errors.

10. Answer is (d)

Explanation: The Master Control Station *MCS* of the Control Segment is located at Schriever (formerly Falcon) Air Force Base in Colorado Springs, Colorado. The station is manned by the 2nd Space Operations Squadron. Almanac and ephemeris information are uploaded by the Control Segment. The ephemeris is informed by orbital perturbations of the satellites. The locations of the other satellites in the constellation are included in the almanac. The Control Segment also provides each satellite's clock correction for rebroadcast. However, the P and C/A codes are actually generated in the satellites themselves.

3 The Framework

TECHNOLOGICAL FORERUNNERS

CONSOLIDATION

In the early 1970s the Department of Defense, DOD, commissioned a study to define its future positioning needs. That study found nearly 120 different types of positioning systems in place, all limited by their special and localized requirements. The study called for consolidation and NAVSTAR GPS (navigation system with timing and ranging, global positioning system) was proposed. Specifications for the new system were developed to build on the strengths and avoid the weaknesses of its forerunners. Here is a brief look at the earlier systems and their technological contributions toward the development of GPS.

TERRESTRIAL RADIO POSITIONING

Radar

Long before the satellite era the developers of RADAR (radio detecting and ranging) were working out many of the concepts and terms still used in electronic positioning today. For example, the classification of the radio portion of the electromagnetic spectrum by letters, such as the L-band now used in naming the GPS carriers, was introduced during World War II to maintain military secrecy about the new technology.

Actually, the 23-cm wavelength that was originally used for search radar was given the L designation because it was long compared to the shorter 10-cm wavelengths introduced later. The shorter wavelength was called S-band, the S for short. But the Germans used even shorter wavelengths of 1.5 cm. They were called K-band, for kurtz, meaning short in German. Wavelengths that were neither long nor short were given the letter C, for compromise, and P-band, for previous, was used to refer to the very first meter-length wavelengths used in radar. There is also an X-band radar used in fire-control radars and other applications.

In any case, the concept of measuring distance with electromagnetic signals (ranging in GPS) had one of its earliest practical applications in radar. Since then there have been several incarnations of the idea.

Distance by Timing

Shoran (short range navigation), a method of electronic ranging using pulsed VHF (very-high-frequency) signals, was originally designed for bomber navigation, but was later adapted to more benign uses. The system depended on a signal, sent by a mobile transmitter-receiver-indicator unit being returned to it by a fixed transponder. The elapsed time of the round trip was then converted into distances.

Shoran Surveying

It was not long before the method was adapted for use in surveying. Using Shoran from 1949 to 1957, Canadian geodesists were able to achieve precisions as high as 1:56,000 on lines of several hundred kilometers. Shoran was used in hydrographic surveys in 1945 by the Coast and Geodetic Survey. In 1951 Shoran was used to locate islands off Alaska in the Bering Sea that were beyond positioning by visual means. Also in the early 1950s, the U.S. Air Force created a Shoran measured trilateration net between Florida and Puerto Rico that was continued on to Trinidad and South America.

Shoran's success led to the development of Hiran (high-precision Shoran). Hiran's pulsed signal was more focused, its amplitude more precise, and its phase measurements more accurate.

Hiran Surveying

Hiran, also applied to geodesy, was used to make the first connection between Africa, Crete, and Rhodes in 1943. But its most spectacular application were the arcs of triangulation joining the North American Datum (1927) with the European Datum (1950) in the early 1950s. By knitting together continental datums, Hiran might be considered to be the first practical step toward positioning on a truly global scale.

Sputnik

These and other radio navigation systems proved that ranges derived from accurate timing of electromagnetic radiation were viable. But useful as they were in geodesy and air-navigation, they only whet the appetite for a higher platform. In 1957, the development of Sputnik, the first earth-orbiting satellite, made that possible.

Some of the benefits of earth-orbiting satellites were immediately apparent. It was clear that the potential coverage was virtually unlimited. But other advantages were less obvious.

Satellite Advantages

One advantage is that satellite technology allowed a more flexible choice of frequencies. The coverage of a terrestrial radio navigation system is limited by the propagation characteristics of electromagnetic radiation near the ground. To achieve long ranges, the basically spherical shape of the earth favors low frequencies that stay close to the surface. One system, *Loran-C* (long-range navigation-C), can be used to determine positions up to 3000 km from a fixed transmitter, but its frequency must be in the *LF* (low-frequency range) at 100 kHz or so.

Like several of the systems mentioned, Loran is still in use. In general it is used in applications where determination of position within a quarter of a nautical mile or so is sufficient. However, with modern technology, accuracies of 10 m are achievable today. *Omega*, another hyperbolic radio navigation system, can be used at ranges of

9000 km, but its 10- to 14-kHz frequency is so low it is actually audible (the range of human hearing is about 20 Hz to 15 kHz). These low frequencies have drawbacks. They can be profoundly affected by unpredictable ionospheric disturbances. And modeling the reduced propagation velocity of a radio signal over land can be difficult. But earth-orbiting satellites could allow the use of a broader range of frequencies and signals emanating from space are simply more reliable.

Using satellites, one could use high-frequency signals to achieve virtually limitless coverage. But development of the technology for launching transmitters with sophisticated frequency standards into orbit was not accomplished immediately. Therefore, some of the earliest extraterrestrial positioning was done with optical systems.

OPTICAL SYSTEMS

Optical tracking of satellites is a logical extension of astronomy. The astronomic determination of a position on the earth's surface from star observations, certainly the oldest method, is actually very similar to extrapolating the position of a satellite from a photograph of it crossing the night sky. In fact, the astronomical coordinates, right ascension α and declination δ, of such a satellite image are calculated from the background of fixed stars.

Triangulation with Photographs

Photographic images that combine reflective satellites and fixed stars are taken with *ballistic cameras* whose chopping shutters open and close very fast. The technique causes both the satellites, illuminated by sunlight or earth-based beacons, and the fixed stars to appear on the plate as a series of dots. Comparative analysis of photographs provides data to calculate the orbit of the satellite. Photographs of the same satellite made by cameras thousands of kilometers apart can thus be used to determine the camera's positions by triangulation. The accuracy of such networks has been estimated as high as ±5 meters.

Laser Ranging

Other optical systems are much more accurate. One called *SLR* (satellite laser ranging) is similar to measuring the distance to a satellite using a sophisticated EDM. A laser aimed from the earth to satellites equipped with retro reflectors yields the range. It is instructive that two current GPS satellites carry onboard corner cube reflectors for exactly this purpose. The GPS space vehicles numbered 36 (PRN 06) and 37 (PRN 07), launched in 1994 and 1993, respectively, can be tracked with SLR, thereby allowing ground stations to separate the effect of errors attributable to satellite clocks from errors in the satellite's ephemerides.

The same technique, called *LLR* (lunar laser ranging), is used to measure distances to the moon using corner cube reflectors left there during manned missions. These techniques can achieve positions of centimeter precision when information is gathered from several stations. However, one drawback is that the observations must be spread over long periods, up to a month, and they, of course, depend on two-way measurement.

Optical Drawbacks

While some optical methods, like SLR, can achieve extraordinary accuracies, they can at the same time be subject to some chronic difficulties. Some methods require skies to be clear simultaneously at widely spaced sites. Even then, local atmospheric turbulence causes the images of the satellites to scintillate. The bulky equipment is expensive and optical refraction is difficult to model. In the case of photographic techniques, emulsions cannot be made completely free from irregularities. Still optical tracking remains a significant part of the satellite management programs of the U.S. National Aeronautics and Space Administration (NASA) and other agencies.

EXTRATERRESTRIAL RADIO POSITIONING

Satellite Tracking

The earliest American extraterrestrial systems were designed to assist in satellite tracking and satellite orbit determination, not geodesy. Some of the methods used did not find their way into the GPS technology at all. Some early systems relied on the reflection of signals and transmissions from ground stations that would either bounce off the satellite or stimulate onboard transponders. But systems that required the user to broadcast a signal from the earth to the satellite were not favorably considered in designing the then-new GPS system. Any requirement that the user reveal his position was not attractive to the military planners responsible for developing GPS. They favored a passive system that allowed the user to simply receive the satellite's signal. So, with their two-way measurements and utilization of several frequencies to resolve the cycle ambiguity, many early extraterrestrial tracking systems were harbingers of the modern EDM technology more than GPS.

Prime Minitrack

But elsewhere there were ranging techniques useful to GPS. NASA's first satellite tracking system, *Prime Minitrack*, relied on phase difference measurements of a single-frequency carrier broadcast by the satellites themselves and received by two separate ground-based antennas. This technique is called *interferometry*. Interferometry is the measurement of the difference between the phases of signals that originate at a common source but travel different paths to the receivers. The combination of such signals, collected by two separate receivers, invariably finds them out of step since one has traveled a longer distance than the other. Analysis of the signal's phase difference can yield very accurate ranges, and interferometry has become an indispensable measurement technique in several scientific fields.

Very Long Baseline Interferometry

For example, very long baseline interferometry (VLBI) did not originate in the field of satellite tracking or aircraft navigation but in radio astronomy. The technique was so successful it is still in use today. Radio telescopes, sometimes on different

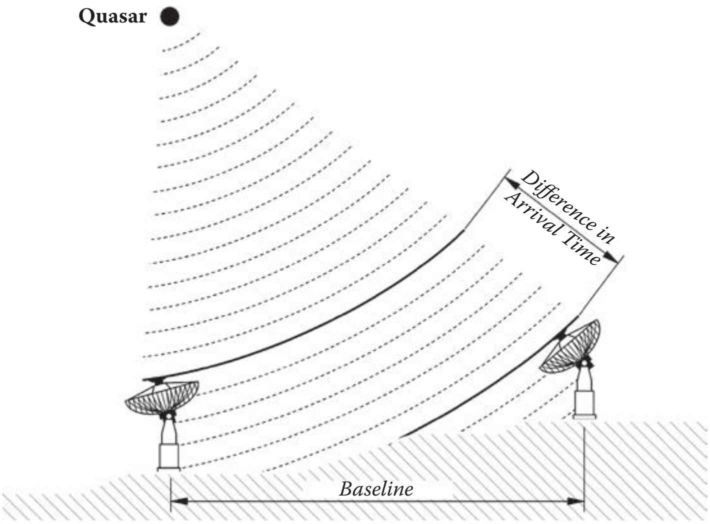


FIGURE 7.1 VLBI

continents, tape-record the microwave signals from quasars, star-like points of light billions of light-years from earth (Figure 3.1).

VERY LONG BASELINE INTERFEROMETRY

These recordings are encoded with time tags controlled by hydrogen masers, the most stable of oscillators (clocks). The tapes are then brought together and played back at a central processor. Cross-correlation of the time tags reveals the difference in the instants of wave front arrivals at the two telescopes. The discovery of the time offset that maximizes the correlation of the signals recorded by the two telescopes yields the distance and direction between them within a few centimeters, over thousands of kilometers.

VLBI's potential for geodetic measurement was realized as early as 1967. But the concept of high-accuracy baseline determination using phase differencing was really proven in the late 1970s. A direct line of development leads from the VLBI work of that era by a group from the Massachusetts Institute of Technology to today's most accurate GPS ranging technique, carrier phase measurement. VLBI, along with other extraterrestrial systems like SLR, also provide valuable information on the earth's gravitational field and rotational axis. Without that data, the high accuracy of the modern coordinate systems that are critical to the success of GPS, like the Conventional Terrestrial System (CT), would not be possible. But the foundation for routine satellite-based geodesy actually came even earlier and from a completely different direction. The first prototype satellite of the immediate precursor of the GPS system that was successfully launched reached orbit on June 29, 1961. Its range measurements were based on the Doppler effect, not phase differencing, and the system came to be known as *TRANSIT,* or more formally the *Navy Navigational Satellite System.*

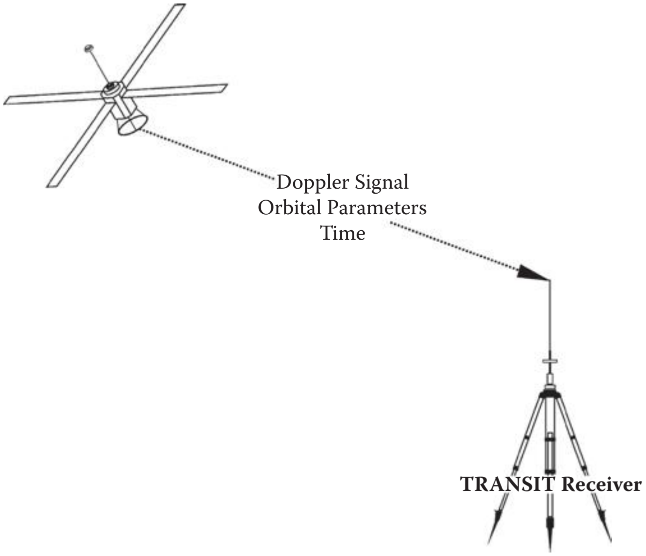


FIGURE 3.2 TRANSIT

TRANSIT

The Doppler Shift

Satellite technology and the Doppler effect were combined in the first comprehensive earth-orbiting satellite system dedicated to positioning. By tracking Sputnik in 1957, experimenters at Johns Hopkins University's Applied Physics Laboratory found that the Doppler shift of its signal provided enough information to determine the exact moment of its closest approach to the earth. This discovery led to the creation of the Navy Navigational Satellite System (NNSS) and the subsequent launch of 6 satellites specifically designed to be used for navigation of military aircraft and ships. This same system, eventually known as TRANSIT, was classified in 1964, declassified in 1967, and was widely used in civilian surveying for many years until it was switched off on December 31, 1996 (Figure 3.2).

TRANSIT Shortcomings

The TRANSIT system had some nagging drawbacks. For example, its primary observable was based on the comparison of the nominally constant frequency generated in the receiver with the Doppler-shifted signal received from 1 satellite at a time. With a constellation of only 6 satellites, this strategy sometimes left the observer waiting up to 90 minutes between satellites and at least two passes were usually required for acceptable accuracy. With an orbit of only about 1100 km above the earth the TRANSIT satellite's orbits were quite low and, therefore, unusually susceptible to atmospheric drag and gravitational variations, making the calculation of accurate orbital parameters particularly difficult.

TRANSIT and GPS

Through the decades of the 1970s and 1980s both the best and the worst aspects of the TRANSIT system were instructive. Some of the most successful strategies of the TRANSIT system have been incorporated into GPS. In both systems, the satellites broadcast their own ephemerides to the receivers. Both systems are divided into three segments: the control segment, including the tracking and upload facilities; the space segment, meaning the satellites themselves; and the user segment, everyone with receivers. In both systems the satellites broadcast two frequencies to allow compensation for the ionospheric dispersion. TRANSIT satellites used the frequencies of 400 MHz and 150 MHz, while GPS uses 1575.42 MHz and 1227.60 MHz. The Doppler effect is utilized in both systems. And in both the TRANSIT and GPS systems each satellite and receiver contains its own frequency standards.

Linking Datums

Perhaps the most significant difference between the TRANSIT system and previous extraterrestrial systems was TRANSIT's capability of linking national and international datums with relative ease. Its facility at strengthening geodetic coordinates laid the groundwork for modern geocentric datums.

NAVSTAR GPS

Orbits and Clocks

In 1973, the early GPS experiments were started. From the beginning of GPS, the plan was to include the best features and improve on the shortcomings of all of the previous work in the field. For example, the GPS satellites have been placed in nearly circular orbits over 20,000 km above the earth where the consequences of gravity and atmospheric drag are much less severe. And while the low frequency signals of the TRANSIT system made the ionospheric delay very troublesome, the much higher frequency of GPS signals reduce the effect. The rubidium, cesium, and hydrogen maser time clocks in GPS satellites are a marked improvement over the quartz oscillators used in TRANSIT satellites.

Increased Accuracy

TRANSIT's shortcomings restricted the practical accuracy of the system. Submeter work could only be achieved with long occupation on a station (at least a day) augmented by the use of a precise ephemerides for the satellites in postprocessing. GPS provides much more accurate positions in a much shorter time than any of its predecessors, but these improvements are only accomplished by standing on the shoulders of the technologies that have gone before.

Military Application

The genesis of GPS was military. It grew out of the congressional mandate issued to the Departments of Defense and Transportation to consolidate the myriad of

TABLE 3.1
Technologies Preceding GPS

Name of System	Range	Positional Accuracy in Meters	Features
OMEGA	Worldwide	2,200 CEP	Susceptible to VLF propagation anomalies
LORAN-C	U.S. and selected overseas	180 CEP	Skywave interference and localized coverage
TRANSIT	Worldwide	Submeter in days	Long waits between satellite passes
GPS	Worldwide	Centimeter in minutes	24 hour worldwide all weather

Note: CEP, Circular Error Probable.

navigation systems. Its application to civilian surveying was not part of the original design. In 1973 the DOD directed the Joint Program Office (JPO) in Los Angeles to establish the GPS system. Specifically, JPO was asked to create a system with high accuracy and continuous availability in real time that could provide virtually instantaneous positions to both stationary and moving receivers—all features that the TRANSIT system could not supply (see Table 3.1).

Secure, Passive, and Global

Worldwide coverage and positioning on a common coordinate grid were required of the new system—a combination that had been difficult, if not impossible, with terrestrial-based systems. It was to be a passive system, which ruled out any transmissions from the users, as had been tried with some previous satellite systems. But the signal was to be secure and resistant to jamming, so codes in the satellite's broadcasts would need to be complex.

Expense and Frequency Allocation

The U.S. Department of Defense (DOD) also wanted the new system to be free from the sort of ambiguity problems that had plagued OMEGA and other radar systems. And DOD did not want the new system to require large expensive equipment like the optical systems. Finally, frequency allocation was a consideration. The replacement of existing systems would take time, and with so many demands on the available radio spectrum, it was going to be hard to find frequencies for GPS.

Large Capacity Signal

Not only did the specifications for GPS evolve from the experience with earlier positioning systems, so did much of the knowledge needed to satisfy them. Providing 24-hour real-time, high-accuracy navigation for moving vehicles in three dimensions was a tall order. Experience showed part of the answer was a signal that was capable of carrying a very large amount of information efficiently and that required a large bandwidth. So, the GPS signal was given a double-sided 10-MHz bandwidth. But that was still not enough, so the idea of simultaneous observation of several satellites was also incorporated into the GPS system to accommodate the requirement. That decision had far-reaching implications.

The Satellite Constellation

Unlike some of its predecessors, GPS needed to have not one, but at least four satellites above an observer's horizon for adequate positioning, even more if possible. And the achievement of full-time worldwide GPS coverage would require this condition be satisfied at all times, anywhere on or near the earth.

Spread spectrum signal. The specification for the GPS system required all-weather performance and correction for ionospheric propagation delay. TRANSIT had shown what could be accomplished with a dual-frequency transmission from the satellites, but it had also proved that a higher frequency was needed. The GPS signal needed to be secure and resistant to both jamming and multipath. A spread spectrum, meaning spreading the frequency over a band wider than the minimum required for the information it carries, helped on all counts. This wider band also provided ample space for pseudorandom noise encoding, a fairly new development at the time. The PRN codes allowed the GPS receiver to acquire signals from several satellites simultaneously and still distinguish them from one another.

The Perfect System?

The absolute ideal navigational system, from the military's point of view, was described in the Army POS/NAV Master Plan in 1990. It should have worldwide and continuous coverage. The users should be passive. In other words, they should not be required to emit detectable electronic signals to use the system. The ideal system should be capable of being denied to an enemy and it should be able to support an unlimited number of users. It should be resistant to countermeasures and work in real time. It should be applicable to joint and combined operations. There should be no frequency allocation problems. It should be capable of working on common grids or map datums appropriate for all users. The positional accuracy should not be degraded by changes in altitude nor by the time of day or year. Operating personnel should be capable of maintaining the system. It should not be dependent on externally generated signals and it should not have decreasing accuracy over time or the distance traveled. Finally, it should not be dependent on the identification of precise locations to initiate or update the system.

A pretty tall order, GPS lives up to most, though not all, of the specifications.

GPS in Civilian Surveying

As mentioned earlier, application to civilian surveying was not part of the original concept of GPS. The civilian use of GPS grew up through partnerships between various public, private, and academic organizations. Nevertheless, while the military was still testing its first receivers, GPS was already in use by civilians. Geodetic surveys were actually underway with commercially available receivers early in the 1980s.

Federal Specifications

The Federal Radionavigation Plan of 1984, a biennial document including official policies regarding radionavigation, set the predictable and repeatable accuracy of civil and commercial use of GPS at 100 meters horizontally and 156 meters vertically. This specification meant that the C/A code ranging for Standard Positioning Service could be defined by a horizontal circle with a radius of 100 meters, 95% of the time. However, that same year civilian users were already achieving results up to six orders of magnitude better than that limit.

And since that time the accuracy available from SPS has increased from ±100 meters, horizontally, 95% of the time, to about ±20 meters to ±40 meters and better since Selective Availability was switched off on May 2, 2000 by presidential order.

Interferometry

By using interferometry, the technique that had worked so well with Prime Minitrack and VLBI, civilian users were showing that GPS surveying was capable of extraordinary results. In the summer of 1982 a research group at the Massachusetts Institute of Technology (MIT) tested an early GPS receiver and achieved accuracies of 1 and 2 ppm of the station separation. Over a period of several years extensive testing was conducted around the world that confirmed and improved on these results. In 1984, a GPS network was produced to control the construction of the Stanford Linear Accelerator. This GPS network provided accuracy at the millimeter level. In other words, by using the carrier phase observable instead of code ranging, private firms and researchers were going far beyond the accuracies the U.S. Government expected to be available to civilian users of GPS.

The interferometric solutions made possible by computerized processing developed with earlier extraterrestrial systems were applied to GPS by the first commercial users. The combination made the accuracy of GPS its most impressive characteristic, but it hardly solved every problem. For many years, the system was restricted by the shortage of satellites as the constellation slowly grew. The necessity of having four satellites above the horizon restricted the available observation sessions to a few, sometimes inconvenient, windows of time. Another drawback of GPS for the civilian user was the cost and the limited application of both the hardware and the software. GPS was too expensive and too inconvenient for everyday use, but the accuracy of GPS surveying was already extraordinary in the beginning.

Civil Applications of GPS

Today, with a mask angle of 10°, there are some periods when up to 10 satellites are above the horizon. And GPS receivers have grown from only a handful to the huge variety of receivers available today. Some push the envelope to achieve ever-higher accuracy; others offer less sophistication and lower cost. The civilian user's options are broader with GPS than any previous satellite positioning system—so broad that, as originally planned, GPS will likely replace more of its predecessors in both the military and civilian arenas, as it has replaced the TRANSIT system. In fact, GPS has developed into a system that has more civilian applications and users than military

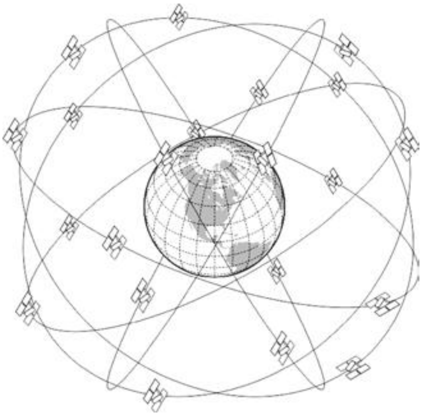


FIGURE 3.3 The GPS Constellation

ones. But the extraordinary range of GPS equipment and software requires the user to be familiar with an ever-expanding body of knowledge including the Global Navigation Satellite System (GNSS). GNSS includes both GPS and satellite navigation systems built by other nations and will present new options for users. However, the three segments of GPS will be presented before elaborating on GNSS.

GPS SEGMENT ORGANIZATION

The Space Segment

Though there has been some evolution in the arrangement, the current GPS constellation under full operational capability consists of 24 satellites However, there are more satellites than that in orbit and broadcasting at any given time. For example, the constellation includes several orbital spares.

As the primary satellites aged and their failure was possible, spares were launched. One reason for the arrangement was to maintain the 24 satellite constellation without interruption.

It was also done to ensure that it is possible to keep four satellites in each of the six orbital planes. Each of these planes is inclined to the equator by 55° (Figure 3.3) in a symmetrical, uniform arrangement. Such a uniform design does cover the globe completely and means that multiple satellite coverage is available even when satellites fail. Even though the satellites routinely outlast their anticipated design lives, they do eventually wear out. However, there is always concern about the balance between cost and satellite availability, and nonuniform arrangements can be used to optimize performance.

GPS CONSTELLATION

Orbital Period

NAVSTAR satellites are more than 20,000 km above the earth in a *posigrade* orbit. A posigrade orbit is one that moves in the same direction as the earth's rotation.

Since each satellite is nearly three times the earth's radius above the surface, its orbital period is 12 sidereal hours.

4 minute difference. When an observer actually performs a GPS survey project, one of the most noticeable aspects of a satellite's motion is that it returns to the same position in the sky about 4 minutes earlier each day. This apparent regression is attributable to the difference between 24 solar hours and 24 sidereal hours, otherwise known as star-time. GPS satellites retrace the same orbital path twice each sidereal day, but since their observers, on earth, measure time in solar units the orbits do not look quite so regular to them. The satellites actually lose 3 minutes and 56 seconds with each successive solar day.

This rather esoteric fact has practical applications; for example, if the satellites are in a particularly favorable configuration for measurement, the observer may wish to take advantage of the same arrangement the following day. However, he or she would be well advised to remember the same configuration will occur about 4 minutes earlier on the solar time scale. Both Universal Time (UT) and GPS time are measured in solar, not sidereal units. It is possible that the satellites will be pushed 50 km higher in the future to remove their current 4-minute regression, but for now it remains.

Design

As mentioned earlier, the GPS constellation was designed to satisfy several critical concerns. Among them were the best possible coverage of the earth with the fewest number of satellites, the reduction of the effects of gravitational and atmospheric drag, sufficient upload and monitoring capability with all control stations located on American soil, and the achievement of maximum accuracy.

Dilution of Precision

The distribution of the satellites above an observer's horizon has a direct bearing on the quality of the position derived from them. Like some of its forerunners, the accuracy of a GPS position is subject to a geometric phenomenon called *dilution of precision (DOP).* This number is somewhat similar to the strength of figure consideration in the design of a triangulation network. DOP concerns the geometric strength of the figure described by the positions of the satellites with respect to one another and the GPS receivers.

A low DOP factor is good, a high DOP factor is bad. In other words, when the satellites are in the optimal configuration for a reliable GPS position the DOP is low, when they are not, the DOP is high (Figure 3.4).

BAD DILUTION OF PRECISION

Four or more satellites must be above the observer's mask angle for the simultaneous solution of the clock offset and three dimensions of the receiver's position. But if all of those satellites are crowded together in one part of the sky, the position would be likely to have an unacceptable uncertainty and the DOP, or dilution of precision, would be high. In other words, a high DOP is a like a warning that the actual errors

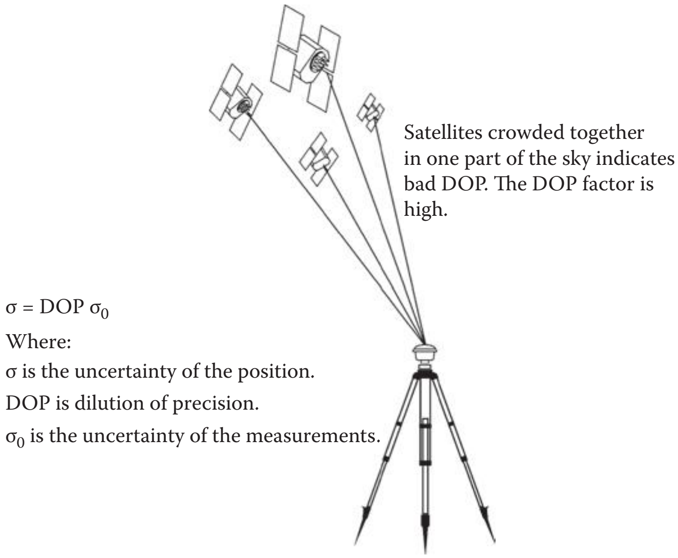


FIGURE 3.4 Bad DOP

in a GPS position are liable to be larger than you might expect. But remember, it is not the errors themselves that are directly increased by the DOP factor, it is the uncertainty of the GPS position that is increased by the DOP factor.

Here is an approximation of the effect:

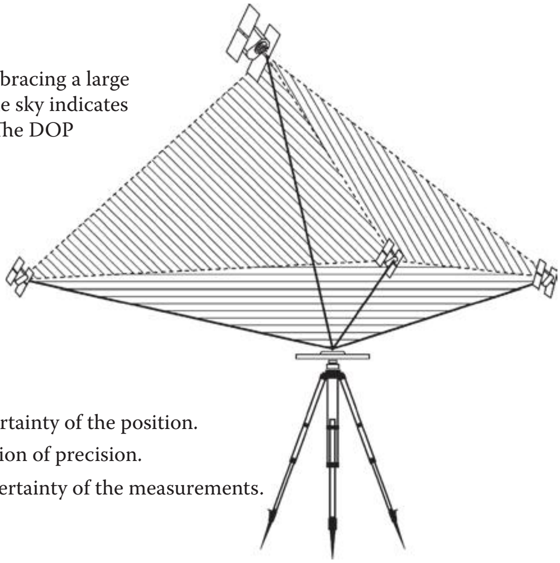
$$\sigma = \text{DOP}\ \sigma_o$$

where

σ = the uncertainty of the position
DOP = the dilution of precision factor
σ_o = the uncertainty of the measurements

In other words, the standard deviation of the GPS position is the dilution of precision factor multiplied by the standard deviations of the biases in the observables.

Now since a GPS position is derived from a three dimensional solution there are several DOP factors used to evaluate the uncertainties in the components of a GPS position. For example, there is horizontal dilution of precision (HDOP) and vertical dilution of precision (VDOP) where the uncertainty of a solution for positioning has been isolated into its horizontal and vertical components, respectively. When both horizontal and vertical components are combined, the uncertainty of three-dimensional positions is called position dilution of precision (PDOP). There is also time dilution of precision (TDOP), which indicates the uncertainty of the clock. There is geometric dilution of precision (GDOP), which is the combination of all of the above. And finally, there is relative dilution of precision (RDOP), which includes the number of receivers, the number of satellites they can handle, the length of the observing session, as well as the geometry of the satellites' configuration.



Satellites embracing a large
volume of the sky indicates
good DOP. The DOP
factor is low.

$\sigma = \text{DOP}\ \sigma_0$

Where:

σ is the uncertainty of the position.

DOP is dilution of precision.

σ_0 is the uncertainty of the measurements.

FIGURE 3.5 Good DOP

GOOD DILUTION OF PRECISION

The larger the volume of the body defined by the lines from the receiver to the satellites, the better the satellite geometry and the lower the DOP (Figure 3.5). An ideal arrangement of four satellites would be one directly above the receiver, the others 120° from one another in azimuth near the horizon. With that distribution the DOP would be nearly 1, the lowest possible value. In practice, the lowest DOPs are generally around 2. For example, if the standard deviation of a position were ±5 meters and the DOP 2, then the actual uncertainty of the position would be 2 times ± 5 meters or ±10 meters.

The users of most GPS receivers can set a PDOP mask to guarantee that data will not be logged if the PDOP goes above the set value. A typical PDOP mask is 6. As the PDOP increases the accuracy of the positions probably deteriorate, and as it decreases they probably improve, but in neither case is that outcome certain.

Outages

When a DOP factor exceeds a maximum limit in a particular location, indicating an unacceptable level of uncertainty exists over a period of time, that period is known as an *outage*. This expression of uncertainty is useful both in interpreting measured baselines and planning a GPS survey.

Satellite Positions in Mission Planning

The position of the satellites above an observer's horizon is a critical consideration in planning a GPS survey. So, most software packages provide various methods of illustrating the satellite configuration for a particular location over a specified period

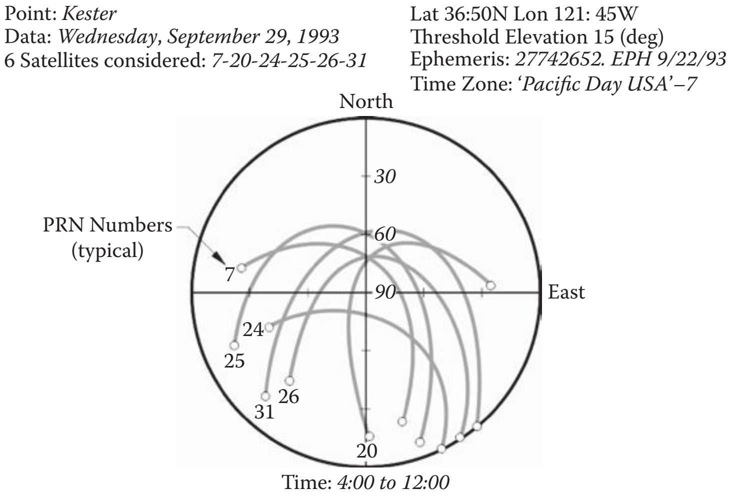


FIGURE 3.6 Polar Plot

of time. For example, the configuration of the satellites over the entire span of the observation is important; as the satellites move, the DOP changes. Fortunately the dilution of precision can be worked out in advance. DOP can be predicted. It depends on the orientation of the GPS satellites relative to the GPS receivers. And since most GPS software allow calculation of the satellite constellation from any given position and time, they can also provide the accompanying DOP factors.

Another commonly used plot of the satellite's tracks is constructed on a graphical representation of the half of the celestial sphere. The observer's zenith is shown in the center and the horizon on the perimeter. The program usually draws arcs by connecting the points of the instantaneous azimuths and elevations of the satellites above a specified mask angle. These arcs then represent the paths of the available satellites over the period of time and the place specified by the user.

In Figure 3.6, the plot of the polar coordinates of the available satellites with respect to time and position is just one of several tables and graphs available to help the GPS user visualize the constellation.

Figure 3.7 another useful graph that is available from many software packages. It shows the correlation between the number of satellites above a specified mask angle and the associated PDOP for a particular location during a particular span of time.

There are four spikes of unacceptable PDOP, labeled here for convenience. It might appear at first glance that these spikes are directly attributable to the drop in the number of available satellites. However, please note that while spike 1 and 4 do indeed occur during periods of 4 satellite data, spikes 2 and 3 are during periods when there are 7 and 5 satellites available, respectively. It is not the number of satellites above the horizon that determine the quality of GPS positions, one must also look at their position relative to the observer, the DOP, among other things. The variety of the tools to help the observer predict satellite visibility underlines the importance of their configuration to successful positioning.

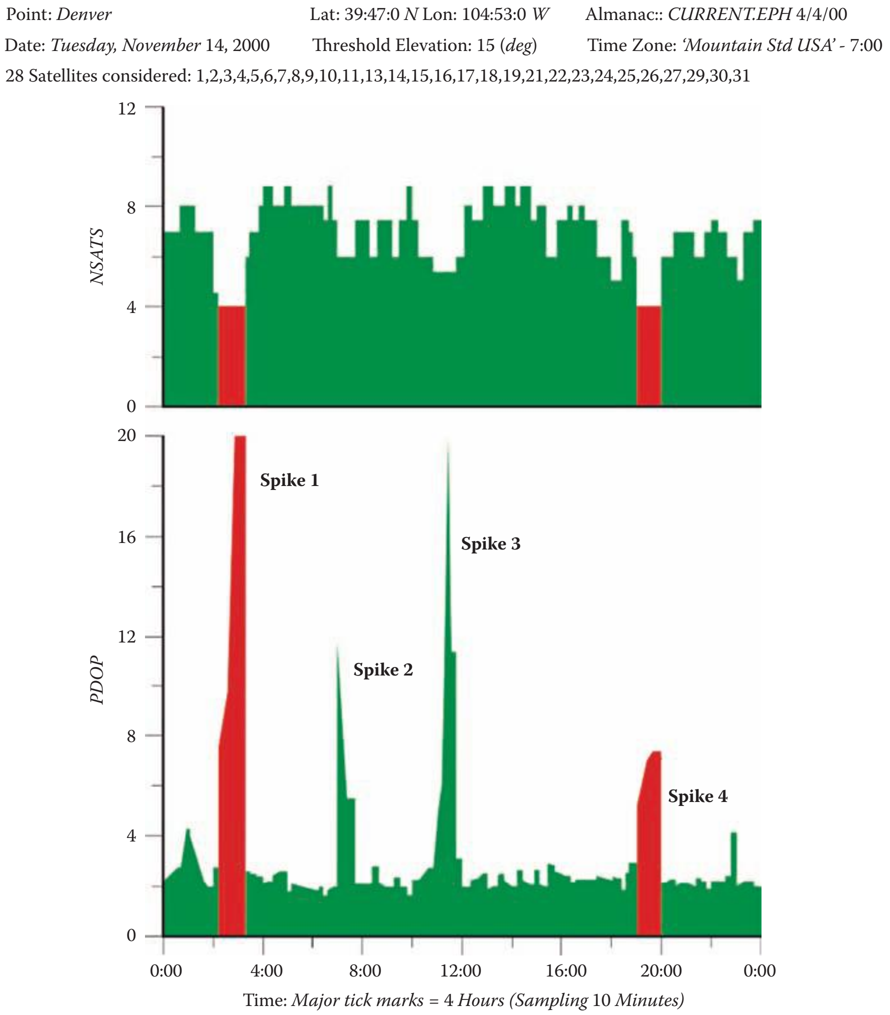


FIGURE 3.7 Number of Space Vehicles (SVs) and PDOP

Satellite Names

The 11 GPS satellites launched from Vandenberg Air Force Base between 1978 and 1985 were known as *Block I satellites*. Ten of the satellites built by Rockwell International achieved orbit on Atlas F rockets. There was one launch failure. All were prototype satellites built to validate the concept of GPS positioning. This test constellation of Block I satellites was inclined by 63° to the equator instead of the current specification of 55°. They could be maneuvered by hydrazine thrusters operated by the control stations.

The first GPS satellite was launched February 22, 1978 and was known as *Navstar 1*. An unfortunate complication is that this satellite was also known as PRN 4 just as Navstar 2 was known as PRN 7. The Navstar number, or Mission number, includes the Block name and the order of launch, for example I-1, meaning the first satellite of Block I, and the PRN number refers to the weekly segment of the P code that has been assigned to the satellite, and there are still more identifiers. Each GPS satellite has a Space Vehicle number, an Interrange Operation Number, a NASA catalog number, and an orbital position number as well. However, in most literature, and to the GPS receivers themselves, the PRN number is the most important.

Block I

The Block I satellites weigh 845 kg in final orbit. They were powered by three rechargeable nickel-cadmium batteries and 7.25 square meters of single-degree solar panels. These experimental satellites served to point the way for some of the improvements found in subsequent generations. For example, even with the backup systems of two rubidium and two cesium oscillators onboard each satellite, the clocks proved to be the weakest components. The satellites themselves could only store sufficient information for 3½ days of independent operation. And the uploads from the control segment were not secure; they were not encrypted. Still, all 11 achieved orbit, except Navstar 7. The design life for these satellites was 4½ years, but their actual average lifetime was 8.76 years. There are no Block I satellites operating today. The last Block I satellite was retired in late 1995.

Block II and Block IIA

The next generation of GPS satellites are known as *Block II satellites*. There have been 28 built. The first left Cape Canaveral on February 14, 1989, almost 14 years after the first GPS satellite was launched. It was about twice as heavy as the first Block I satellite and is expected to have a design life of 7½ years *with a Mean Mission Duration, MMD,* of 6 years. The Block II satellites can operate up to 14 days without an upload from the control segment and their uploads are encrypted. The satellites themselves are radiation hardened, and their signals were subject to selective availability.

These satellites were also built by Rockwell International. Block II included the launch of 9 satellites between 1989 and 1990. Nineteen Block IIA satellites, with several navigational improvements, were launched between 1990 and 1997.

Block IIR

The third generation of GPS satellites is known as *Block IIR satellites*; the R stands for *replenishment*. There are two significant advancements in these satellites. The Block IIR satellites have enhanced autonomous navigation capability because of their use of intersatellite linkage. These GPS satellites are not only capable of self-navigation, they also provide other spacecraft equipped with an onboard GPS receiver with the data they need to define their own positions.

Block IIF

The fourth generation of GPS satellites is known as *Block IIF satellites*; the F stands for *follow-on*. The program includes the procurement of 33 satellites and the

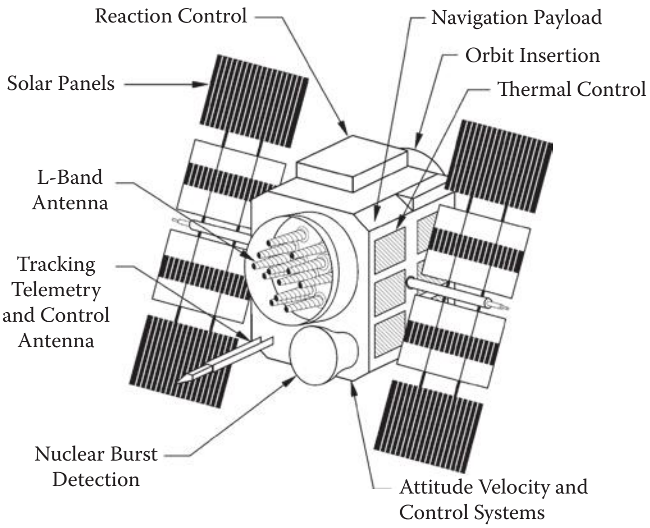


FIGURE 3.8 GPS Block II Satellite

operation and support of a new GPS operations control segment. They will include a third civilian frequency on the Block IIF satellites. The frequency is called L5. The Block IIF satellites are expected to have a design life of 12 years, and a *Mean Mission Duration, MMD* life of 9.9 years.

BLOCK II GPS SATELLITE

Signal Deterioration

While the signals from Block I satellites were not subject to any officially sanctioned deterioration, the same cannot be said of the Block II satellites (Figure 3.8). In the interest of national security the signals from the operational constellation of GPS satellites, including the Block II/IIA and IIR satellites, were intentionally degraded periodically. Selective availability (SA) of the C/A code was implemented by disrupting the satellite clock frequency from time to time from April 1990 until May 2, 2000 at approximately 4:00 UT when it was discontinued.

While SA has been switched off, the P code does continue to be intermittently supplanted by the encrypted Y code in a procedure known as *antispoofing (AS)*. However, this procedure does not significantly affect relative positioning methods that rely on the carrier beat phase observable.

GPS Satellites

All GPS satellites have some common characteristics. They weigh about a ton and with solar panels extended are about 27 feet long. They generate about 700 watts of power. They all have three-dimensional stabilization to ensure that their solar arrays point toward the sun and their 12 helical antennae to the earth. GPS satellites move at a speed of about 8700 miles per hour. Even so, the satellites must pass through the shadow of the earth from time to time and onboard batteries provide power. All satellites are equipped with thermostatically controlled heaters and reflective insulation to maintain the optimum temperature for the oscillator's operation. Prior to launch, GPS satellites are checked out at a facility at Cape Canaveral, Florida.

THE CONTROL SEGMENT

As mentioned in Chapter 1, there are several government tracking and uploading facilities distributed around the world. These facilities not only monitor the L-band signals from the GPS satellites and update their Navigation Messages but also track the satellite's health, their maneuvers, and even battery recharging. Taken together these facilities are known as the *Control Segment*. Here is how they work together.

Master Control Station

The Master Control Station (MCS), once located at Vandenberg Air Force Base in California, now resides at the Consolidated Space Operations Center (CSOC) at Schriever Air Force Base near Colorado Springs, Colorado, and has been manned by the 2nd Space Operations Squadron, *2SOPS,* since 1992. The squadron not only operates and maintains the Master Control Station, but also the network of ground antennas and monitoring stations. There is also a backup MCS at Gaithersburg, Maryland.

The monitoring stations track the navigation signals of all the satellites. They provide meteorological, pseudorange, and carrier phase data to the MCS where they are smoothed using a *Kalman* filter. The carrier phase measurements are used to smooth the pseudorange data to reduce the measurement noise, more about that in a moment. That information is then processed at the MCS and used to update the satellites' navigation messages. The MCS sends updated information, including the broadcast ephemeris and satellite clock, to the satellites through ground antennas at the uploading stations.

Other Stations

The Air Force monitoring stations are located at Ascension Island, Colorado Springs, Diego Garcia, Hawaii, Kwajalein Atoll, and Cape Canaveral. In 2005 during the Legacy Accuracy Improvement Initiative, L-AII, an additional six stations administered by the NGA have been added at Washington, D.C., England, Argentina, Ecuador, Bahrain, and Australia. Together they monitor all GPS satellites in view and collect ranging and satellite clock data from all available satellites. Their data is then passed on to MCS and operators in the MCS calculate each satellite's status, ephemeris, and clock data. This information is sent to ground antennas located at Ascension Island, Diego Garcia, Kwajalein Atoll, and Cape Canaveral, there is also a ground antenna in Colorado. There the data is uploaded back to each satellite for inclusion in their broadcast messages. The 12-station network ensures that each satellite in the GPS constellation is monitored almost continuously from at least two stations. In the future 5 more NGA stations are set to be added to the network and every satellite will be monitored from three stations. And a further improvement in the Control Segment, the Architecture Evolution Plan, AEP, to accommodate the Block IIF satellites and accompanying new signals is coming.

Kalman Filtering

Kalman filtering, named for R.E. Kalman's recursive solution for least-squares filtering (Kalman 1960), has been applied to the results of radionavigation for several decades. It is a statistical method of smoothing and condensing large amounts of data. One of its uses in GPS is reduction of the pseudoranges measured at 1.5-second intervals between a monitoring station and a satellite. Kalman filtering is used to condense a smoothed set of pseudoranges for a 15-minute period that is transmitted to the Master Control Station.

An Analogy

Kalman filtering can be illustrated by the example of an automobile speedometer. Imagine the needle of an automobile's speedometer has a bent cable and is fluctuating between 64 and 72 mph as the car moves down the road. The driver might estimate the actual speed at 68 mph. Although not accepting the speedometer's measurements literally, he has taken them into consideration and constructed an internal model of his velocity. If the driver further depresses the accelerator and the needle responds by moving up, his reliance on the speedometer increases. Despite its vacillation, the needle has reacted as the driver thought it should. It went higher as the car accelerated. This behavior illustrates a predictable correlation between one variable, acceleration, and another, speed. Now he is more confident in his ability to predict the behavior of the speedometer. The driver is illustrating *adaptive gain*, meaning that he is fine tuning his model as he receives new information about the measurements. As he does, a truer picture of the relationship between the readings from the speedometer and his actual speed emerges, without recording every single number as the needle jumps around. The driver in this analogy is like the Kalman filter.

Without this ability to take the huge amounts of satellite data and condense it into a manageable number of components, GPS processors would be overwhelmed. Kalman filtering is used in the uploading process to reduce the data to the satellite clock offset and drift, 6 orbital parameters, 3 solar radiation pressure parameters, biases of the monitoring stations clock, and a model of the tropospheric effect and earth rotational components.

Constant Tracking

Every GPS satellite is being tracked by at least one of the control segment's monitoring stations at all times. The MCS sends its updates around the world to strategically located uploading stations. They in turn transmit the new Navigation messages to each satellite. The Block II satellites can function without new uploads for 14 days and Block IIA satellites for only 180 days. In any event, the older their Navigation message gets, the more its veracity deteriorates.

Postcomputed Ephemerides

This system is augmented by other tracking networks that produce post-computed ephemerides. Their impetus has been several: the necessity of timely orbital information with more precision than the broadcast ephemeris and the correlation of the terrestrial coordinate systems with the orbital system through VLBI and SLR sites, among others. The International GPS Geodynamics Service (IGS) is an example of a global tracking network. There are also regional organizations such as the network of eight tracking stations that form the Australian Fiducial Network (AFN) administered by the Australian Surveying and Land Information Group (AUSLIG).

In the United States, the National Oceanic and Atmospheric Administration (NOAA) and, more specifically, the National Geodetic Survey (NGS) have been given the job of providing accurate GPS satellite ephemerides, or orbits. Their precise ephemerides are derived using 24 hour data segments from the global GPS network coordinated by the IGS. It is important to note that the former Cooperative International GPS Network or CIGNET is now a part of the IGS.

The IGS consists of 41 continuously operating GPS receivers with more being added to improve the global distribution. Wherever possible, the IGS network receivers are collocated with the very long baseline interferometry (VLBI) radio telescopes. The IGS reference frame is the International Earth Rotation Service Terrestrial Reference Frame (ITRF).

The precise ephemerides from NGS are available on-line along with a summary file to document the computation and to convey relevant information about the observed satellites, such as maneuvers or maintenance. The precise ephemerides are usually available 2 to 6 days after the date of observation. Each data set provides one week's ephemeris information at 15-minute intervals. The online address is http://www.ngs.noaa.gov/orbits.

GPS for Orbit Calculation

It is important to note that GPS receivers are installed on other earth orbiting satellites and are thereby becoming a utility for orbit calculation for those space systems. Satellites from those in low Earth and extending out to geosynchronous orbits are utilizing GPS technology on board more and more.

THE USER SEGMENT

The military plans to build a GPS receiver into virtually all of its ships, aircraft, and terrestrial vehicles. In fact, the Block IIR satellites may be harbingers of the incorporation of more and more receivers into extraterrestrial vehicles as well. But even with such widespread use in the military, civilian GPS will be still more extensive.

Constantly Increasing Application of GPS

The uses the general public finds for GPS will undoubtedly continue to grow as the cost and size of the receivers continues to shrink. The number of users in surveying will be small when compared with the large numbers of trains, cars, boats, and airplanes with GPS receivers. GPS will be used to position all categories of civilian transportation, as well as law enforcement and emergency vehicles. Nevertheless, surveying and geodesy have the distinction of being the first practical application of GPS and the most sophisticated uses and users are still under its purview. That situation will likely continue for some time.

Next chapter. The number, range, and complexity of GPS receivers available to surveyors has exploded in recent years. There are widely varying prices and features that sometimes make it difficult to match the equipment with the application. Chapter 4 will be devoted to a detailed discussion of the GPS receivers themselves.

EXERCISES

1. Which GPS satellites carry corner cube reflectors, and what is their purpose?

    (a) SVN 32 and SVN 33 carry onboard corner cubes to allow photographic tracking. The purpose of the reflectors is to allow ground stations to distinguish the satellites, illuminated by earth-based beacons, from the background of fixed stars.
    (b) All GPS satellites carry corner cube reflectors. The purpose of the reflectors is to allow the users to broadcast signals to the satellites that will activate the onboard transponders.
    (c) SVN 36 and SVN 37 carry onboard corner cubes to allow SLR. The purpose of the reflectors is to allow ground stations to distinguish between satellite clock errors and satellite ephemeris errors.
    (d) No GPS satellites carry corner cube reflectors. Such an arrangement would require the user to broadcast a signal from the earth to the satellite. Any requirement that the user reveal his position is not allowed by the military planners responsible for developing GPS. They have always favored a passive system that allowed the user to simply receive the satellite's signal.

2. Which of the following statements concerning the L-band designation is not true?

    (a) The frequency bands used in radar were given letters to preserve military secrecy.
    (b) The GPS carriers, L1 and L2, are named for the L-band radar designation.
    (c) The frequencies broadcast by the TRANSIT satellites were within the L-band.
    (d) The L in L-band stands for long.

3. Which of the following is an aspect of the NAVSTAR GPS system that is an improvement on the retired TRANSIT system?

    (a) Rubidium, cesium, and hydrogen maser frequency standards.
    (b) Satellite that broadcast two frequencies.
    (c) A passive system, one that does not require transmissions from the users.
    (d) Satellites that broadcast their own ephemerides.

4. Which of the following requirements for the ideal navigational system, from the military point of view, described in the Army POS/NAV Master Plan in 1990 does GPS not currently satisfy?

   (a) The users should be passive.
   (b) It should be resistant to countermeasures.
   (c) It should be capable of working in real time.
   (d) It should not be dependent on externally generated signals.

5. Which of the following statements best explains the fact that for a stationary receiver a GPS satellite appears to return to the same position in the sky about 4 minutes earlier each day?

   (a) Over the same period of time VDOP, vertical dilution of precision, is frequently larger than HDOP, horizontal dilution of precision, for a stationary receiver.
   (b) The apparent regression is due to the difference between the star-time and solar time over a 24-hour period.
   (c) The loss of time is attributable to the satellite's pass through the shadow of the earth.
   (d) The apparent regression is due to the cesium and rubidium clocks of earlier GPS satellites. They will be replaced in the Block IIR satellites by hydrogen masers. Hydrogen masers are much more stable than earlier oscillators.

6. All of the following concepts were developed in other contexts and all are now utilized in GPS. Which of them has been around the longest?

   (a) Orbiting transmitters with accurate frequency standards.
   (b) Kalman filtering.
   (c) Measuring distances with electromagnetic signals.
   (d) The Doppler shift.

7. Practically speaking, which of the following was the most attractive aspect of first civilian GPS surveying in the early 1980s?

   (a) Satellite availability.
   (b) GPS hardware.
   (c) Accuracy.
   (d) GPS software.

8. Which of the following satellite identifiers is most widely used?

   (a) Interrange Operation Number
   (b) NASA catalog number
   (c) PRN number
   (d) NAVSTAR number

9. Satellites in which of the following categories are currently providing signals for positioning and navigation?

    (a) Block IIR
    (b) TRANSIT
    (c) Block IIF
    (d) Block I

10. Which of the following is not the location of a GPS Control Segment monitoring station?

    (a) Ascension Island
    (b) Diego Garcia
    (c) Cape Canaveral
    (d) Kwajalein Atoll

ANSWERS AND EXPLANATIONS

1. Answer is (c)

Explanation: Two current GPS satellites carry onboard corner cube reflectors, SVN 36 (PRN 06) and SVN 37 (PRN 07), launched in 1994 and 1993, respectively. The purpose of the corner cube reflectors is to allow SLR tracking. The ground stations can use this exact range information to separate the effect of errors attributable to satellite clocks from errors in the satellite's ephemerides. Remember the satellite's ephemeris can be likened to a constantly updated coordinate of the satellite's position. The Satellite Laser Ranging (SLR) can nail down the difference between the satellite's broadcast ephemeris and the satellite's actual position.

This allows, among other things, more proper attribution of the appropriate portion of the range error to the satellite's clock.

2. Answer is (c)

Explanation: The original letter designations were assigned to frequency bands in radar to maintain military secrecy. The L-band was given the letter L to indicate that its wavelength was long. The frequencies within the L-band are from 3900 MHz to 1550 MHz, approximately.

Stated another way, the wavelengths within the L-band are approximately from 76 cm to 19 cm.

The GPS carrier frequencies L1 at 1575.42 MHz and L2 at 1227.60, with wavelengths of 19 cm and 24 cm respectively, and are both close to the L-band range. While one might say that the L1 frequency is not exactly within original L-band designation, the frequencies broadcast by the TRANSIT satellites certainly are not. The frequencies used in the TRANSIT system are 400 MHz and 150 MHz. These frequencies fall much below the L-band and into the VHF range.

It is interesting to note that the old L-band designation has actually been replaced; it is now known as the D-band. However, there does not appear to be any intention to change the names of the GPS carrier frequencies.

3. Answer is (a)

Explanation: Many of the innovations used in the TRANSIT system informed decisions in creating the NAVSTAR GPS system. Both have satellites that broadcast two frequencies to allow compensation for the ionospheric dispersion. TRANSIT satellites used the frequencies of 400 MHz and 150 MHz, while GPS uses 1575.42 MHz and 1227.60 MHz. And while the low frequencies used in the TRANSIT system were not as effective at eliminating the ionospheric delay, the idea is the same. Both systems are passive, meaning there is no transmission from the user required. Both systems use satellites that broadcast their own ephemerides to the receivers.

And there are more similarities; both systems are divided into three segments: the control segment, including the tracking and upload facilities; the space segment, meaning the satellite themselves; and the user segment, everyone with receivers. And in both the TRANSIT and GPS systems each satellite and receiver contains its own frequency standards. However, the standards used in NAVSTAR GPS satellites are much more sophisticated than those that were used in TRANSIT. The rubidium, cesium, and hydrogen maser frequency standards used in GPS satellites far surpass the quartz oscillators that were used in the TRANSIT satellites. The TRANSIT navigational broadcasts were switched off on December 31, 1996.

4. Answer is (d)

Explanation: Not only does GPS have worldwide 24-hour coverage, it is also capable of providing positions on a huge variety of grids and datums. The system allows the user's receivers to be passive; there is no necessity for them to emit any electronic signal to use the system. GPS codes are complex and there is more than one strategy that the Department of Defense can use to deny GPS to an enemy. At the same time, there can be a virtually an unlimited number of users without over-taxing the system. However, GPS is dependent on externally generated signals for the continued health of the system. While satellites can operate for periods without uploads and orbital adjustments from the Control Segment, they certainly cannot do without them entirely.

The Control Segment's network of ground antennas and monitoring stations track the navigation signals of all the satellites. The MCS uses that information to generate updates for the satellites which it uploads through ground antennas.

5. Answer is (b)

Explanation: The difference between 24 solar hours and 24 sidereal hours, otherwise known as star-time, is 3 minutes and 56 seconds, or about 4 minutes. GPS satellites retrace the same orbital path twice each sidereal day, but since their observers, on earth, measure time in solar units the orbits do not look quite so regular to them and both Universal Time (UT) and GPS time are measured in solar, not sidereal units.

6. Answer is (d)

Explanation: The concept of measuring distance with electromagnetic signals had its earliest practical applications in radar in the 1940s and during World War II.

Development of the technology for launching transmitters with onboard frequency standards into orbit was available soon after the launch of Sputnik in 1957. TRANSIT 1B launched on April 13, 1960 was the first successfully launched navigation satellite.

Kalman filtering, named for R.E. Kalman's recursive solution for least-squares filtering, was developed in 1960. It is a statistical method of smoothing and condensing large amounts of data and has been used in radionavigation ever since. It is an integral part of GPS.

However, the Doppler shift was discovered in 1842, and certainly has the longest history of any of the ideas listed. The Doppler shift describes the apparent change in frequency when an observer and a source are in relative motion with respect to one another. If they are moving together the frequency of the signal from the source appears to rise, and if they are moving apart the frequency appears to fall. The Doppler effect came to satellite technology during the tracking of Sputnik in 1957. It occurred to observers at Johns Hopkins University's Applied Physics Laboratory that the Doppler shift of its signal could be used to find the exact moment of its closest approach to the earth. However, the phenomenon had been described 115 years earlier by Christian Doppler using the analogy of a ship on the ocean.

7. Answer is (d)

Explanation: The interferometric solutions made possible by computerized processing developed with earlier extraterrestrial systems were applied to GPS by the first commercial users, but the software was cumbersome by today's standards. There were few satellites up in the beginning and the necessity of having at least four satellites above the horizon restricted the available observation sessions to difficult periods of time. GPS receivers were large, unwieldy, and very expensive. Nevertheless, in the summer of 1982 a research group at the Massachusetts Institute of Technology (MIT) tested an early GPS receiver and achieved accuracies of 1 and 2 ppm of the station separation. In 1984, a GPS network was produced to control the construction of the Stanford Linear Accelerator. This GPS network provided accuracy at the millimeter level. GPS was inconvenient and expensive, but the accuracy was remarkable from the outset.

8. Answer is (c)

Explanation: The satellite that has currently been given the number Space Vehicle 32, or SV32 is also known as PRN 1. This particular satellite which was launched on November 22, 1992 currently occupies orbital slot F-1. It has a Navstar number, or Mission number of IIA-16. This designation includes the Block name and the order of launch of that mission.

This same GPS satellite also has an Interrange Operation Number and a NASA catalog number. However, the most often used identifier for this satellite is PRN 1.

9. Answer is (a)

Explanation: The first of 6 TRANSIT satellites to reach orbit was launched on June 29, 1961. The constellation of satellites known as the Navy Navigational Satellite System functioned until it was switched off on December 31, 1996 replaced by the GPS system. The first of 11 Block I GPS satellites was launched on February 22, 1978 and the last on October 9, 1985, however, no Block I satellites are operating today.

However, several Block IIR GPS satellites are currently operational. The Block IIR are the third generation of GPS satellites. These satellites use hydrogen masers frequency standards which are much more stable than earlier oscillators. The Block IIR satellites also have enhanced autonomous navigation capability. The first satellite of this category was launched on January 17, 1997.

10. Answer is (c)

Explanation: The monitoring stations are located at Ascension Island, Colorado Springs, Diego Garcia, Hawaii, and Kwajalein Atoll. A back-up station planned for Onizuka Air Force Base, California. However, a Control Segment monitoring station is not located at Cape Canaveral.

The monitoring stations track all GPS satellites in view and collect ranging and satellite clock data using the P code pseudoranges and integrated Doppler measurements. Their data is then passed on to MCS where each satellite's status, ephemeris and clock error is calculated. This information is sent to antennas located at the monitoring stations (except Hawaii) where the data is uploaded back to each satellite for inclusion in their broadcast messages.

4 Receivers and Methods

COMMON FEATURES OF GPS RECEIVERS

A BLOCK DIAGRAM OF A CODE CORRELATION RECEIVER

Receivers for GPS Surveying

The most important hardware in a GPS surveying operation are the receivers (Figure 4.1). Their characteristics and capabilities influence the techniques available to the user throughout the work, from the initial planning to processing. There are literally hundreds of different GPS receivers on the market. Only a portion of that number is appropriate for GPS surveying and they share some fundamental elements.

They are generally capable of accuracies from submeter to subcentimeter. They are capable of differential GPS (DPGS), real-time GPS, static GPS, and other hybrid techniques. They usually are accompanied by postprocessing software and network adjustment software. And many are equipped with capacity for extra batteries, external data collectors, external antennas, and tripod mounting hardware. These features, and others, distinguish GPS receivers used in the various aspects of surveying from handheld GPS units designed primarily for recreational use.

A GPS receiver, like any GPS receiver, must collect and then convert signals from GPS satellites into measurements. It is not easy. The GPS signal has low power to start with. An orbiting GPS satellite broadcasts this weak signal across a cone of approximately 28° of arc. From the satellite's point of view, about 11,000 miles up, that cone covers the whole planet. It is instructive to contrast this arrangement with a typical communication satellite that not only has much more power, but a very directional signal as well. Its signals are usually collected by a large dish antenna, but the typical GPS receiver has a very small, relatively nondirectional antenna. Fortunately, antennas used for GPS receivers do not even have to be pointed directly at the signal source.

Stated another way, a GPS satellite spreads a low power signal over a large area rather than directing a high power signal at a very specific area. In fact, the GPS signal would be completely obscured by the huge variety of electromagnetic noise that surrounds us if it were not a spread spectrum coded signal. The GPS signal intentionally occupies a broader frequency bandwidth than it must to carry its information. This characteristic is used to prevent jamming, mitigate multipath, and allow unambiguous satellite tracking.

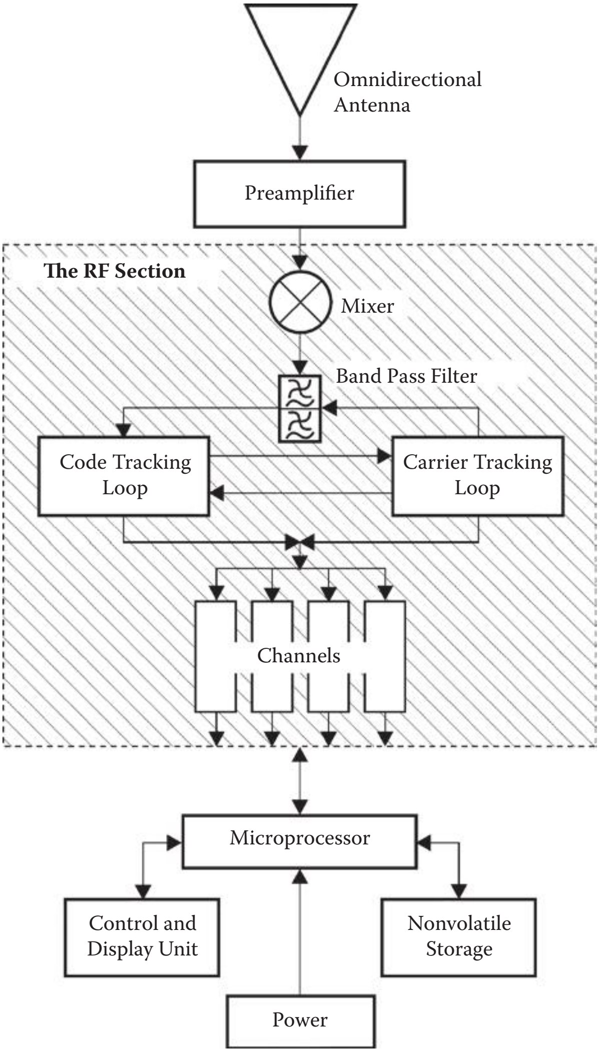


FIGURE 4.1 A Block Diagram of a GPS Receiver

THE ANTENNA

The antenna collects the satellite's signals. Its main function is the conversion of electromagnetic waves into electric currents sensible to the *radio frequency* (*RF*) section of the receiver, more about the RF section later. Several antenna designs are possible in GPS, but the satellite's signal has such a low power density, especially after propagating through the atmosphere, that antenna efficiency is critical. Therefore GPS antennas must have high sensitivity, also known as *high gain*. They can be designed to collect only the L1 frequency, or both L1 and L2, but in all cases they must be *Right Hand Circular Polarized,* (*RHCP*) as are the GPS signals broadcast from the satellites.

Most receivers have an antenna built in, but many can accommodate a separate tripod-mounted or range pole-mounted antenna as well. These separate antennas with their connecting coaxial cables in standard lengths are usually available from the receiver manufacturer. The cables are an important detail. The longer the cable, the more of the GPS signal is lost traveling through it.

As mentioned earlier the wavelengths of the GPS carriers are 19 cm (L1) and 24 cm (L2) and antennas that are a quarter or half wavelength tend to be the most practical and efficient so GPS antenna elements can be as small as 4 or 5 cm. Most of the receiver manufacturers use a *microstrip* antenna. These are also known as patch antennas. The microstrip can receive one or both GPS frequencies although dual frequency microstrip antennas often have two patches, one for each frequency. Microstrip antennas are durable, compact, have a simple construction and a low profile. The next most commonly used antenna is known as a *dipole*. A dipole antenna has a stable phase center and simple construction, but needs a good ground plane. A ground plane also facilitates the use of a microstrip antenna where it not only ameliorates multipath, but also tends to increase the antenna's zenith gain. A *quadrifilar* antenna is a single frequency antenna that has two orthogonal bifilar helical loops on a common axis. Quadrifilar antennas perform better than a microstrip on crafts that pitch and roll like boats and airplanes. They are also used in many recreational handheld GPS receivers. Such antennas have a good gain pattern, do not require a ground plane, but are not azimuthally symmetric. The least common design is the *helix* antenna. A helix is a dual frequency antenna. It has a good gain pattern, but a high profile.

Bandwidth

An antenna ought to have a bandwidth commensurate with its application. In general, the larger the bandwidth, the better the performance. However, there is a downside; increased bandwidth degrades the signal to noise ratio by including more interference. GPS microstrip antennas usually operate in a range from about 2 to 20 MHz, which corresponds with the null-to-null bandwidth of both new and legacy GPS signals. For example L2C and the central lobe of the C/A code span 2.046 MHz, whereas L5 and the P(Y) code have a bandwidth of 20.46 MHz. Therefore the antenna and front-end of a receiver designed to collect the P(Y) code on both L1 and L2 would have a bandwidth of 20.46 MHz, but a system tracking the C/A code or the future L2C or the C/A code may have a narrower bandwidth. It would need 2.046 MHz for the central lobe of the C/A code, or if it were designed to track the future L1C signal its bandwidth would need to be 4.092 MHz. A dual frequency microstrip antenna would likely operate in a bandwidth from 10 to 20 MHz.

Nearly Hemispheric Coverage

Since a GPS antenna is designed to be omnidirectional, its gain pattern, that is the change in gain over a range of azimuths and elevations, ought to be nearly a full hemisphere, but not perfectly hemispheric. First, most surveying applications filter the signals from very low elevations to reduce the effects of multipath and atmospheric delays. In other words, a portion of the GPS signal may come into the antenna from below the mask angle; therefore the antenna's gain pattern is specifically designed to reject such signals. Second, the contours of equal phase around the antenna's electronic center, that is, the *phase center*, are not themselves perfectly spherical.

A gain of about 3 to 5 *decibels* (*dB*) is typical for a GPS antenna. The dB is a unit for the logarithmic measure of the relative power of a signal. It is one-tenth of a bel, a unit named for Alexander Graham Bell. A dBW indicates actual power of a signal compared to a reference of 1 Watt (W), but the decibel alone is dimensionless. It is a unit of comparison like a ratio. In the context of an antenna if the decibel is defined by comparison with an idealized *isotropic* antenna it is written *dBi* or *dBic*. An isotropic antenna is a hypothetical lossless antenna that has equal capabilities in all directions.

Antenna Orientation

In a perfect GPS antenna, the phase center of the gain pattern would be exactly coincident with its actual, physical, center. If such a thing were possible, the centering of the antenna over a point on the earth would ensure its electronic centering as well. But that absolute certainty remains elusive for several reasons.

It is important to remember that the position at each end of a GPS baseline is the position of the phase center of the antenna at each end, not their physical centers, and the phase center is not an immovable point. The location of the phase center actually changes slightly with the satellite's signal. For example, it is different for the L2 than for L1. In addition, as the azimuth, intensity, and elevation of the received signal changes, so does the difference between the phase center and the physical center. Small azimuthal effects can also be brought on by the local environment around the antenna. But most phase center variation is attributable to changes in satellite elevation. In the end, the physical center and the phase center of an antenna may differ as much as a couple of centimeters. On the other hand with today's patch antennas it can be as little as a few millimeters.

It is also fortunate that the errors are systematic, and to compensate for some of this offset error most receiver manufacturers recommend users take care when making simultaneous observations on a network of points that their antennas are all oriented in the same direction. Several manufacturers even provide reference marks on the antenna so that each one may be rotated to the same azimuth, usually north to maintain the same relative position between their physical and phase, electronic, centers. Another approach to the problem's reduction is adjusting the phase center offset out of the solution in postprocessing.

Height of Instrument

The antenna's configuration also affects another measurement critical to successful GPS surveying—the height of instrument. The measurement of the height of the instrument in a GPS survey is normally made to some reference mark on the antenna. However, it sometimes must include an added correction to bring the total vertical distance to the antenna's phase center.

THE RADIO FREQUENCY (RF) SECTION

Different receiver types use different techniques to process the GPS signals, but go through substantially the same steps that are contained in this section.

The preamplifier increases the signal's power, but it is important that the gain in the signal coming out of the preamplifier is considerably higher than the noise. Since signal processing is easier if the signals arriving from the antenna are in a common frequency band, the incoming frequency is combined with a signal at a harmonic frequency. This latter, pure sinusoidal signal is the previously mentioned reference signal generated by the receiver's oscillator. The two frequencies are multiplied together in a device known as a *mixer*. Two frequencies emerge: one of them is the sum of the two that went in, and the other is the difference between them.

The sum and difference frequencies then go through a *bandpass filter*, an electronic filter that removes the unwanted high frequencies and selects the lower of the two. It also eliminates some of the noise from the signal. For tracking the P-code this filter will have a bandwidth of about 20 MHz, but it will be around 2 MHz if the C/A code is required. In any case, the signal that results is known as the *intermediate frequency* (*IF*), or beat frequency signal. This beat frequency is the difference between the Doppler-shifted carrier frequency that came from the satellite and the frequency generated by the receiver's own oscillator. In fact, to make sure that it embraces the full range of the Doppler effect on the signals coming in from the GPS satellites, the bandwidth of the IF itself can vary from 5 to 10 kHz Doppler. That spread is typically lessened after tracking is achieved.

There are usually several IF stages before copies of it are sent into separate channels, each of which extract the code and carrier information from a particular satellite.

As mentioned before, a replica of the C/A or P code is generated by the receiver's oscillator and now that is correlated with the IF signal. It is at this point that the pseudorange is measured. Remember, the pseudorange is the time shift required to align the internally generated code with the IF signal, multiplied by the speed of light.

The receiver also generates another replica, this time a replica of the carrier. That carrier is correlated with the IF signal and the shift in phase can be measured. The continuous phase observable, or observed cycle count, is obtained by counting the elapsed cycles since lock-on and by measuring the fractional part of the phase of the receiver generated carrier.

Channels

The antenna itself does not sort the information it gathers. The signals from several satellites enter the receiver simultaneously. But in the *channels* of the RF section the undifferentiated signals are identified and segregated from one another.

A channel in a GPS receiver is not unlike a channel in a television set. It is hardware, or a combination of hardware and software, designed to separate one signal from all the others. At any given moment, only one frequency from one satellite can be on one channel at a time. A receiver may have as few as 3 or as many as 40 physical channels. Today, 12 channels are typical. A receiver with 12 channels is also known as a 12 channel *parallel* receiver. Such a receiver may actually have 24 channels if it is a dual frequency receiver.

Multiplexing and Sequencing

While a parallel receiver has dedicated separate channels to receive the signals from each satellite that it needs for a solution, a *multiplexing* receiver gathers some data from one satellite for awhile and then switches to another satellite and gathers more data and so on. Such a receiver can usually perform this switching quickly enough that it appears to be tracking all of the satellites simultaneously. And when a GPS receiver's channels are not continuously dedicated to just one satellite's signal or one frequency, it is either known as a multiplexing or a *sequencing* receiver. These are also known as *fast-switching* or *fast-multiplexing* receivers. A multiplexing receiver must still dedicate one frequency from one satellite to one channel at a time, it just makes that time very short. For example, one channel may be used to track the signal from one satellite for only 20 milliseconds, leave that signal and track another for 20 milliseconds, and then return to the first, or even move on to a third.

Even though multiplexing is generally less expensive, this strategy of switching channels is now virtually obsolete. There are four reasons. While a parallel receiver does not necessarily offer more accurate results, parallel receivers with dedicated channels are faster; a parallel receiver has a more certain phase lock; and there is redundancy if a channel fails and they possess a superior signal-to-noise ratio (SNR). Whether continuous or switching channels are used, a receiver must be able to discriminate between the incoming signals. They may be differentiated by their unique C/A codes on L1, their Doppler shifts, or some other method, but in the end each signal is assigned to its own channel.

Tracking Loops

There are *code tracking loops, the delay lock loops* (*DLL*) and *carrier tracking loops,* and *the phase locking loops* (*PLL*). The tracking loops are connected to each of the receiver's channels and also work cooperatively with each other. As mentioned, dual frequency receivers have dedicated channels and tracking loops for each frequency.

Pseudoranging

Today, in most receivers the first procedure in processing an incoming satellite signal is synchronization of the C/A code from the satellite's L1 broadcast, with a replica C/A code generated by the receiver itself. The details of this process, also known as a *code-phase* measurement, are more fully described in Chapter 1. But for the purpose of this discussion, recall that when there is no initial match between the satellite's code and the receiver's replica, the receiver time shifts, or *slews*, the code it is generating until the optimum correlation is found. Then a code tracking loop, the delay lock loop, keeps them aligned.

The time shift discovered in that process is a measure of the signal's travel time from the satellite to the phase center of the receiver's antenna. Multiplying this time delay by the speed of light gives a range. But it is called a pseudorange in recognition of the fact that it is contaminated by the errors and biases set out in Chapter 2.

Carrier Phase Measurement

Once the receiver acquires the C/A code it has access to the Navigation message. It can read the ephemeris and the almanac information, use GPS time, and, for those receivers that do utilize the P code, use the hand-over word on every subframe as a stepping stone to tracking the more precise code. But the code's pseudoranges alone are not adequate for the majority of surveying applications. Therefore, the next step in signal processing for most receivers involves the carrier phase observable.

As stated earlier, just as they produce a replica of the incoming code, receivers also produce a replica of the incoming carrier wave. And the foundation of carrier phase measurement is the combination of these two frequencies. Remember, the incoming signal from the satellite is subject to an ever-changing Doppler shift while the replica within the receiver is nominally constant.

Carrier Tracking Loop

The process begins after the PRN code has done its job and the code tracking loop is locked. By mixing the satellite's signal with the replica carrier, this process eliminates all the phase modulations, strips the codes from the incoming carrier, and simultaneously creates two intermediate or beat-frequencies—one is the sum of the combined frequencies, and the other is the difference. The receiver selects the latter, the difference, with a device known as a *bandpass filter*. Then this signal is sent on to the *carrier tracking loop* also known as the phase locking loop, PLL, where the voltage-controlled oscillator is continuously adjusted to follow the beat frequency exactly.

Doppler Shift

As the satellite passes overhead, the range between the receiver and the satellite changes. That steady change is reflected in a smooth and continuous movement of the phase of the signal coming into the receiver. The rate of that change is reflected in the constant variation of the signal's Doppler shift. But if the receiver's oscillator frequency is matching these variations exactly, as they are happening, it will duplicate the incoming signal's Doppler shift and phase. This strategy of making measurements using the carrier beat phase observable is a matter of counting the elapsed cycles and adding the fractional phase of the receiver's own oscillator. Some of the details of the process are more fully described in Chapter 1.

Range rate. Doppler information has broad applications in signal processing. It can be used to discriminate between the signals from various GPS satellites, to determine integer ambiguities in kinematic surveying, as a help in the detection of cycle slips, and as an additional independent observable for autonomous point positioning. But perhaps the most important application of Doppler data is the determination of the *range rate* between a receiver and a satellite. *Range rate* is a term used to mean the rate at which the range between a satellite and a receiver changes over a particular period of time.

TYPICAL GPS DOPPLER SHIFT

With respect to the receiver, the satellite is always in motion even if the receiver is *static*. But the receiver may be in motion in another sense, as it is in kinematic GPS.

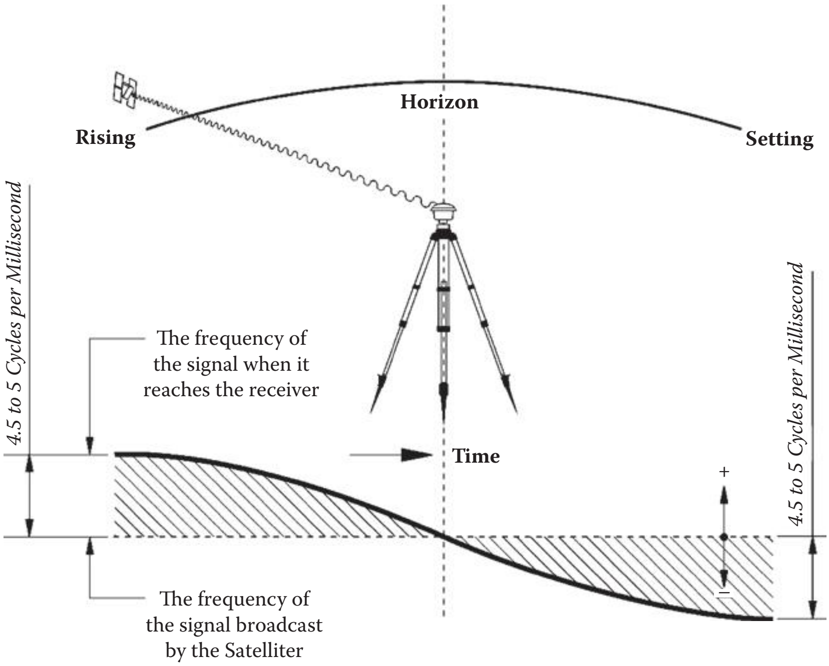


FIGURE 4.2 Typical Doppler Shift

The ability to determine the instantaneous velocity of a moving vehicle has always been a primary application of GPS and is based on the fact that the Doppler-shift frequency of a satellite's signal is nearly proportional to its range rate (Figure 4.2).

The Typical Change in the Doppler Shift

To see how it works, let's look at a *static*, that is, stationary, GPS receiver. The signal received would have its maximum Doppler shift, 4.5 to 5 cycles per millisecond, when the satellite is at its maximum range, just as it is rising or setting. But the Doppler shift continuously changes throughout the overhead pass. Immediately after the satellite rises, relative to a particular receiver, its Doppler shift gets smaller and smaller, until the satellite reaches its closest approach. At the instant its radial velocity with respect to the receiver is zero, the Doppler shift of the signal is zero as well. But as the satellite recedes, it grows again, negatively, until the Doppler shift once again reaches its maximum extent just as the satellite sets.

Continuously Integrated Doppler

The Doppler-shift and the carrier phase are measured by first combining the received frequencies with the nominally constant reference frequency created by the receiver's oscillator. The difference between the two is the often mentioned *beat frequency,* an intermediate frequency, and the number of beats over a given time interval is known as the *Doppler count* for that interval. Since the beats can be counted much more precisely than their continuously changing frequency can be measured, most GPS receivers just keep track of the accumulated cycles, the Doppler count. The sum of consecutive Doppler counts from an entire satellite pass is often stored, and the data can then be treated like a sequential series of biased range differences. *Continuously integrated Doppler* is such a process. The rate of the change in the continuously integrated Doppler shift of the incoming signal is the same as that of the reconstructed carrier phase.

Integration of the Doppler frequency offset results in an accurate measurement of the advance in carrier phase between epochs. And as stated earlier, using double-differences in processing the carrier phase observables removes most of the error sources other than multipath and receiver noise. But, as presented in Chapter 2, after the double difference, the integer ambiguity still remains.

Integer Ambiguity

The solution of the integer ambiguity, the number of whole cycles on the path from satellite to receiver, is not easy. But it would be much harder if it was not preceded by pseudoranges, or code phase measurements in most receivers. This allows the centering of the subsequent double-difference solution.

After the code-phase measurements narrows the field there are three methods used to solve the integer ambiguity. The first is a sort of geometric method. The carrier phase data from multiple epochs is processed and the constantly changing satellite geometry is used to find an estimate of the actual position of the receiver. This approach is also used to show the error in the estimate by calculating how its results hold up as the geometry of the constellation changes. It works pretty well, but depends on a significant amount of satellite motion, and therefore, takes time to converge on a solution.

The second approach uses filtering. Here independent measurements are averaged to find the estimated position with the lowest noise level.

The third uses a search through the range of possible integer combinations, and then calculates the one with the lowest residual. The search and filtering methods depend on *heuristic* calculations, in other words, trial and error. These approaches cannot assess the correctness of a particular answer, but can calculate the probability, given certain conditions, that the answer is within a specified set of limits.

In the end most GPS processing software uses some combination of all three ideas. All of these methods narrow the field by beginning at an initial position estimate provided by code-phase measurements.

Signal Squaring

There is a method that does not use the codes carried by the satellite's signal. It is called codeless tracking, or *signal squaring*. It was first used in the earliest civilian GPS receivers, supplanting proposals for a TRANSIT-like Doppler solution. It makes no use of pseudoranging and relies exclusively on the carrier phase observable. Like other methods it also depends on the creation of an intermediate or beat frequency. But with signal squaring the beat frequency is created by multiplying the incoming carrier by itself. The result has double the frequency and half the wavelength of the original. It is squared.

There are some drawbacks to the method. For example, in the process of squaring the carrier, it is stripped of all its codes. The chips of the P code, the C/A code, and the Navigation message normally impressed onto the carrier by 180° phase shifts are eliminated entirely. As discussed in Chapter 1, the signals broadcast by the satellites have phase shifts called *code states* that change from +1 to –1 and vice versa, but squaring the carrier converts them all to exactly 1. The result is that the codes themselves are wiped out. Therefore, this method must acquire information such as almanac data and clock corrections from other sources. Other drawbacks of squaring the carrier include the deterioration of the signal-to-noise ratio because when the carrier is squared the background noise is squared too. And cycle slips occur at twice the original carrier frequency.

But signal squaring has its up side as well. It reduces susceptibility to multipath. It has no dependence on PRN codes and is not hindered by the encryption of the P code. The technique works as well on L2 as it does on L1, and that facilitates dual-frequency ionospheric delay correction. Therefore, signal squaring can provide high accuracy over long baselines.

So there is a cursory look at some of the different techniques used to process the signal in the RF section. Now, onto the microprocessor of the receiver.

THE MICROPROCESSOR

The microprocessor controls the entire receiver, managing its collection of data. It controls the digital circuits that in turn manage the tracking and measurements, extract the ephemerides and other information from the Navigation message, and mitigate multipath and noise among other things.

The GPS receivers used in surveying often send these data to the storage unit. But more and more they are expected to process the ranging data, do datum conversion, and produce their final positions instantaneously, that is, in *real-time*. And then serve up the position through the control and display unit (CDU). There is a two-way street between the microprocessor and the CDU each can receive information from or send information to the other.

Differential Positioning

There are applications of GPS in which the receiver's microprocessor is expected to provide autonomous single-point positioning using unsmoothed code pseudoranges such as those from inexpensive handheld GPS units. Even though some manufacturers and users make extraordinary claims for their handheld coarse/acquisition (C/A) code pseudorange receivers, such autonomous point solutions are not accurate by surveying standards. But the positions from such receivers have improved. For example, SA's distortion of the satellite clock used to reduce single C/A receiver autonomous positioning to an accuracy of ±100 meters. Since it has been switched off, such positioning can achieve typical accuracies of ±20 to ±40 meters. Still, these are not surveying instruments.

However, code-based pseudoranges using DGPS (differential GPS) can achieve good real-time, or postprocessed, mapping results. For example, DGPS is often used in collecting data for Geographical Information Systems (GIS).

Differential GPS

The type of differential positioning known as DGPS still depends on code pseudorange observations, but requires at least two receivers. One receiver is placed on a control station and another on an unknown position. They simultaneously track the same codes from the same satellites and because many of the errors in the observations are common to both receivers, the errors are correlated and tend to cancel each other to some degree.

The data from such an arrangement is usually postprocessed, although with a radio link results can be had in real-time. Improvements in this technology have refined the technique's accuracy markedly, and meter- or even submeter results are possible. Still the positions are not as reliable as those achieved with the carrier phase observable.

Positional Accuracies

With GIS, corner search and mapping work excepted, much GPS surveying requires a higher standard of accuracy. Certainly GPS control surveying often employs several static receivers that simultaneously collect and store data from the same satellites for a period of time known as a *session*. After all the sessions for a day are completed their data are usually downloaded in a general binary format to the hard disk of a PC for postprocessing.

Kinematic and Real-Time Kinematic

However, not *all* GPS surveying is handled this way. For example, methods such as kinematic surveying and real-time kinematic surveying also use radio links and the carrier phase observable between GPS receivers. These methods can provide very good results indeed, more about that later.

THE CONTROL AND DISPLAY UNIT (CDU)

Every GPS receiver has a control and display unit. From handheld keyboards to soft keys around a screen to digital map displays and interfaces to other instrumentation there are a variety of configurations. Nevertheless they all have the same fundamental purpose, facilitation of the interaction between the operator and the receiver's microprocessor. The CDU may be used to select different surveying methods and/or set its parameters, that is, epoch interval, mask angle, and antenna height. The CDU may offer any combination of help menus, prompts, datum conversions, readouts of survey results, and so forth.

Typical Displays

The information available from the CDU varies from receiver to receiver. But when four or more satellites are available they can generally be expected to display the PRN numbers of the satellites being tracked, the receiver's position in three dimensions, and velocity information. Most of them also display the dilution of precision and GPS time.

THE STORAGE

Most GPS receivers today have internal data logging. They use solid state memory cards. Most allow the user the option of connecting to a computer and having the data downloaded directly to the hard drive. Cassettes, floppy disks, and computer tapes are mostly things of the past in GPS receiver storage.

The amount of storage required for a particular session depends on several things: the length of the session, the number of satellites above the horizon, the epoch interval, and so forth. For example, presuming the amount of data received from a single GPS satellite is ~100 bytes per epoch, a typical 12-channel dual-frequency receiver observing 6 satellites and using a 1-second epoch interval over the course of a 1-hour session would require ~2MB of storage capacity for that session.

$$3600 \text{ epochs per hour} \times 1 \text{ hour} \times 6 \text{ satellites} \times 100 \text{ bytes per satellite} = \sim 2 \text{ MB}$$

Downloading

A large number of range measurements and other pertinent data are sent to the receiver's storage during observation sessions. These data are subsequently downloaded through a serial port to a PC or laptop computer to provide the user with the option of postprocessing.

THE POWER

Battery Power

Since most receivers in the field operate on battery power, batteries and their characteristics are fundamental to GPS surveying. A variety of batteries are used and there are various configurations. For example, some GPS units are powered by camcorder batteries, and handheld recreational GPS units often use disposable AA or even AAA batteries. However, in surveying applications rechargeable batteries are the norm. Lithium, Nickel Cadmium, and Nickel Metal-Hydride may be the most common categories, but lead-acid car batteries still have an application as well.

The obvious drawbacks to lead-acid batteries are size and weight. And there are a few others—the corrosive acid, the need to store them charged, and their low cycle life. Nevertheless lead-acid batteries are especially hard to beat when high power is required. They are economical and long lasting.

Nickel Cadmium batteries (NiCd) cost more than lead-acid batteries but are small and operate well at low temperatures. Their capacity does decline as the temperature drops. Like lead-acid batteries, NiCd batteries are quite toxic. They self-discharge at the rate of about 10% per month and even though they do require periodic full discharge these batteries have an excellent cycle life. Nickel Metal-Hydride (NiMH) batteries self-discharge a bit more rapidly than NiCd batteries and have a less robust cycle life, but are not as toxic.

Lithium-ion batteries overcome several of the limitations of the others. They have a relatively low self-discharge rate. They do not require periodic discharging and do not have a memory issues as do NiCd batteries. They are light, have a good cycle life, and low toxicity. On the other hand, the others tolerate overcharging and the lithium-ion battery does not. It is best to not charge lithium-ion batteries in temperatures at or below freezing. These batteries require a protection circuit to limit current and voltage but are widely used in powering electronic devices, including GPS receivers.

About half of the available GPS carrier phase receivers have an internal power supply and most will operate 5½ hours or longer on fully charged 6-amp-hour battery. Most code-tracking receivers, those that do not also use the carrier phase observable, could operate for about 15 hours on the same size battery.

It is fortunate that GPS receivers operate at low power; from 9 to 36 volts DC is generally required. This allows longer observations with fewer, and lighter, batteries than might be otherwise required. It also increases the longevity of the GPS receivers themselves.

CHOOSING A GPS RECEIVER

The first question is what observable is to be tracked? There are receivers that use only the C/A code on the L1 frequency and receivers that cross-correlate with the P code, or encrypted Y code, on both frequencies. There are L1 carrier phase tracking receivers and there are dual-frequency carrier phase tracking receivers.

The C/A code only. Generally speaking recreational handheld GPS receivers' positions are accurate to approximately ±10 to ±30 meters with SA turned off. Nevertheless, their use is growing, and while most are not capable of tracking the carrier phase observable some have differential capability. Still these receivers were developed with the needs of navigation in mind. In fact, they are sometimes categorized by the number of waypoints they can store.

Waypoint is a term that grew out of military usage. It means the coordinate of an intermediate position a person, vehicle, or airplane must pass to reach a desired destination. With such a receiver, a navigator may call up a distance and direction from his present location to the next waypoint.

These receivers do not have substantial memory. It is not needed for their designed applications. However, considerable memory is required for differential GPS, DGPS receivers. It is still a receiver that only tracks the C/A code. But for many applications it must be capable of collecting the same information as is simultaneously collected at a base station on a known point, and storing it for postprocessing. Such an arrangement is capable of meter-level accuracy.

The C/A code and phase on L1. Relative GPS on lines of less than ten miles can be accomplished using at least two receivers that track the C/A-code and phase of L1. Using double differencing this application can provide accuracy of approximately ±(1 cm + 2 ppm).

Carrier phase on both frequencies plus the C/A-code. A receiver that is capable of observing the carrier phase of both frequencies and the C/A-code is appropriate for collecting positions on long base lines, more than 10 miles.

Receiver productivity. Still choosing the right instrument for a particular application is not easy. Receivers are generally categorized by their physical characteristics, the elements of the GPS signal they can use with advantage, and by the claims about their accuracy. But the effects of these features on a receiver's actual productivity are not always obvious.

For example, it is true that the more aspects of the GPS signal a receiver can employ, the greater its flexibility, but so too the greater its cost. And separating the capabilities a user needs to do a particular job from those that are really unnecessary is more complicated than simply listing tracking characteristics, storage capacity, power consumption, and other physical statistics. The user must first have some information about how these features relate to a receiver's performance in particular GPS surveying methods.

Typical concerns. A surveyor generally wants to be able to answer questions like the following. Can this receiver give me the accuracy I need for my work? What is the likely rate of productivity on a project like this? What is the actual cost per point? This section is intended to provide some of the information needed to answer such questions. It offers some general guidelines that are useful in choosing a GPS receiver.

TRENDS IN RECEIVER DEVELOPMENT

In the early years of GPS, the military concentrated on testing navigation receivers. But civilians got involved much sooner than expected and took a different direction: receivers with geodetic accuracy.

High Accuracy Early

The first GPS receivers in commercial use were single frequency, six channel, and codeless instruments. Their measurements were based on interferometry. As early as the 1980s, those receivers could measure short baselines to millimeter accuracy and long baselines to 1 ppm. It is true the equipment was cumbersome, expensive, and, without access to the Navigation message, dependent on external sources for clock and ephemeris information. But they were the first at work in the field and their accuracy was impressive.

Another Direction

During the same era a parallel trend was underway. The idea was to develop a more portable, dual-frequency, four-channel receiver that could use the Navigation message. Such an instrument did not need external sources for clock and ephemeris information and could be more self-contained.

More Convenience

Unlike the original codeless receivers that required all units on a survey brought together and their clocks synchronized twice a day, these receivers could operate independently. And while the codeless receivers needed to have satellite ephemeris information downloaded before their observations could begin, this receiver could derive its ephemeris directly from the satellite's signal. Despite these advantages, the instruments developed on this model still weighed more than 40 pounds, were very expensive, and were dependent on P code tracking.

Multichannel and Code-Correlating

A few years later a different kind multichannel receiver appeared. Instead of the P code, it tracked the C/A code. Instead of using both the L1 and L2 frequencies, it depended on L1 alone. And on that single frequency, it tracked the C/A code and also measured the carrier phase observable. This type of receiver established the basic design for the many of the GPS receivers that surveyors use today. They are multichannel receivers, and they can recover all of the components of the L1 signal. The C/A code is used to establish the signal lock and initialize the tracking loop. Then the receiver not only reconstructs the carrier wave, it also extracts the satellite clock correction, ephemeris, and other information from the Navigation message. Such receivers are capable of measuring pseudoranges, along with the carrier phase and integrated Doppler observables.

Dual-Frequency

Still, as some of the earlier instruments illustrated, the dual-frequency approach does offer significant advantages. It can be used to limit the effects of ionospheric delay, it can increase the reliability of the results over long baselines, and it certainly increases the scope of GPS procedures available to a surveyor. For these reasons, a substantial number of receivers utilize both frequencies.

Receiver manufacturers are currently using several configurations in building dual-frequency receivers. In order to get both carriers without knowledge of the P code, some use a combination of C/A code-correlation and codeless methods. These receivers use code correlation on L1 and then, borrowing an idea from the first GPS receivers, they add signal squaring on L2.

Adding Codeless Capability

By adding codeless technology on the second frequency, such receivers can avail themselves of the advantages of a dual-frequency capability while avoiding the difficulties of the P code.

Antispoofing, or AS, is the encryption of the P code on both L1 and L2. These encrypted codes are known as the Y codes, Y1 and Y2, respectively. But even though the P code has been encrypted, the carrier phase and pseudorange P code observables have been recovered successfully by many receiver manufacturers.

Adding P Code Tracking

Dual-frequency receivers are fast becoming the standard for geodetic applications of GPS, and some do utilize the P code or the encrypted Y code. There are no civilian receivers that rely solely on the P code. Nearly all receivers that use the P code use the C/A code on L1. Some use the C/A code on L1 and codeless technology on L2. Some use the P code only when it is not encrypted and become codeless on L2 when AS is activated. Finally, some track all the available codes on both frequencies. In any of these configurations, the GPS surveying receivers that use the P code in combination with the C/A code and/or codeless technology are among the most expensive.

Typical GPS Surveying Receiver Characteristics

These examples represent the general scope of receivers that provide a level of accuracy acceptable for most surveying applications. Most share some practical characteristics: they have multiple independent channels that track the satellites continuously, and they begin acquiring satellites' signals from a few seconds to less than a minute from the moment they are switched on. Most acquire all the satellites above their mask angle in a very few minutes, with the time usually lessened by a warm start, and most provide some sort of audible tone to alert the user that data is being recorded, and so forth. About three-quarters of them can have their sessions preprogrammed in the office before going to their field sites. And nearly all allow the user to select the *sampling rate* of their phase measurements, from 0 to 3600 seconds.

GPS Receiver Costs

Comparing the cost of GPS receivers over time is complicated by the fact that, today, postprocessing software is often included in their prices. Nevertheless, it is clear that the cost of GPS receivers and GPS technology overall has been through a remarkable decline. The first GPS receivers were five times more expensive than the highest-priced receiver available today. At the same time, the capabilities of even an average receiver have come to outstrip those of the best of the early instruments. These trends will undoubtedly continue, but perhaps the more important aspect of receiver development is the increase in their variety. This growth in diversity has been driven by the rapid expansion in the scope of the uses of GPS.

As previously mentioned, the GPS work done in geodetic and land surveying relies on post-processed relative positioning. There are several very different techniques available to GPS surveyors. And each method makes unique demands on the receivers used to support it.

SOME GPS SURVEYING METHODS

Static

This was the first method of GPS surveying used in the field and it continues to be the primary technique today (see Table 4.1). Relative static positioning involves several stationary receivers simultaneously collecting data from at least 4 satellites during observation sessions that usually last from 30 minutes to 2 hours. A typical application of this method would be the determination of vectors, or baselines as they are called, between several static receivers to accuracies from 1 ppm to 0.1 ppm over tens of kilometers (Figure 4.3).

Prerequisites for Static GPS

There are few absolute requirements for relative static positioning. The requisites include: more than one receiver, four or more satellites, and a mostly unobstructed sky above the stations to be occupied. But as in most of surveying, the rest of the elements of the system are dependent on several other considerations.

TABLE 4.1
Comparison of GPS Survey Methods

Technique	Accuracy	Observation Time	Drawbacks	Strengths
Static	1/100,000 to 1/5,000,000	1 to 2 hrs	Slow	High accuracy
Kinematic	1/100,000 to 1/750,000	1 to 2 mins	Requires constant lock on at least 4 satellites; needs initialization	Fast
Rapid static	1/100,000 to 1/1,000,000	5 to 20 mins	Requires the most sophisticated equipment	Very fast and accurate
On-the-fly (OTF)	1/100,000 to 1/1,000,000	Virtually instantaneous positioning		Allows highly accurate positions from a receiver in motion
Pseudokinematic	1/50,000 to 1/500,000	10 to 20 mins; 2 observations, 1 to 4 hrs apart	Requires 2 separate observations per point	More productive than static

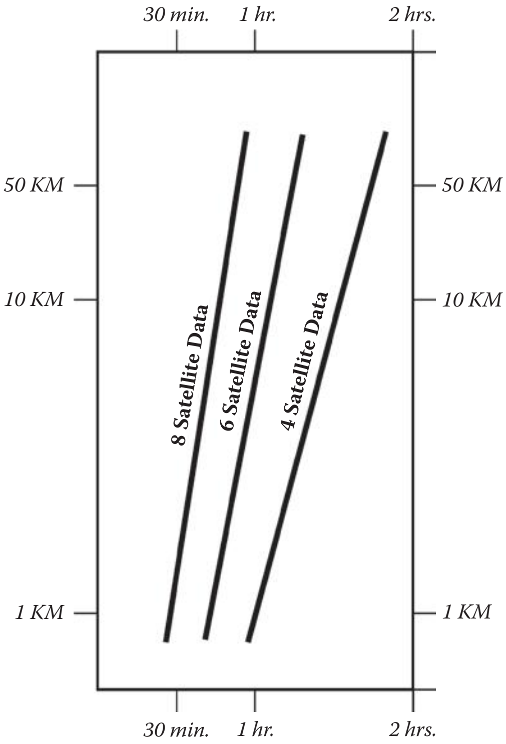


FIGURE 4.3 Generalized Session Lengths for Static Baselines

Productivity

The assessment of the productivity of a GPS survey almost always hinges, in part at least, on the length of the observation sessions required to satisfy the survey specifications. The determination of the session's duration depends on several particulars, such as the length of the baseline and the relative position, that is the geometry, of the satellites, among others.

Session Length

Generally speaking, the larger the constellation of satellites, the better the available geometry, the lower the positioning dilution of precision (PDOP), and the shorter the length of the session needed to achieve the required accuracy. For example, given 6 satellites and good geometry, baselines of 10 km or less might require a session of 45 minutes to 1 hour, whereas, under exactly the same conditions, a baseline over 20 km might require a session of 2 hours or more. Alternatively, 45 minutes of 6-satellite data may be worth an hour of 4-satellite data, depending on the arrangement of the satellites in the sky.

Large Amounts of Data

A static receiver gathers a huge amount of data during an observation session of 1 or 2 hours. Why does it need so much information? The answer may be a bit of a surprise. One might expect there to be some sort of time-consuming process that gradually refines the receiver's measurements down to the last millimeter. The situation is quite different. In fact, the receiver actually achieves the millimeter level of accuracy in a matter of seconds, in as little as a single epoch. It is the resolution of the larger divisions of the measurement, the meters, for example, that requires long observation so typical of relative static positioning.

Resolution of the Cycle Ambiguity

The receiver does not need long sessions to make the fine distinctions between millimeters. It needs long sessions to solve the integer number of cycles between itself and the satellites, the so-called *cycle ambiguity* problem. In fact, it is the unique handling of precisely this difficulty that allows the kinematic method to achieve high accuracy with very much shorter occupation times. In the static application, the receivers must resolve the phase ambiguity anew with each occupation, so the sessions are long. But in the kinematic method the receiver resolves the phase ambiguity once, and only once, at the beginning of the project. Then by keeping a continuous lock on the satellite's signals, it maintains that solution throughout the work. This strategy allows for short occupations without sacrificing accuracy. (See Chapter 1 for a more thorough presentation of the cause of the cycle ambiguity problem.)

Preprogrammed Observations

Many receivers offer the user facility to preprogram parameters at the occupied station. Some are so automatic they do not require operator interaction at all once they are programmed and on-site, a feature that can be somewhat of a mixed blessing. This will be investigated in more detail in the section on planning a GPS survey.

Observation Settings

The selection of satellites to track, start and stop times, mask elevation angle, assignment of data file names, reference position, bandwidth, and sampling rate are some options useful in the static mode, as well as other GPS surveying methods. These features may appear to be prosaic, but their practicality is not always obvious. For example, satellite selection can seem unnecessary when using a receiver with sufficient independent channels to track all satellites above the receiver's horizon without difficulty. However, it is quite useful when the need arises to eliminate data from a satellite that is known to be unhealthy before the observation begins.

Data Interval

Another example is the selectable sampling rate, also known as the *data interval*. This feature allows the user to stipulate the short period of time between each of the microprocessor's downloads to storage. The fastest rate available is usually between 0 and 1 second, and the slowest, 999 and 3600 seconds. The faster the data-sampling rate, the larger the volume of data a receiver collects and the larger the amount of storage it needs. A fast rate is helpful in cycle slip detection, and that improves the receiver's performance on baselines longer than 50 km, where the detection and repair of cycle slips can be particularly difficult.

Compatible Receivers

Relative static positioning, just as all the subsequent surveying methods discussed here, involves several receivers occupying many sites. Problems can be avoided as long as the receivers on a project are compatible. For example, it is helpful if they have the same number of channels and signal processing techniques. And the Receiver Independent Exchange Format (RINEX), developed by the Astronomical Institute in 1989, allows different receivers and postprocessing software to work together. Almost all GPS processing software will output RINEX files.

Receiver Capabilities and Baseline Length

The number and type of channels available to a receiver is a consideration because, generally speaking, the more satellites the receiver can track continuously the better. Another factor that ought to be weighed is whether a receiver has single or dual-frequency capability. Single-frequency receivers are best applied to relatively short baselines, say, under 25 km. The biases at one end of such a vector are likely to be similar to those at the other. Dual-frequency receivers, on the other hand, have the capability to nearly eliminate the effects of ionospheric refraction, and can handle longer baselines.

Static GPS surveying has been used on control surveys from local to statewide extent, and will probably continue to be the preferred technique in that category.

DIFFERENTIAL GPS

DGPS is a method that improves GPS pseudorange accuracy. A GPS receiver at a base station, a known position, measures pseudoranges to the same satellites at the same time as other roving GPS receivers. The roving receivers occupy unknown positions in the same geographic area. Occupying a known position, the base station receiver finds corrective factors that can either be communicated in real-time to the roving receivers using a radio link, or may be applied in postprocessing.

Real-time DGPS is somewhat limited to near line-of-sight, whereas post-processed DGPS can provide meter or submeter positions up to about 300 km from the base station. DGPS is widely used to collect data for GIS applications. There will be more about DGPS in Chapter 6.

KINEMATIC

Reference Receiver and Rover Receivers

As previously mentioned, kinematic positioning is faster than the static method. The term kinematic is sometimes applied to GPS surveying methods where receivers are in continuous motion, but for relative positioning the more typical arrangement is a stop-and-go technique, a method developed by Dr. Benjamin Remondi.

This latter approach involves the use of at least one stationary reference receiver and at least one moving receiver, called a *rover*. The technique is similar to static GPS in that all the receivers observe the same satellites simultaneously, and the reference receivers occupy the same control points throughout the survey. However, the kinematic method differs from static GPS in the movement of the rovers from point to point across the network. They stop momentarily at each new point, usually very briefly, and their data eventually provide vectors between themselves and the reference receivers.

Leapfrog Kinematic

There is a variation of this technique involving two receivers. It is known as the *leapfrog* approach. Instead of either of the receivers remaining a motionless reference throughout the survey, they both move, one at a time. In its simplest form, the receivers take the roving in turns. First receiver A is stationary and receiver B moves; then receiver B waits while receiver A moves. But since they never move at the same time they each take a turn to ensure that at any given moment one of them is providing a static anchor for each vector in the survey, however briefly.

Kinematic Positional Accuracy

When compared with the other relative positioning methods, there is little question that the very short sessions of the kinematic, or real-time kinematic, method can produce the largest number of positions in the least amount of time. The remarkable thing is that this technique can do so with only slight degradation in the accuracy of the work. Recall that in static GPS, the long sessions are not required to resolve the finest aspect of the carrier phase measurement. The millimeters are resolved in a matter of seconds. It is the determination of the meters, the resolution of the integer number of cycles or so-called *phase ambiguity* that takes time.

Maintaining Lock

Once the phase ambiguity is resolved with one of these initialization techniques, the receivers are kept running and their lock on the satellite's signals is carefully maintained throughout the survey by avoiding any obstructions that may interrupt them. The drawback of kinematic GPS is that not only must the occupied stations be free of overhead obstructions, the route between them must also be clear. Otherwise the lock could be lost while the roving receiver is moving from point to point. In that case, either the original initialization procedure would have to be repeated or the receiver can be returned to the last trouble-free position collected for reintialization on a known baseline.

Reconnaissance

Therefore, kinematic surveys require considerable reconnaissance. Trees, utility poles, buildings, fences—any sort of structure, natural or man made—that might prevent the satellite's signals from reaching the receiver for even an instant must be avoided to achieve maximum efficiency with the kinematic method. For this reason it is best to consider 5, rather than 4, satellites to be the absolute minimum number in planning a kinematic survey. Despite the best reconnaissance, the signal from one satellite may easily be lost during the work. If five are considered the minimum from the outset, such a loss is not nearly so disastrous, since four will remain.

Applications

Kinematic GPS is best suited for work done in wide-open areas. Several kinematic procedures have been developed for airborne and marine work, including photogrammetry without ground control. Land-based kinematic relative positioning applications range from gravity field surveys to construction work.

Most Carrier Phase Receivers Capable of Kinematic GPS

The majority of the receivers that measure the carrier phase observable are quite adequate to most kinematic applications and generally offer it, along with their static capability. Still, even within that wide selection, some receiver characteristics are more useful than others. For example, a fast data-sampling rate is essential in kinematic surveys. From 1 to 5 seconds is usual, but even faster rates are available in work that requires the receivers to be in continuous motion. An option of bandwidth selection in the tracking loops is also helpful in kinematic applications.

Wideband

A wideband signal is transmitted by the GPS satellites because it provides some immunity from jamming and is capable of high-resolution ranging. However, after it has been collected by the receiver, the spread-spectrum signal is usually compressed, or *despread*, into a smaller bandwidth by demodulating it with an identical wideband waveform. Having the option to select the bandwidth is helpful because the narrower it is the better the signal-to-noise ratio (SNR). But there is a trade-off: although the SNR for the narrower bandwidth is better, a wider bandwidth provides a more secure signal lock. In the end, receivers that can be adapted to find the best balance provide the best results.

Practical Considerations in Kinematic GPS

The physical characteristics of the receiver also need to be considered. When the work is likely to encounter heavy vegetation, a receiver with a separate antenna is a boon, particularly in the kinematic mode. Mounted on a range pole or mast, it can sometimes be elevated above the obstructions.

In kinematic work where the equipment must be packed in, two obvious factors in determining a receiver's suitability are its weight and rate of power consumption. A receiver that provides the user with an audible warning when lock is lost is also quite useful in kinematic applications. This feature may help a user avoid the continuation of a kinematic survey that must be reinitialized.

PSEUDOKINEMATIC

This technique has several names. It has come to be known as *pseudokinematic*, *pseudostatic,* and *intermittent-static*. The variety in its titles is indicative of its standing between static and kinematic in terms of productivity. It is less productive than kinematic, but more productive than the static method. While it is also somewhat less accurate than either relative static or kinematic positioning, its primary advantage is its flexibility. There is no need for the receiver to maintain a lock on the satellite's signals while moving from station to station.

Double Occupation

The pseudokinematic observation procedures can be virtually identical to those previously described for the kinematic relative positioning. However, there is a fundamental difference between the two. In pseudokinematic work, it is necessary to occupy the unknown stations twice. Each one of the two occupations may be as brief, but they should be separated by at least 1 hour and not more than 4 hours. The time between the two occupations allows the satellites of the second session to reach a configuration different enough from that of the first, for resolution of the phase ambiguity. This removes the necessity of maintaining a constant lock on the satellite's signals, as required in kinematic work. It also allows the user to occupy many more stations than would be possible with static positioning over the same time.

No Need for Continuous Lock

In pseudokinematic relative positioning, at least two stationary receivers collect up to 10 minutes of data from the same satellites at the same time. A different baseline is occupied by moving one or both of the receivers, without continuous lock. In fact, the receivers may even be switched off during the move. After the move they collect up to 10 minutes of data from the ends of the new baseline. This procedure continues until all of the baselines have been occupied. So far the method is very like other relative positioning techniques, but pseudokinematic differs from the others in that, following the initial occupations, the receivers must return to every one of the baselines in the same manner a second time.

Best Used in Easy Access Situations

While the accuracy of the technique is usually at the subcentimeter level, the production advantage of the procedure over the static mode is somewhat diminished by the reoccupation requirement. Therefore, the best application of pseudokinematic GPS is over relatively short baselines, under 10 km, where the access to the stations is quick and easy.

Mostly Radial Surveys

The majority of receivers that measure the carrier phase observable offer both kinematic and pseudokinematic surveying along with their static capability. Pseudokinematic, as well as kinematic, procedures tend to encourage radial surveys. While the radial approach may be useful for mapping, topography, photocontrol, and some low-order mining surveys, its open-ended quality may not be adequate in other work, like control surveys.

RAPID-STATIC

Also known as *fast-static*, this field procedure is like pseudokinematic in that it can succeed with very short occupation times. However, it is unlike pseudokinematic in that baselines positioned with rapid-static need be occupied only once. Neither does this procedure require continuous lock on the satellite's signals, as does the kinematic method.

Wide Laning

These advantages are accomplished by relying on the sophisticated hardware of the most expensive receivers. Success in rapid-static depends on the use of receivers that can combine code and carrier phase measurements from both GPS frequencies. The field procedures in rapid-static are very like those of the static method. But using a technique known as *wide laning*, rapid-static can provide the user with nearly the same accuracy available from 1- and 2-hour sessions of relative static positioning with observations of 5 to 20 minutes.

Wide laning is based on the linear combination of the measured phases from both GPS frequencies, L1 and L2. Carrier phase measurements can be made on L1 and L2 separately, of course, but when they are combined two distinct signals result. One is called a *narrow lane* and has a short wavelength of 10.7 cm. The other is known as the *wide lane*. Its frequency, only 347.82 MHz, is more than three times slower than the original carriers. Furthermore, its 86.2 cm wavelength is about 4 times longer than the 19.0 cm and 24.4 cm of L1 and L2, respectively. These changes greatly increase the spacing of the phase ambiguity, thereby making its resolution much easier.

However, the demands on the receiver are extensive and expensive. It must be capable of dual-frequency tracking and the noise on the signal increases by nearly 6 times. Nevertheless, wide laning is very useful for resolving the integer ambiguity very quickly, *on-the-fly* (*OTF*).

ON-THE-FLY

Wide laning is also at the heart of a GPS surveying technique called on-the-fly (OTF). This method allows initialization while the receiver is actually in motion.

Accurate Initial Positions

Just as in the rapid-static method, P code pseudoranges provide for the wide-lane integer bias resolution. That solution is then used, in turn, to solve the base carrier's integer ambiguity. The OTF process differs from rapid static in that the P code pseudoranges are electronically refined. At the top of the ambiguity resolution process, several consecutive P code pseudoranges are run through a Kalman filter, a kind of averaging process. This improves the initial position from which the wide-lane and base-carrier solutions work, giving the receiver a very accurate, virtually instantaneous idea of where it is. The high-quality starting point makes phase ambiguity resolution very fast.

In fact, these improved P code pseudoranges offer the user an initialization process that is so fast and accurate it can be accomplished while the receiver is in continuous motion. If the receiver momentarily loses lock on the satellite's signals, with OTF the lock can be reacquired without stopping. Fixing of the integer ambiguities for carrier phase observations on-the-fly is at the heart of precise real-time GPS positioning.

Photogrammetry without Ground Control

The accuracy of OTF is comparable with kinematic GPS, and it lends itself to radial surveys, but it is not crippled by momentary loss of four-satellite data. OTF is a great boon to aerial and hydrographic work and has made the prospect of photogrammetry without ground control a reality, and that is only the beginning of the possible applications of this method.

REAL-TIME KINEMATIC (RTK)

Like kinematic, here the base station occupies a known position. However, in the real-time kinematic method the base station actually transmits corrections to the roving receiver or receivers using a radio link. The procedure offers high accuracy immediately, in real-time. The results are not postprocessed.

In the earliest use of GPS, kinematic and rapid static positioning were not frequently used because ambiguity resolution methods were still inefficient. But now with integer ambiguity resolution such as OTF available, real-time kinematic, and similar surveying methods have become very widely used. In fact, today, many surveyors rely heavily on real-time kinematic surveying in their day to day work, more about that in Chapter 6.

EXERCISES

1. Which of the following is not a consideration in antenna design for GPS receivers?

    (a) Efficient conversion of electromagnetic waves into electric currents.
    (b) Directional capability.
    (c) Coincidence of the phase center and the physical center.
    (d) Reduction of the effects of multipath.

2. Which of the ideas listed below is intended to limit the effect of the difference between the phase center and the physical center of GPS antennas on a baseline measurement?

    (a) Ground plane antennas at each end of the baseline.
    (b) Choke ring antennas at each end of the baseline.
    (c) Rotation of the antenna's reference marks to north at each end of the baseline.
    (d) The use of the receiver's built-in antennas at each end of the baseline.

3. Which of the following does not describe an advantage of a 12-channel parallel continuous-tracking receiver with dedicated channels over a 1-channel multiplexing, or sequencing receiver?

    (a) The parallel continuous-tracking receiver has a superior signal-to-noise ratio.
    (b) The parallel continuous-tracking receiver is more accurate.
    (c) If a channel stops working, there is redundancy with a parallel continuous-tracking receiver.
    (d) The parallel continuous-tracking receiver has less frequent cycle slips.

4. Which of the following statements concerning the intermediate frequency (IF) in GPS signal processing is correct?

    (a) In the radio frequency (RF) section of a GPS receiver two frequencies go through a *bandpass filter,* which selects the higher of the two. This signal is then known as the *IF*, or intermediate frequency.
    (b) GPS signal processing usually has a single IF stage.
    (c) The intermediate frequency is a beat frequency.
    (d) The intermediate frequency is the sum of the Doppler-shifted carrier from the satellite and the signal generated by the receiver's own oscillator.

5. Which of the following is not a drawback of signal squaring?

   (a) The effect of multipath is increased.
   (b) The signal-to-noise ratio deteriorates.
   (c) The codes are stripped from the carrier.
   (d) The receiver must acquire information such as almanac data and clock corrections somewhere other than the Navigation message.

6. Which of the following is the closest to the length of a C/A code period that is 1,023 chips and 1 millisecond in duration?

   (a) 300 meters
   (b) 300 kilometers
   (c) 66 meters
   (d) 20,000 kilometers

7. What is an isotropic antenna?

   (a) A hypothetical lossless antenna that has equal capabilities in all directions.
   (b) A durable, compact antenna with a simple construction and a low profile.
   (c) An antenna built with several concentric rings and designed to reduce the effects of multipath.
   (d) An antenna that has a stable phase center and simple construction, but needs a good ground plane.

8. Which of the following is still used to store data in GPS receivers?

   (a) Cassette
   (b) Floppy disk
   (c) Tapes
   (d) Internal onboard memory

9. Which of the following two numbers correctly show the proper relationship between the wavelength of the L2 carrier and the wide lane, first, and the increase in the signal noise, second?

   (a) about 2 times longer, about 3 times noisier
   (b) about 4 times longer, about 6 times noisier
   (c) about 10 times slower, about 10 times noisier
   (d) 9.23 times faster, 18.16 times noisier

10. Which of the following procedures do kinematic GPS, DGPS, and RTK have in common?

    (a) Each method uses a radio link between the base station and the rover or rovers.
    (b) Each method uses a base station and rover arrangement.
    (c) Each method must maintain continuous lock on at least 4 satellites for continuous uninterrupted positioning.
    (d) Each method's results must be postprocessed.

ANSWERS AND EXPLANATIONS

1. Answer is (b)

Explanation: The GPS signal is actually quite weak and it is broadcast over a very large area. The efficiency of GPS antennas is an important consideration. Limiting the effects of multipath is also a very important consideration. The perfect coincidence of the phase center with the physical center in GPS antennas has yet to be achieved, but it is much sought after and is certainly a design consideration.

However, the GPS signal is a spread-spectrum coded signal that occupies a broader frequency bandwidth than it must to carry its information. This fact allows GPS antennas to be omnidirectional. They do not require directional orientation to properly receive the GPS signal.

2. Answer is (c)

Explanation: Ground planes used with many GPS antennas and the choke ring antenna are designs that attack the same problem, limiting the effects of multipath. The coincidence of the phase center with the physical center in GPS antennas is not yet perfected and since the measurements made by a GPS receiver are made to the phase center of its antenna, its orientation is paramount. The rotation of the antennas at each end of a baseline to north per an imprinted reference is a strategy to reduce the effect of the difference between the two points.

3. Answer is (b)

Explanation: Multiplexing receivers are at somewhat of a disadvantage when compared with continuous-tracking receivers. Continuous tracking receivers with dedicated channels have a superior signal-to-noise ratio (SNR). They are faster. There is redundancy if a channel fails. They have a more certain phase lock. There are fewer cycle slips.

A multiplexing receiver is not necessarily less accurate than a parallel continuous-tracking receiver.

4. Answer is (c)

Explanation: The intermediate frequency is definitely a beat frequency. It is the combination of the frequency coming from the satellite and a sinusoidal signal generated by the receiver's oscillator. These frequencies are multiplied together in a mixer which produces the sum of the two and the difference. Both go through a bandpass filter which selects the lower of the two and it is known as the intermediate frequency, or beat frequency signal. This beat frequency is the difference between the Doppler-shifted carrier frequency that came from the satellite and the frequency generated by the receiver's own oscillator. There are usually several IF stages in a GPS receiver.

5. Answer is (a)

Explanation: With signal squaring the beat or intermediate frequency is created by multiplying the incoming carrier by itself. The result has double the frequency and half the wavelength of the original. It is squared. The process of squaring the carrier strips off all the codes. The P code, the C/A code, and the Navigation message are not available to the receiver. And with the codes wiped out the receiver must acquire information such as almanac data and clock corrections from other sources. Also, the signal-to-noise ratio is degraded, increasing cycle slips. On the other hand, the effect of multipath is actually decreased. And since squaring works as well on L2 as it does on L1, dual-frequency ionospheric delay correction is possible.

6. Answer is (b)

Explanation: The C/A code "chipping rate," the rate at which each chip is modulated onto the carrier, is 1.023 Mbps. That means, at light speed, the chip length is approximately 300 m. But, the whole C/A code period is 1023 chips, 1 millisecond long. That is approximately 300 km. And, of course, each satellite repeats its whole 300 km C/A code over and over.

7. Answer is (a)

Explanation: An isotropic antenna is a hypothetical, lossless antenna that has equal capabilities in all directions. This theoretical antenna is used as a basis of comparison when expressing the gain of GPS and other antennas. A gain of about 3 dB (decibels) is typical for the usual omnidirectional GPS antenna. The gain, or gain pattern, describes the success of a GPS antenna in collecting more energy from above the mask angle, and less from below the mask angle. The decibel here does not indicate the power of the antenna; it refers to a comparison. In this case, 3 dB indicates that the GPS antenna has about 50% of the capability of an isotropic antenna.

8. Answer is (d)

Explanation: Most GPS receivers today have internal data logging. They use solid state memory or memory cards. Most also allow the user the option of connecting to a computer and having the data downloaded directly to the hard drive. Cassette, floppy disks, and computer tapes are mostly things of the past in GPS receiver storage.

9. Answer is (b)

Explanation: Wide laning uses the combination of GPS frequencies, L1 and L2. Carrier phase measurements can be made on L1 and L2 separately, of course, but when they are combined two distinct signals result. One is called a *narrow lane* and has a short wavelength of 10.7 cm. The other is known as the *wide lane*. Its frequency, only 347.82 MHz, is more than 3 times slower than the original carriers. Furthermore, its 86.2 cm wavelength is about 4 times longer than the 19.0 cm and 24.4 cm of L1 and L2, respectively and the noise on the signal increases by nearly 6 times.

10. Answer is (b)

Explanation: DGPS is a method that improves GPS pseudorange accuracy. A GPS receiver at a base station, usually at a known position, measures pseudoranges to the same satellites at the same time as other roving GPS receivers. The roving receivers occupy unknown positions in the same geographic area. Occupying a known position, the base station receiver finds corrective factors that can either be communicated in real-time to the roving receivers, or may be applied in postprocessing.

Kinematic GPS is a version of relative positioning in which one receiver is a stationary reference, the base station and at least one other roving receiver coordinates unknown positions with short occupation times while both track the same satellites and maintain constant lock. If lock is lost re-initialization is necessary to fix the integer ambiguity.

Real-time kinematic is a method of determining relative positions between known control and unknown positions using carrier phase measurements. A base station at the known position transmits corrections to the roving receiver or receivers. The procedure offers high accuracy immediately, in real-time. The results need not be post-processed. In the earliest use of GPS kinematic and rapid static positioning were not frequently used because ambiguity resolution methods were still inefficient. Later when ambiguity resolution such as on-the-fly (OTF) became available, real-time kinematic and similar surveying methods became more widely used.

5 Coordinates

A FEW PERTINENT IDEAS ABOUT GEODETIC DATUMS FOR GPS

Plane Surveying

Plane surveying has traditionally relied on an imaginary flat reference surface, or *datum*, with Cartesian axes. This rectangular system is used to describe measured positions by ordered pairs, usually expressed in northings and eastings, or y- and x-coordinates. Even though surveyors have always known that this assumption of a flat earth is fundamentally unrealistic, it provided, and continues to provide, an adequate arrangement for small areas. The attachment of elevations to such horizontal coordinates somewhat acknowledges the topographic irregularity of the earth, but the whole system is always undone by its inherent inaccuracy as surveys grow large.

Development of State Plane Coordinate Systems

The purpose of the state plane coordinate system was to overcome some of the limitations of the horizontal plane datum when they are applied over large areas while avoiding the imposition of geodetic methods and calculations on local surveyors. Using the conic and cylindrical models of the Lambert and Mercator map projections, the flat datum was curved, but only in one direction. By curving the datums and limiting the area of the zones the distortion can be limited to a scale ratio of about 1 part in 10,000 without disturbing the traditional system of ordered pairs of Cartesian coordinates.

The state plane coordinate system was a step ahead at that time. To this day, it provides surveyors with a mechanism for coordination of surveying stations that approximates geodetic accuracy more closely than the commonly used methods of small-scale plane surveying. However, the state plane coordinate systems were organized in a time of generally lower accuracy and efficiency in surveying measurement. Its calculations were designed to avoid the lengthy and complicated mathematics of geodesy. It was an understandable compromise in an age when such computation required sharp pencils, logarithmic tables, and lots of midnight oil. These ideas will be addressed more completely later in this chapter.

The distortion of positions attributable to the transformation of geodetic coordinates into the plane grid coordinates of any one of these projections is generally small. Most GPS and land surveying software packages provide routines for automatic transformation of latitude and longitude to and from these mapping projections. Similar programs are available from the *National Geodetic Survey* (*NGS*).

Therefore, for most applications of GPS, there ought to be no technical compunction about expressing the results in grid coordinates. However, given the long traditions of plane surveying, it can be easy for some to lose sight of the geodetic context of the entire process that produced the final product of a GPS survey is presented in plane coordinates. The following is an effort to provide some context to that relationship.

GPS Surveyors and Geodesy

Today, GPS has thrust surveyors into the thick of geodesy, which is no longer the exclusive realm of distant experts. Thankfully, in the age of the microcomputer, the computational drudgery can be handled with software packages. Nevertheless, it is unwise to venture into GPS believing that knowledge of the basics of geodesy is, therefore, unnecessary. It is true that GPS would be impossible without computers, but blind reliance on the data they generate eventually leads to disaster.

SOME GEODETIC COORDINATE SYSTEMS

Three-Dimensional (3-D) Cartesian Coordinates

A spatial Cartesian system with three axes lends itself to describing the terrestrial positions derived from space-based geodesy. Using three rectangular coordinates instead of two, one can unambiguously define any position on the earth, or above it for that matter. The three-dimensional Cartesian coordinates (x,y,z) derived from this system are known as Earth-Centered-Earth-Fixed (ECEF) coordinates. It is a right-handed orthogonal system that rotates with and is attached to the earth.

A three-dimensional Cartesian coordinate system is right-handed if it can be described by the following model: the extended forefinger of the right hand symbolizes the positive direction of the x-axis. The middle finger of the same hand extended at right angles to the forefinger symbolizes the positive direction of the y-axis. The extended thumb of the right hand, perpendicular to them both, symbolizes the positive direction of the z-axis. But such a system is only useful if its origin (0,0,0) and its axes (x,y,z) can be fixed to the planet with certainty, something easier said than done (Figure 5.1).

The usual arrangement is known as the *Conventional Terrestrial Reference System* (*CTRS*), and the *Conventional Terrestrial System* (*CTS*). The latter name will be used here. The origin is the center of mass of the earth, the *geocenter*. The x-axis is a line from the geocenter through the intersection of the zero meridian with the equator toward 0° longitude, specifically, the intersection of the *World Geodetic System* (*WGS84*), Reference Meridian Plane, and the plane of the Conventional Terrestrial Pole Equator, more about that in a moment. The y-axis is extended from the geocenter along a line perpendicular from the x-axis in the same mean equatorial plane toward 90° East longitude. Finally, the z-axis is coincident with the earth's axis of rotation and there is a difficulty.

Polar Motion

The earth's rotational axis will not hold still. It actually wanders slightly with respect to the solid earth in a very slow oscillation called *polar motion*. The largest

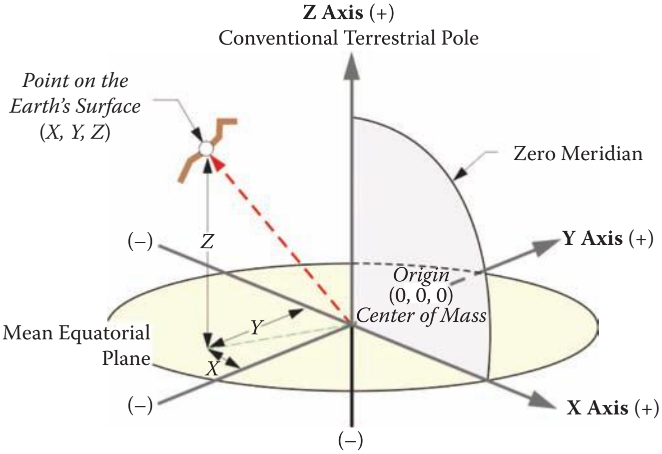


FIGURE 5.1 Three-Dimensional Coordinate (ECEF)

component of the movement relative to the earth's crust has a 430-day cycle known as the *Chandler period*. The actual displacement caused by the wandering generally does not exceed 12 meters. Nevertheless, the conventional terrestrial system of coordinates would be useless if its third axis was constantly wobbling. Therefore, an average stable position was chosen for the position of the pole and the *z*-axis.

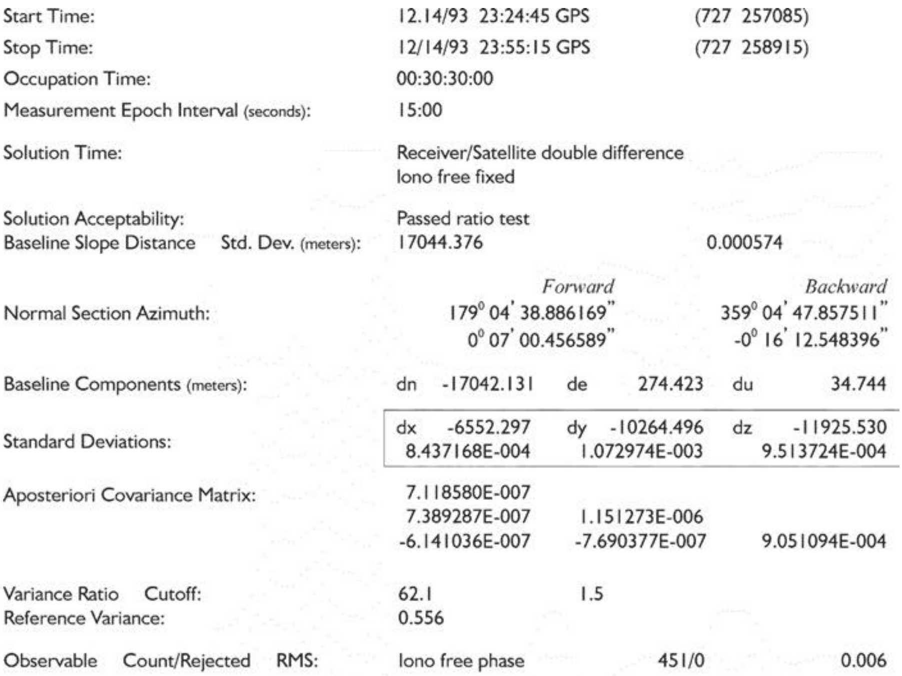
Between 1900 and 1905, the mean position of the earth's rotational pole was designated as the *Conventional International Origin* (*CIO*). This was originally defined by the *Bureau International de l'Heure* (*BIH*). It has, since 1989, been refined by the *International Earth Rotation Service* (*IERS*) using very long baseline interferometry (VLBI) and satellite laser ranging (SLR). The name of the *z*-axis has been changed to the CTP, epoch 1984, but its role has remained the same. Also the CTP provides a geometrically stable and clear definition on the earth's surface for the *z*-axis as regards WGS84. So, by international agreement, the *z*-axis of the *Conventional Terrestrial System* (*CTS*) is a line from the earth's center of mass through the CTP.

The Conventional Terrestrial System (CTS) in GPS

It is important to note that the Control Segment generates the position and velocity of the satellites themselves in ECEF coordinates. It follows that most modern GPS software provide the GPS positions in ECEF as well. Further, the ends of baselines determined by GPS observation are typically given in ECEF coordinates so that the vectors themselves become the difference between those *x, y,* and *z* coordinates. The display of these differences as *DX, DY,* and *DZ* is a usual product of these post-processed calculations (Figure 5.2).

Latitude and Longitude

Despite their utility, such 3-D Cartesian coordinates are not the most common method of expressing a geodetic position. Latitude and longitude have been the coordinates of choice for centuries. The application of these angular designations rely on



```
Start Time:                                  12.14/93  23:24:45 GPS          (727  257085)
Stop Time:                                   12/14/93  23:55:15 GPS          (727  258915)
Occupation Time:                             00:30:30:00
Measurement Epoch Interval (seconds):        15:00

Solution Time:                               Receiver/Satellite double difference
                                             Iono free fixed

Solution Acceptability:                      Passed ratio test
Baseline Slope Distance   Std. Dev. (meters): 17044.376                       0.000574

                                                  Forward                        Backward
Normal Section Azimuth:                        179° 04' 38.886169"       359° 04' 47.857511"
                                                 0° 07' 00.456589"        -0° 16' 12.548396"

Baseline Components (meters):                dn   -17042.131   de     274.423   du        34.744

                                             dx    -6552.297   dy  -10264.496   dz     -11925.530
Standard Deviations:                          8.437168E-004     1.072974E-003      9.513724E-004

Aposteriori Covariance Matrix:                7.118580E-007
                                              7.389287E-007     1.151273E-006
                                             -6.141036E-007    -7.690377E-007      9.051094E-004

Variance Ratio   Cutoff:                     62.1              1.5
Reference Variance:                          0.556

Observable   Count/Rejected   RMS:           Iono free phase             451/0             0.006
```

FIGURE 5.2 DX, DY, and DZ from GPS

the same two standard lines as 3-D Cartesian coordinates: the mean equator and the zero meridian. Unlike the CTS, they require some clear representation of the terrestrial surface. In modern practice, latitude and longitude cannot be said to uniquely define a position without a clear definition of the earth itself.

ELEMENTS OF A GEODETIC DATUM

How can latitude, φ, and longitude, λ, be considered inadequate in any way for the definition of a position on the earth? The reference lines—the mean equator and the zero meridian—are clearly defined. The units of degrees, minutes, seconds, and decimals of seconds, allow for the finest distinctions of measurement. Finally, the reference surface is the earth itself.

The answer to the question relies, in the first instance, on the fact that there are several categories of latitude and longitude, the geographical coordinates. From the various options, astronomic, geocentric and geodetic, the discussion here will concern geodetic latitude and longitude as they are typical. Its definition has a good deal to do with where down is.

The Deflection of the Vertical

Down seems like a pretty straightforward idea. A hanging plumb bob certainly points down. Its string follows the direction of gravity. That is one version of the idea. There are others.

Imagine an optical surveying instrument set up over a point. If it is centered precisely with a plumb bob and leveled carefully, the plumb line and the line of the level telescope of the instrument are perpendicular to each other. In other words, the level line, the horizon of the instrument, is perpendicular to gravity. Using an instrument so oriented it is possible to determine the latitude and longitude of the point. Measuring the altitude of a circumpolar star is one good method of finding the latitude. The measured altitude would be relative to the horizontal level line of the instrument. A latitude found this way is called *astronomic latitude*.

One might expect that this astronomic latitude would be the same as the geocentric latitude of the point, but they are different. The difference is due to the fact that a plumb line coincides with the direction of gravity; it does not point to the center of the earth where the line used to derive geocentric latitude originates.

Astronomic latitude also differs from the most widely used version of latitude, geodetic. The line from which geodetic latitude is determined is perpendicular with the surface of the ellipsoidal model of the earth that does not match a plumb line either, more about that in the next section. In other words, there are three different versions of down and each with its own latitude. For geocentric latitude, down is along a line to the center of the earth. For geodetic latitude down is along a line perpendicular to the ellipsoidal model of the earth. For astronomic latitude down is along a line in the direction of gravity. And more often than not these are three completely different lines.

GEOCENTRIC, GEODETIC, AND ASTRONOMIC LATITUDE

Each can be extended upward too, toward the zenith, and there are small angles between them. The angle between the vertical extension of a plumb line and the vertical extension of a line perpendicular to the ellipsoid is called the *deflection of the vertical*. It sounds better than the difference in down. This deflection of the vertical defines the actual angular difference between the astronomic latitude and longitude of a point and its geodetic latitude and longitude; latitude and longitude because, even though the discussion has so far been limited to latitude, the deflection of the vertical usually has both a north–south and an east–west component. The deflection of the vertical also has an effect on azimuths; for example, there will be a slight difference between the azimuth of a GPS baseline and the astronomically determined azimuth of the same line.

It is interesting to note that that optical instrument set up so carefully over a point on the earth cannot be used to measure geodetic latitude and longitude directly because they are not relative to the actual earth, but rather a model of it (see Figure 5.3). Gravity does not even come into the ellipsoidal version of down. On the model of the earth down is a line perpendicular to the ellipsoidal surface at a particular point. On the real earth down is the direction of gravity at the point. They are most often not the same thing, but fortunately the difference is usually very small. But since the ellipsoidal model of the earth is imaginary, it is quite impossible to actually set up an instrument on the ellipsoid. On the other hand, the measurement of latitude and longitude by astronomic observations has a very long history indeed. And yet the most commonly used coordinates are not astronomic latitudes and longitudes,

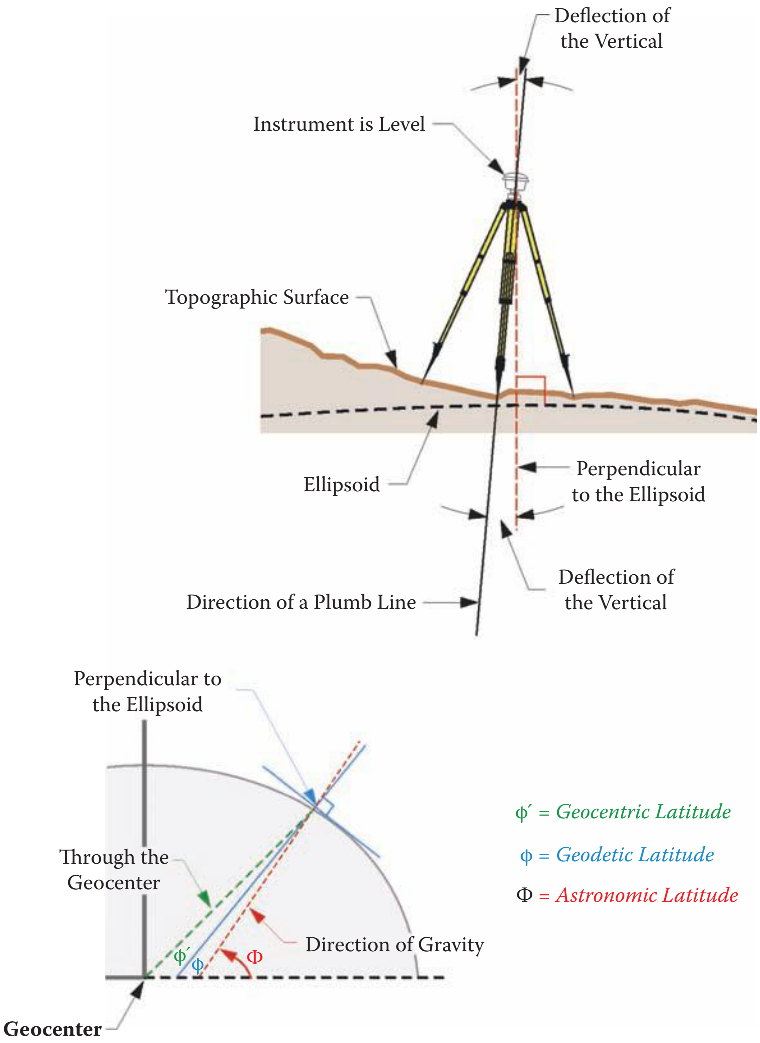


FIGURE 5.3 Deflection of the Vertical and Three Latitudes

but geodetic latitudes and longitudes. So conversion from astronomic latitude and longitude to geodetic latitude and longitude has a long history as well. Therefore, until the advent of GPS geodetic latitudes and longitudes were often values ultimately derived from astronomic observations by post-observation calculation. And in a sense that is still true, the change is a modern GPS receiver can display the geodetic latitude and longitude of a point to the user immediately because the calculations can be completed with incredible speed. But a fundamental fact remains unchanged: the instruments by which latitudes and longitudes are measured are oriented to gravity, the ellipsoidal model on which geodetic latitudes and longitudes are determined is not. And that is just as true for the antenna of a GPS receiver, an optical surveying instrument, a camera in an airplane taking aerial photography, or even the GPS satellite themselves.

Datums

The second part of the answer to the question posed earlier is this: if geographic coordinates are to have meaning they must have a context, a datum.

Despite the certainty of the physical surface of the earth, the lithosphere, it remains notoriously difficult to define in mathematical terms. The dilemma is illustrated by the ancient struggle to represent its curved surface on flat maps. There have been a whole variety of map projections developed over the centuries that rely on mathematical relationships between positions on the earth's surface and points on the map. Each projection serves a particular application well, but none of them can represent the earth without distortion. For example, no modern surveyor would presume to promise a client a high-precision control network with data scaled from a map. As the technology of measurement has improved, the pressure for greater exactness in the definition of the earth's shape has increased. Even with electronic tools that widen the scope and increase the precision of the data, perfection is nowhere in sight.

Development of the Ellipsoidal Model

Despite the fact that local topography is obvious to an observer standing on the earth, efforts to grasp the more general nature of the planet's shape and size have been occupying scientists for at least 2300 years. There have, of course, been long intervening periods of unmitigated nonsense on the subject. Ever since 200 B.C. when Eratosthenes almost calculated the planet's circumference correctly, geodesy has been getting ever closer to expressing the actual shape of the earth in numerical terms. A leap forward occurred with Newton's thesis that the earth was an ellipsoid rather than a sphere in the first edition of his *Principia* in 1687.

Newton's idea that the actual shape of the earth was slightly ellipsoidal was not entirely independent. There had already been some other suggestive observations. For example, 15 years earlier, astronomer J. Richter had found that to maintain the accuracy of the one-second clock he used in his observations in Cayenne, French Guiana, he had to shorten its pendulum significantly. The clock's pendulum, regulated in Paris, tended to swing more slowly as it approached the equator. Newton reasoned that the phenomenon was attributable to a lessening of the force of gravity. Based on his own theoretical work, he explained the weaker gravity by the proposition, "the earth is higher under the equator than at the poles, and that by an excess of about 17 miles" (*Philosophiae naturalis principia mathematica*, Book III, Proposition XX).

Although Newton's model of the planet bulging along the equator and flattened at the poles was supported by some of his contemporaries, notably Huygens, the inventor of Richter's clock, it was attacked by others. The director of the Paris Observatory, Jean Dominique Cassini, for example, took exception to Newton's concept. Even though the elder Cassini had himself observed the flattening of the poles of Jupiter in 1666, neither he nor his equally learned son Jacques were prepared to accept the same idea when it came to the shape of the earth. It appeared they had some empirical evidence on their side.

For geometric verification of the earth model, scientists had employed arc measurements at various latitudes since the early 1500s. Establishing the latitude of their beginning and ending points astronomically, they measured a cardinal line to discover the length of one degree of longitude along a meridian arc. Early attempts assumed a spherical earth, and the results were used to estimate its radius by simple multiplication. In fact, one of the most accurate of the measurements of this type, begun in 1669 by the French Abbé J. Picard, was actually used by Newton in formulating his own law of gravitation. However, Cassini noted that close analysis of Picard's arc measurement, and others, seemed to show the length of one degree of longitude actually *decreased* as it proceeded northward. He concluded that the earth was not flattened as proposed by Newton, but was rather elongated at the poles.

The argument was not resolved until two expeditions between about 1733 and 1744 were completed. They were sponsored by the Paris Académie Royale des Sciences and produced irrefutable proof. One group which included Clairaut and Maupertuis was sent to measure a meridian arc near the Arctic Circle, 66°20' Nφ, in Lapland. Another expedition with Bouguer and Godin, to what is now Ecuador, measured an arc near the equator, 01°31' Sφ. Newton's conjecture was proved correct, and the contradictory evidence of Picard's arc was charged to errors in the latter's measurement of the astronomic latitudes.

The ellipsoidal model (Figure 5.4), bulging at the equator and flattened at the poles, has been used ever since as a representation of the general shape of the earth's surface. It is called an oblate spheroid. In fact, several reference ellipsoids have been established for various regions of the planet. They are precisely defined by their semimajor axis and flattening. The relationship between these parameters are expressed in the formula:

$$f = \frac{a-b}{a}$$

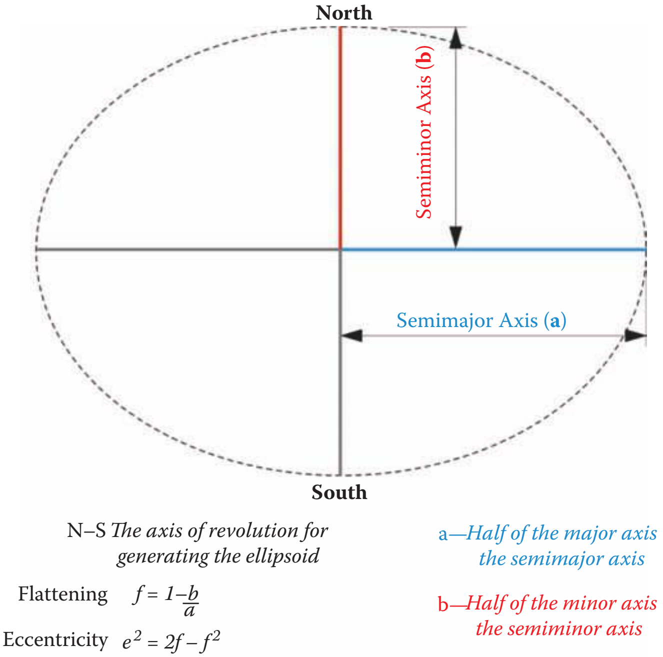
where

f = flattening
a = semimajor axis
b = semiminor axis.

BIAXIAL ELLIPSOIDAL MODEL OF THE EARTH

The Role of an Ellipsoid in a Datum

The semimajor and flattening can be used to completely define an ellipsoid of revolution. The ellipsoid is revolved around the minor access. However, six additional elements are required if that ellipsoid is to be used as a geodetic datum: three to specify its center and three more to clearly indicate its orientation around that center. The Clarke 1866 spheroid is one of many reference ellipsoids. Its shape is completely



Parameters of a Biaxial Ellipsoid

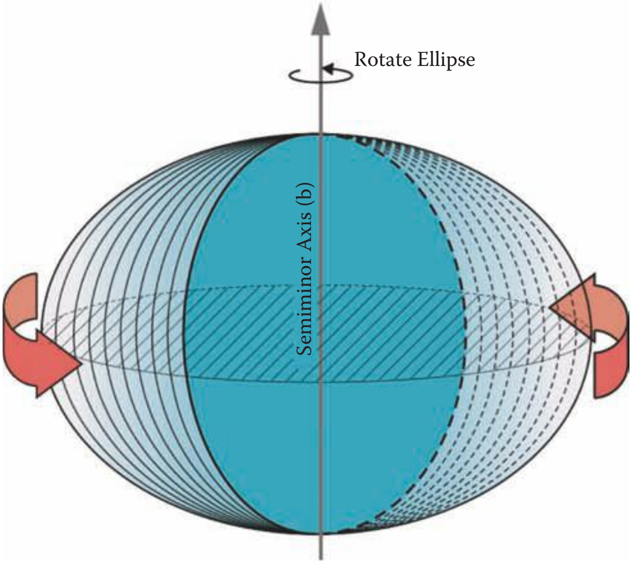


FIGURE 5.4 A Ellipsoid

defined by a semimajor axis, a, of 6378.2064 km and a flattening, f, of 1/294.9786982. It is the reference ellipsoid of the datum known to surveyors as the North American Datum of 1927 (NAD27), but it is not the datum itself.

For the Clarke 1866 spheroid to become NAD27, it had to be attached at a point and specifically oriented to the actual surface of the earth. However, even this ellipsoid, which fits North America best of all, could not conform to that surface perfectly. Therefore, the initial point was chosen near the center of the anticipated geodetic network to best distribute the inevitable distortion. The attachment was established at Meades Ranch, Kansas, 39°13'26".686 Nφ, 98°32'30".506 Wλ and *geoidal height* zero (we will discuss geoidal height in a later section). Those coordinates were not sufficient, however. The establishment of directions from this initial point was required to complete the orientation. The azimuth from Meades Ranch to station Waldo was fixed at 75°28'09".64 and the deflection of the vertical set at zero.

Once the initial point and directions were fixed, the whole orientation of NAD27 was established, including the center of the reference ellipsoid. Its center was imagined to reside somewhere around the center of mass of the earth. However, the two points were certainly not coincident, nor were they intended to be. In short, NAD27 does not use a geocentric ellipsoid.

REGIONAL ELLIPSOIDS

Measurement Technology and Datum Selection

In the period before space-based geodesy was tenable, a regional datum was not unusual. The Australian Geodetic Datum 1966, the European Datum 1950, and the South American Datum 1969, among others, were also designed as nongeocentric systems. Achievement of the minimum distortion over a particular region was the primary consideration in choosing their ellipsoids, not the relationship of their centers to the center of mass of the earth (Figure 5.5). For example, in the Conventional Terrestrial System (CTS) the 3-D Cartesian coordinates of the center of the Clarke 1866 spheroid as it was used for NAD27 are about $X = -4$ m, $Y = +166$ m, and $Z = +183$ m.

This approach to the design of datums was bolstered by the fact that the vast majority of geodetic measurements they would be expected to support were of the classical variety. That is, the work was done with theodolites, towers, and tapes. They were earth-bound. Even after the advent of electronic distance measurement, the general approach involved the determination of horizontal coordinates by measuring from point to point on the earth's surface and adding heights, otherwise known as *elevations*, through a separate leveling operation. As long as this methodological separation existed between the horizontal and vertical coordinates of a station, the difference between the ellipsoid and the true earth's surface was not an overriding concern. Such circumstances did not require a geocentric datum.

However, as the sophistication of satellite geodesy increased, the need for a truly global, geocentric datum became obvious. The horizontal and vertical information were no longer separate. Since satellites orbit around the center of mass of the earth, a position derived from space-based geodesy can be visualized as a vector originating from that point.

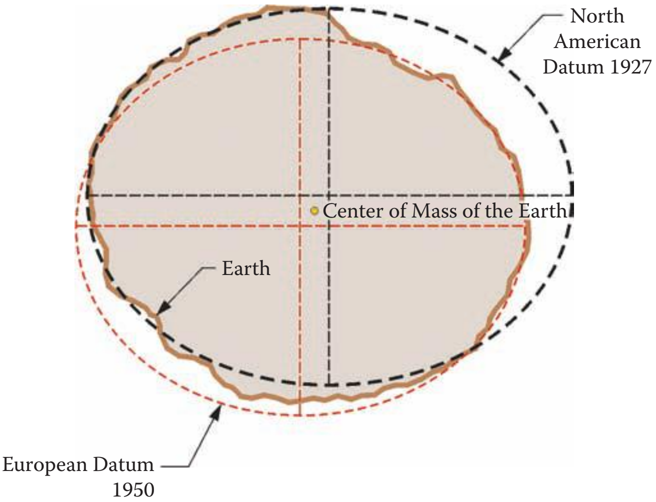


FIGURE 5.5 Nongeocentric Datums

So, today, not only are the horizontal and vertical components of a position derived from precisely the same vector, the choice of the coordinate system used to express them is actually a matter of convenience. The position vector can be transformed into the 3-D Cartesian ECEF system of CTS, the traditional latitude, longitude and height, or virtually any other well-defined coordinate system. However, since the orbital motion and the subsequent position vector derived from satellite geodesy are themselves earth-centered, it follows that the most straightforward representations of that data are earth-centered as well (Figure 5.6).

Position Derived from GPS

The Development of a Geocentric Model

Satellites have not only provided the impetus for a geocentric datum, they have also supplied the means to achieve it. In fact, the orbital perturbations of man-made near-earth satellites have probably brought more refinements to the understanding of the shape of the earth in a shorter span of time than was ever before possible. For example, the analysis of the precession of Sputnik 2 in the late 1950s showed researchers that the earth's semiminor axis was actually 85 meters shorter than had been previously thought. In 1958, while studying the tracking data from the orbit of Vanguard I, Ann Bailey of the Goddard Spaceflight Center discovered that the planet is shaped a bit like a pear. There is a slight protuberance at the North Pole, a little depression at the South Pole, and a small bulge just south of the equator.

These formations and others have been discovered through the observation of small distortions in satellites' otherwise elliptical orbits, little bumps in their road, so to speak. The deviations are caused by the action of Earth's gravity on the satellites as they travel through space. Just as Richter's clock reacted to the lessening of gravity at the equator and thereby revealed one of the largest features of the earth's

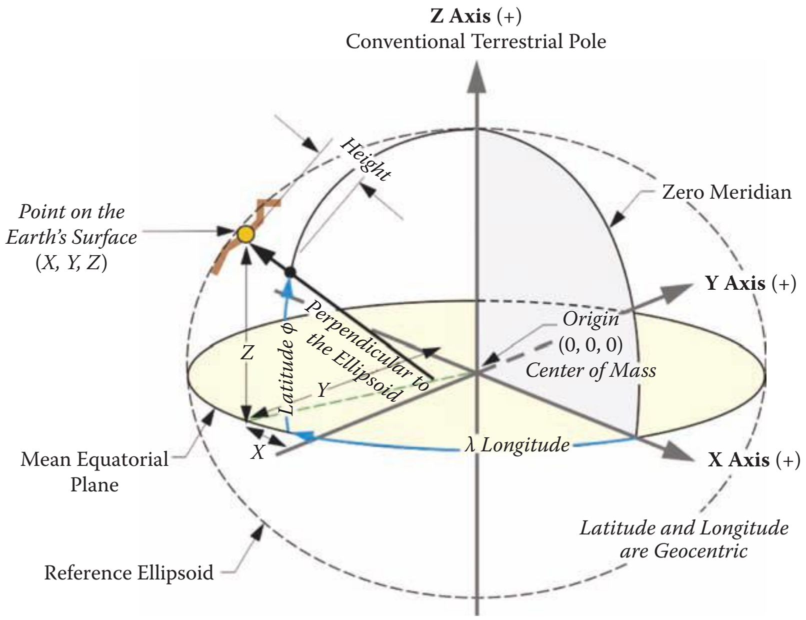


FIGURE 5.6 A Few Fundamentals

shape to Newton, small perturbations in the orbits of satellites, also responding to gravity, reveal details of Earth's shape to today's scientists. The common aspect of these examples is the direct relationship between direction and magnitude of gravity and the planet's form. In fact, the surface that best fits the earth's gravity field has been given a name. It is called the *geoid*.

THE GEOID

An often-used description of the geoidal surface involves idealized oceans. Imagine the oceans of the world utterly still, completely free of currents, tides, friction, variations in temperature and all other physical forces, except gravity. Reacting to gravity alone, these unattainable calm waters would coincide with the figure known as the geoid. Admitted by small frictionless channels or tubes and allowed to migrate across the land, the water would then, theoretically, define the same geoidal surface across the continents, too.

Of course, the 70% of the earth covered by oceans is not so cooperative, nor is there any such system of channels and tubes. In addition, the physical forces eliminated from the model cannot be avoided in reality. These unavoidable forces actually cause mean sea level to deviate up to several meters from the geoid. This is one of the reasons that Mean Sea Level and the surface of the geoid are not the same. And it is a fact frequently mentioned to emphasize the inconsistency of the original definition of the geoid as it was offered by J.B. Listing in 1872. Listing thought of the geoidal surface as equivalent to mean sea level. Even though his idea does not stand up to scrutiny today, it can still be instructive.

An Equipotential Surface

Gravity is not consistent across the topographic surface of the earth. At every point it has a magnitude and a direction. In other words, anywhere on the earth, gravity can be described by a mathematical vector. Along the solid earth, such vectors do not have all the *same* direction or magnitude, but one can imagine a surface of constant gravity potential. Such an *equipotential* surface would be *level* in the true sense. It would coincide with the top of the hypothetical water in the previous example. Despite the fact that mean sea level does not define such a figure, the geoidal surface is not just a product of imagination. For example, the vertical axis of any properly leveled surveying instrument and the string of any stable plumb bob are perpendicular to the geoid. Just as pendulum clocks and earth-orbiting satellites, they clearly show that the geoid is a reality.

Geoidal Undulation

The geoid does not precisely follow mean sea level, nor does it exactly correspond with the topography of the dry land. It is irregular like the terrestrial surface, and it has similar peaks and valleys. It is bumpy. Uneven distribution of the mass of the planet makes it maddeningly so. Maddening, because if the solid earth had no internal anomalies of density, the geoid would be smooth and almost exactly ellipsoidal. In that case, the reference ellipsoid could fit the geoid to near perfection and the lives of geodesists would be much simpler. But like the earth itself, the geoid defies such mathematical consistency and departs from true ellipsoidal form by as much as 100 meters in places (Figure 5.7).

THE MODERN GEOCENTRIC DATUM

Three distinct figures are involved in a geodetic datum for latitude, longitude, and height: the geoid, the reference ellipsoid, and the earth itself. Due in large measure to the ascendancy of satellite geodesy, it has become highly desirable that they share a common center.

While the level surface of the geoid provides a solid foundation for the definitions of heights (more about that later in this chapter) and the topographic surface of the earth is necessarily where measurements are made, neither can serve as the

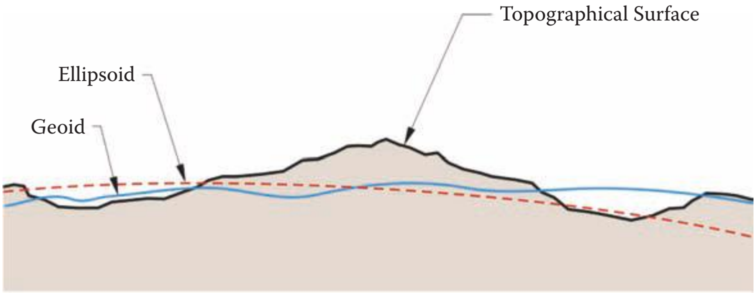


FIGURE 5.7 Three Surfaces

reference surface for geodetic positions. From the continents to the floors of the oceans, the solid earth's actual surface is too irregular to be represented by a simple mathematical statement. The geoid, which is sometimes under, and sometimes above, the surface of the earth, has an overall shape that also defies any concise geometrical definition. But the ellipsoid not only has the same general shape as the earth, but, unlike the other two figures, can be described simply and completely in mathematical terms.

Therefore, a global geocentric system has been developed based on the ellipsoid adopted by the *International Union of Geodesy and Geophysics* (*IUGG*) in 1979. It is called the *Geodetic Reference System 1980* (*GRS80*). Its semimajor axis, *a*, is 6378.137 km and is probably within a few of meters of the earth's actual equatorial radius. Its flattening, *f*, is 1/298.25722 and likely deviates only slightly from the true value, a considerable improvement over Newton's calculation of a flattening ratio of 1/230. But then he did not have orbital data from near-earth satellites to check his work.

World Geodetic System 1984 (WGS84)

However, GRS80 is not the reference ellipsoid for the GPS system that is the WGS84 ellipsoid. The WGS84 ellipsoid is actually very similar to the GRS80 ellipsoid. The difference between them is chiefly found in the flattening. The WGS84 ellipsoid is the foundation of the coordinate system, known as the *World Geodetic System 1984 (WGS84)*. This datum has been used by the U.S. Military since January 21, 1987. However, there have been four incarnations of WGS84 since then, the original and three updates.

While WGS84 has always been the basis for the GPS Navigation message computations, the particular version of the datum has changed. As of this writing the latest version of WGS84 is *WGS84 (G1150)*. There will be more updates in the future. It is worth noting that the number following the letter G is the number of the GPS week during which the coordinates first were used in the National Geospatial Intelligence Agency, NGA precise ephemeris estimations.

Therefore, coordinates provided today by GPS receivers are based in WGS84 (G1150), which is the same as ITRF00. However, most available GPS software can transform those coordinates to a number of other datums as well. The one that is probably of greatest interest to surveyors in the United States today is the *North American Datum 1983* (*NAD83*). Originally, the difference between WGS84 as originally rolled out in 1987 and NAD83 as first introduced in 1986 coordinates was so small that transformation was unnecessary. That is no longer the case when it comes to NAD83 (CORS96) and WGS84 (G1150); the difference can be up to 1 or 2 meters.

This is the third update to the realization of the WGS84 Reference Frame. The original WGS 84, which was based on observations from more than 1900 Doppler Stations, was revised to become WGS84 (G730) to incorporate GPS observations. It was implemented in GPS by the *Operational Control Segment, OCS* on June 29, 1994. More GPS based realizations of WGS84 followed, WGS84 (G873) on January 29, 1997, and WGS84 (G1150) was implemented by GPS *Operational Control Segment* January 20, 2002.

NORTH AMERICAN DATUM 1983

NAD27

The Clarke 1866 ellipsoid was the foundation of NAD27, and the blocks that built that foundation were made by geodetic triangulation. After all, an ellipsoid, even one with a clearly stated orientation to the earth, is only an abstraction until physical, identifiable control stations are available for its practical application. During the tenure of NAD27, control positions were tied together by ten of thousands of miles of triangulation and some traverses. Its measurements grew into chains of figures from Canada to Mexico and coast to coast, with their vertices perpetuated by bronze disks set in stone, concrete, and other permanent media.

These tri-stations, also known as brass caps, and their attached coordinates have provided a framework for all types of surveying and mapping projects for many years. They have served to locate international, state, and county boundaries. They have provided geodetic control for the planning of national and local projects, development of natural resources, national defense, and land management. They have enabled surveys to be fitted together, provided checks, and assisted in the perpetuation of their marks. They have supported scientific inquiry, including crustal monitoring studies and other geophysical research. But even as application of the nationwide control network grew, the revelations of local distortions in NAD27 were reaching unacceptable levels.

Judged by the standards of newer measurement technologies, the quality of some of the observations used in the datum were too low. That, and its lack of an internationally viable geocentric ellipsoid, finally drove its positions to obsolescence. The monuments remain, but it was clear early on that NAD27 had some difficulties. There were problems from too few base lines, Laplace azimuths and other deficiencies. By the early 1970s, the NAD27 coordinates of the national geodetic control network were no longer adequate.

The Development of the North American Datum 1983 (NAD83)

While a committee of the National Academy of Sciences advocated the need for a new adjustment in its 1971 report, work on the new datum, NAD83, did not really begin until after July 1, 1974. Leading the charge was an old agency with a new name. Called the *U.S. Coast & Geodetic Survey* in 1878, and then the *Coast and Geodetic Survey* (*C&GS*) from 1899, the agency is now known as the *National Geodetic Survey* (*NGS*). It is within the *National Oceanic and Atmospheric Administration* (*NOAA).* The first ancestor of today's NGS was established back in 1807 and was known as the *Survey of the Coast.* Its current authority is contained in U.S. Code, Title 33, USC 883a.

NAD83 includes not only the United States, but also Central America, Canada, Greenland, and Mexico. The NGS and the Geodetic Survey of Canada set about the task of attaching and orienting the GRS80 ellipsoid to the actual surface of the earth, as it was defined by the best positions available at the time. It took more than 10 years to readjust and redefine the horizontal coordinate system of North America into what is now NAD83. More than 1.7 million weighted observations derived from classical surveying techniques throughout the Western Hemisphere were involved in the least-squares adjustment. They were supplemented by approximately 30,000 EDM measured baselines, 5000 astronomic azimuths, and more than 650 Doppler stations positioned by the TRANSIT satellite system. Over 100 Very Long Baseline Interferometry (VLBI) vectors were also included. But GPS, in its infancy, contributed only five points.

GPS was growing up in the early 80's and some of the agencies involved in its development decided to join forces. NOAA, the *National Aeronautics and Space Administration* (*NASA*), the *U.S. Geological Survey* (*USGS*), and the Department of Defense coordinated their efforts. As a result each agency was assigned specific responsibilities. NGS was charged with the development of specifications for GPS operations, investigation of related technologies, and the use of GPS for modeling crustal motion. It was also authorized to conduct its subsequent geodetic control surveys with GPS. So, despite an initial sparseness of GPS data in the creation of NAD83, the stage was set for a systematic infusion of its positions as the datum matured. The work was officially completed on July 31, 1986.

The International Terrestrial Reference System (ITRS)

As NAD83 has aged there has been constant improvement in geodesy. When NAD83 was created it was intended to be geocentric. It is now known that NAD83 is about 2 meters from the true geocenter. Other difficulties have arisen, such as the limitations of the Doppler observations used in the definition of NAD83; GPS only contributed 5 positions to the original work.

The geocentric realization of ITRS and the best reference frame currently available is the *International Terrestrial Reference Frame* (*ITRF*). Its origin is at the center of mass of the whole earth including the oceans and atmosphere. The unit of length is the meter. The orientation of its axes was established as consistent with that of the IERS's predecessor, Bureau International de l'Heure, BIH, at the beginning of 1984.

Today, the ITRF is maintained by the International Earth Rotation Service (IERS), which monitors Earth Orientation Parameters (EOP) for the scientific community through a global network of observing stations. This is done with GPS, Very Long Baseline Interferometry (VLBI), Lunar Laser Ranging (*LLR*), Satellite Laser Ranging (SLR), the Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS), and the positions of the observing stations are now considered to be accurate to the centimeter level. Recognizing that the several hundred control stations worldwide for which it publishes yearly coordinates are actually in motion due to the shifting of approximately 20 tectonic plates worldwide, the IERS also provides velocities for them.

The International Terrestrial Reference Frame is actually a series of realizations. The first was in 1988. In other words, it is revised and published on a regular basis. And today NAD83 can be defined in terms of a best-fit transformation from ITRF96.

ITRF00, WGS84, and NAD83

The North American Datum of 1983 (NAD83) is used everywhere in North America except Mexico. The latest version as of this writing is *NAD83* (*CORS96*). This datum is realized in the coterminous United States and Alaska through the National *CORS* (*Continuously Operating Reference Stations*). There are hundreds of National CORS and Cooperative CORS sites participating in the network, and this number continuously grows with the addition of several new stations each month.

As mentioned earlier, in the not too distant past we did not have to be concerned with the shift between NAD83 (1986) and WGS84 as introduced in 1987, because the discrepancy easily fell within our overall error budget. NAD83 and WGS84, originally differed by only a centimeter or two. That is no longer true. In their new definitions—NAD83 (CORS96) and WGS84 (G1150)—differ up to one or two meters within the continental United States. Further, ITRF00 and WGS84 (G1150) are virtually identical, and, therefore, the same divergence applies when ITRF00 is compared against NAD83 (CORS96). NGS has developed a software package called Horizontal Time Dependent Positioning (HTDP) to transform ITRF positions of a given epoch to NAD 83 (CORS96) and vice versa, more about this in a moment.

It is important to note that the current ITRF/WGS84 system is global. It is in constant motion due to the shifting of tectonic plates around the world. However, the current NAD83 system is fixed to one plate, the North American plate, and moves with it. Consequently, NAD83, in the continental United States moves approximately 10 to 20 millimeters per year in relation to the ITRF/WGS84 system.

The Management of NAD83

Since geodetic accuracy with GPS depends on relative positioning, surveyors continue to rely on NGS stations to control their work just as they have for generations. Today, it is not unusual for surveyors to find that some NGS stations have published coordinates in NAD83 and others, perhaps needed to control the same project, only have positions in NAD27. In such a situation, it is often desirable to transform the NAD27 positions into coordinates of the newer datum. But, unfortunately, there is no single-step mathematical approach that can do it accurately.

The distortions between the original NAD27 positions are part of the difficulty. The older coordinates were sometimes in error as much as 1 part in 15,000. Problems stemming from the deflection of the vertical, lack of correction for geoidal undulations, low-quality measurements, and other sources contributed to inaccuracies in some NAD27 coordinates that cannot be corrected by simply transforming them into another datum.

Transformations from NAD27 to NAD83

Nevertheless, various approximate methods are used to transform NAD27 coordinates into supposed NAD83 values. For example, the computation of a constant local translation is sometimes attempted using stations with coordinates in both systems as a guide. Another technique is the calculation of two translations, one rotation and one scale parameter, for particular locations based on the latitudes and longitudes of three or more common stations. Perhaps the best results derive from polynomial expressions developed for coordinate differences, expressed in Cartesian ($\Delta x, \Delta y, \Delta z$) or ellipsoidal coordinates ($\Delta \varphi, \Delta \lambda, \Delta h$), using a 3-D Helmert transformation. However, besides requiring seven parameters (three shift, one scale, and three rotation components) this approach is at its best when ellipsoidal heights are available for all the points involved. Where adequate information is available, software packages such as the NGS programs LEFTI or NADCON can provide geodetic quality coordinates.

Even if a local transformation is modeled with these techniques, the resulting NAD27 positions might still be plagued with relatively low accuracy. The NAD83 adjustment of the national network is based on nearly 10 times the number of observations that supported the NAD27 system. This larger quantity of data, combined with the generally higher quality of the measurements at the foundation of NAD83, can have some rather unexpected results. For example, when NAD27 coordinates are transformed into the new system, the shift of individual stations may be quite different from what the regional trend indicates. In short, when using control from both NAD83 and NAD27 simultaneously on the same project, surveyors have come to expect difficulty.

In fact, the only truly reliable method of transformation is not to rely on coordinates at all, but to return to the original observations themselves. It is important to remember, for example, that geodetic latitude and longitude, as other coordinates, are specifically referenced to a given datum and are not derived from some sort of absolute framework. But the original measurements, incorporated into a properly designed least-squares adjustment, can provide most satisfactory results.

Densification and Improvement of NAD83

The inadequacies of NAD27 and even NAD83 positions in some regions are growing pains of a fundamentally changed relationship. In the past, relatively few engineers and surveyors were employed in geodetic work. Perhaps the greatest importance of the data from the various geodetic surveys was that they furnished precise points of reference to which the multitude of surveys of lower precision could then be tied. This arrangement was clearly illustrated by the design of state plane coordinates systems, devised to make the national control network accessible to surveyors without geodetic capability.

However, the situation has changed. The gulf between the precision of local surveys and national geodetic work is virtually closed by GPS, and that has changed the relationship between local surveyors in private practice and geodesists. For example, the significance of state plane coordinates as a bridge between the two groups has been drastically reduced. Today's surveyor has relatively easy and direct access to the geodetic coordinate systems themselves through GPS. In fact, the 1- to 2-ppm probable error in networks of relative GPS-derived positions frequently exceed the accuracy of the first-order NAD83 positions intended to control them.

Fortunately, GPS surveyors have a chance to contribute to the solution of these difficulties. NGS will accept GPS survey data submitted in the correct format with proper supporting documentation. The process, known as *blue-booking*, requires strict adherence to NGS specifications. GPS measurements that can meet the criteria are processed, classified, and incorporated into the NGRS for the benefit all GPS surveyors.

High Accuracy Reference Networks

Other significant work along this line was accomplished in the state-by-state supernet programs. The creation of *High Accuracy Reference Networks* (*HARN*) were cooperative ventures between NGS and the states, and often include other organizations as well. The campaign was originally known as *High Precision Geodetic Networks* (*HPGN*).

A station spacing of not more than about 62 miles and not less than about 16 miles was the objective in these statewide networks. The accuracy was intended to be 1 part-per-million, or better between stations. In other words, with heavy reliance on GPS observations, these networks were intended to provide extremely accurate, vehicle-accessible, regularly spaced control point monuments with good overhead visibility. These stations were intended to provide control superior to the vectors derived from the day-to-day GPS observations that are tied to them. In that way the HARN points provide the user with a means to avoid any need to warp vectors to fit inferior control. That used to sometime happen in the early days of GPS. To further ensure such coherence in the HARN, when the GPS measurements were complete, they were submitted to NGS for inclusion in a statewide readjustment of the existing NGRS covered by the state. Coordinate shifts of 0.3 to 1.0 m from NAD83 values were typical in these readjustments, which were concluded in 1998.

The most important aspect of HARN positions is the accuracy of their final positions. Entirely new orders of accuracy were developed for GPS relative positioning techniques by the *Federal Geodetic Control Committee* (*FGCC*). The FGCC is an organization chartered in 1968 and composed of representatives from 11 agencies of the federal government. It revises and updates surveying standards for geodetic control networks, among other duties. Its 1989 provisional standards and specifications for GPS work include Orders AA, A, and B, which are defined as having minimum geometric accuracies of 3mm ± 0.01 ppm, 5mm ± 0.1 ppm and 8mm ± 1 ppm, respectively, at the 95%, or 2σ, confidence level. The adjusted positions of HARN stations are designed to provide statewide coverage of at least B Order control as set out in these new standards. The orthometric heights are determined by differential leveling from existing control stations.

The publication of up-to-date geodetic data, always one of the most important functions of NGS, is even more crucial today. The original NAD83 adjustment is indicated with a suffix including the year 1986 in parentheses, that is, NAD83 (1986). However, when a newer HARN adjusted value is available the year in the parentheses will be the year of the adjustment, that is, NAD83 (1994). The format of the data published by NGS has also changed somewhat. NAD83 information includes, of course, the new ellipsoidal latitudes, longitudes, and azimuths. However, unlike the NAD27 data that provided the elevations of only some control points, NAD83 data includes elevations, or heights, for all marked stations. The orthometric heights in the HARN are determined by differential leveling from existing control stations.

Height Modernization and Base Networks

In 1998 following a study mandated by Congress another related NGS campaign began with the help of the states and private organizations. The effort to provide more highly accurate vertical control points, more benchmarks, using GPS was undertaken. It is known as the *National Height Modernization Program, HTMOD*. In the future the NGS intends to replace the state-by-state approach to this work with a more national approach.

From 1998 to 2004 NGS introduced another series of observations in each state designed to tie the network to the *Continuously Operating Reference Stations, CORS*. This work resulted in the Federal Base Network (FBN), which is a nationwide network of monumented stations. These spatial reference positions are among the most precise available and are particularly dense in crustal motion areas. In general these points are spaced at approximately 100 kilometers apart. The accuracies intended are: 1 cm-latitudes and longitudes, 2 cm-ellipsoidal heights, and 3 cm-orthometric heights. These stations are few compared to the much more numerous *Cooperative Base Network* (*CBN*). This is a high-accuracy network of monumented control stations spaced at 25 to 50 km apart throughout the United States and its territories. The CBN was created and is maintained by state and private organizations with the help of NGS. The FBN surveys were completed in 2003 and will support a nationwide coordinate readjustment. FBN and CBN are high accuracy surveys that tie the HARN to CORS.

Continuously Operating Reference Stations

In about 1992 the NGS began establishing a network of Continuous Operating Reference Stations (CORS) throughout the country. The original idea was to provide positioning for navigational and marine needs. There were about 50 CORS in 1996. Their positional accuracies are 3 cm horizontal and 5 cm vertical. They also must meet NOAA geodetic standards for installation, operation, and data distribution. In 1999 NGS brought an additional 33 Continuously Operating Reference Stations (CORS) online, resulting in a total of 165 stations in the National CORS Network.

The Continuously Operating Reference Stations in the NGS network are mostly to provide support for carrier phase observations. There are also many CORS available for differential correction of the code observations of DGPS. Tracking information is available for postprocessing on the Internet.

STATE PLANE COORDINATES

NAD83 POSITIONS AND PLANE COORDINATES

The NGS published data for stations also includes state plane coordinates in the appropriate zone. As before, the easting and northing are accompanied by the mapping angle and grid azimuths, but a scale factor is also included for easy conversions. *Universal Transverse Mercator* (*UTM*) coordinates are among the new elements offered by NGS in the published information for NAD83 stations.

These plane coordinates, both state plane and UTM, are far from an anachronism. The UTM projection has been adopted by the IUGG, the same organization that reached the international agreement to use GRS80 as the reference ellipsoid for the modern geocentric datum. The U.S. National Aeronautics and Space Administration (NASA) and other military and civilian organizations worldwide also use UTM coordinates for various mapping needs. UTM coordinates are often useful to those planning work that embraces large areas. In the United States, state plane systems based on the transverse Mercator projection, an oblique Mercator projection, and the Lambert conic map projection, grid every state, Puerto Rico, and the U.S.

Virgin Islands into their own plane rectangular coordinate system. And GPS surveys performed for local projects and mapping are frequently reported in the plane coordinates of one of these systems.

State Plane Coordinates rely on an imaginary flat reference surface with Cartesian axes. They describe measured positions by ordered pairs, expressed in northings and eastings, or x- and y-coordinates. Despite the fact that the assumption of a flat Earth is fundamentally wrong, calculation of areas, angles and lengths using latitude and longitude can be complicated, so plane coordinates persist. Therefore, the projection of points from the earth's surface onto a reference ellipsoid and finally onto flat maps is still viable.

In fact, many agencies of government, particularly those that administer state, county, and municipal databases prefer coordinates in their particular *State Plane Coordinate System* (*SPCS*). The systems are, as the name implies, state specific. In many states the system is officially sanctioned by legislation. Generally speaking, such legislation allows surveyors to use State Plane Coordinates to legally describe property corners. It is convenient.

MAP PROJECTION

State Plane Coordinate Systems are built on *map projections*. Map projection means representing a portion of the actual Earth on a plane. Done for hundreds of years to create paper maps, it continues, but map projection today is most often really a mathematical procedure done in a computer. Nevertheless, even in an electronic world it cannot be done without distortion.

The problem is often illustrated by trying to flatten part of an orange peel. The orange peel stands in for the surface of the earth. A small part, say a square a quarter of an inch on the side, can be pushed flat without much noticeable deformation. But when the portion gets larger problems appear. Suppose a third of the orange peel is involved, as the center is pushed down the edges tear and stretch, or both. And if the peel gets even bigger the tearing gets more severe. So if a map is drawn on the orange before it is peeled, the map gets distorted in unpredictable ways when it is flattened. And it is difficult to relate a point on one torn piece with a point on another in any meaningful way.

These are the problems that a map projection needs to solve to be useful. The first problem is the surface of an ellipsoid, like the orange peel, is *nondevelopable*. In other words, flattening it inevitably leads to distortion. So, a useful map projection ought to start with a surface that is *developable,* a surface that may be flattened without all that unpredictable deformation. It happens that a paper cone or cylinder both illustrate this idea nicely. They are illustrations only, models for thinking about the issues involved.

If a right circular cone is cut up and one of its elements is perpendicular from the base to its apex, the cone can then be made completely flat without trouble. The same may be said of a cylinder cut up a perpendicular from base to base (see Figure 5.8).

Or one could use the simplest case, a surface that is already developed, a flat piece of paper. If the center of a flat plane is brought tangent to the earth, a portion of the planet can be mapped on it, that is it can be projected directly onto the

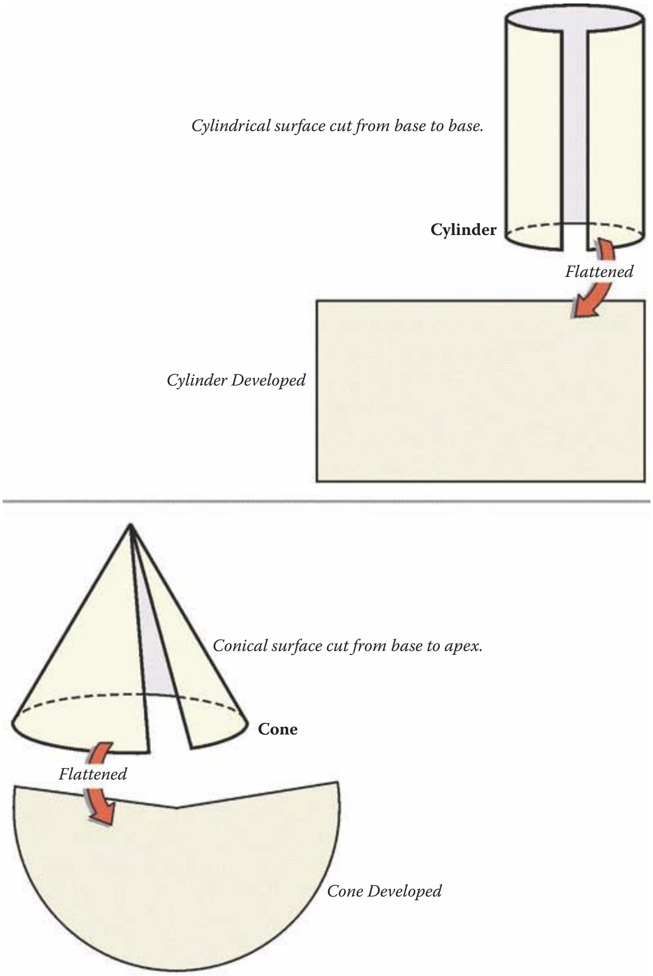


FIGURE 5.8 The Development of a Cylinder and a Cone

flat plane. In fact this is the typical method for establishing an independent *local coordinate system.* These simple Cartesian systems are convenient and satisfy the needs of small projects. The method of projection, onto a simple flat plane, is based on the idea that a small section of the earth, as with a small section of the orange mentioned previously, conforms so nearly to a plane that distortion on such a system is negligible.

Subsequently, local tangent planes have been long used by land surveyors. Such systems demand little if any manipulation of the field observations and the approach has merit as long as the extent of the work is small. But the larger the plane grows the more untenable it becomes. As the area being mapped grows, the reduction of survey observations becomes more complicated since it must take account of the actual shape of the earth. This usually involves the ellipsoid, the geoid, and geographical coordinates, latitude and longitude. At that point surveyors and engineers rely on

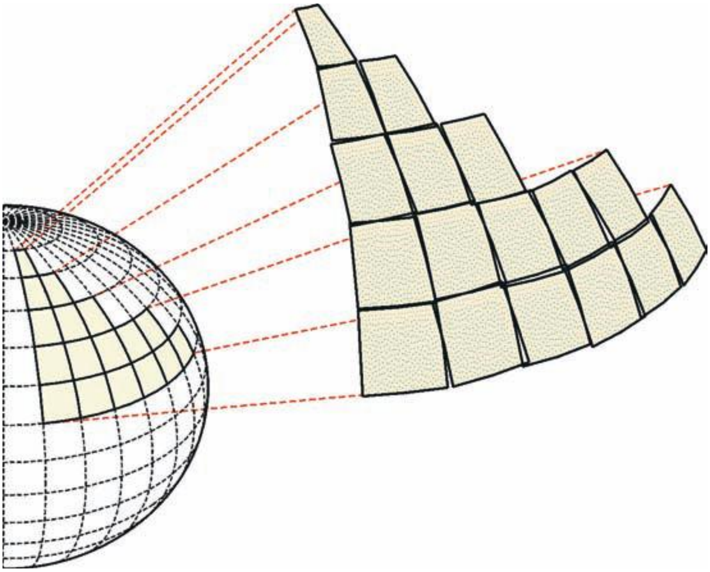


FIGURE 5.9 Local Coordinate Systems Do Not Edge Match

map projections to mitigate the situation and limit the now troublesome distortion. However, a well-designed map projection can offer the convenience of working in plane Cartesian coordinates and still keep the inevitable distortion at manageable levels at the same time.

DISTORTION

The design of such a projection must accommodate some awkward facts. For example, while it would be possible to imagine mapping a considerable portion of the earth using a large number of small individual planes, like facets of a gem, it is seldom done because when these planes are brought together they cannot be edge-matched accurately (see Figure 5.9). They cannot be joined properly along their borders. And the problem is unavoidable because the planes, tangent at their centers, inevitably depart more and more from the reference ellipsoid at their edges, and the greater the distance between the ellipsoidal surface and the surface of the map on which it is represented, the greater the distortion on the resulting flat map. This is true of all methods of map projection. Therefore, one is faced with the daunting task of joining together a mosaic of individual maps along their edges where the accuracy of the representation is at its worst. And even if one could overcome the problem by making the distortion, however large, the same on two adjoining maps, another difficulty would remain. Typically each of these planes has a unique coordinate system. The orientation of the axes, the scale, and the rotation of each one of these individual local systems will not be the same as those elements of its neighbor's coordinate system. Subsequently there are gaps and overlaps between adjacent maps, and their attendant coordinate systems, because there is no common reference system.

So the idea of a self-consistent, local map projection based on small, flat planes tangent to the earth, or the reference ellipsoid, is convenient, but only for small projects that have no need to be related to adjoining work. And as long as there is no need to venture outside the bounds of a particular local system it can be entirely adequate. But, generally speaking, if a significant area needs coverage another strategy is needed.

Decreasing Distortion

Decreasing distortion is a constant and elusive goal in map projection. It can be done in several ways. Most involve reducing the distance between the map projection surface and the ellipsoidal surface. One way this is done is to move the mapping surface from tangency with the ellipsoid and make it actually cut through it. This strategy produces what is know as a *secant* projection. A secant projection is one way to shrink the distance between the map projection surface and the ellipsoid. Thereby the area where distortion is in an acceptable range on the map can be effectively increased (Figure 5.10).

The distortion can be reduced even further when one of those developable surfaces mentioned earlier is added to the idea of a secant map projection plane (Figure 5.11).

SECANT AND CYLINDRICAL PROJECTIONS

Both cones and cylinders have an advantage over a flat map projection plane. They are curved in one direction and can be designed to follow the curvature of the area to be mapped in that direction. Also, if a large portion of the ellipsoid is to be mapped several cones or several cylinders may be used together in the same system to further limit distortion. In that case, each cone or cylinder defines a *zone* in a larger coverage. This is the approach used in State Plane Coordinate systems.

Secant Projections

As mentioned, when a conic or a cylindrical map projection surface is made secant, it intersects the ellipsoid, and the map is brought close to its surface. For example, the conic and cylindrical projections shown in Figure 5.11 cut through the ellipsoid. The map is projected both inward and outward onto it. And two *lines of exact scale*, standard lines, are created along the *small circles* where the cone and the cylinder intersect the ellipsoid. They are called small circles because they do not describe a plane that goes through the center of the earth as do the previously mentioned great circles.

Where the ellipsoid and the map projection surface touch, in this case intersect, there is no distortion. However, between the standard lines the map is under the ellipsoid and outside of them the map is above it. That means that between the standard lines a distance from one point to another is actually longer on the ellipsoid than it is shown on the map, and outside the standard lines a distance on the ellipsoid is shorter than it is on the map. Any length that is measured along a standard line is the same on the ellipsoid and on the map, which is why another name for standard parallels is *lines of exact scale*.

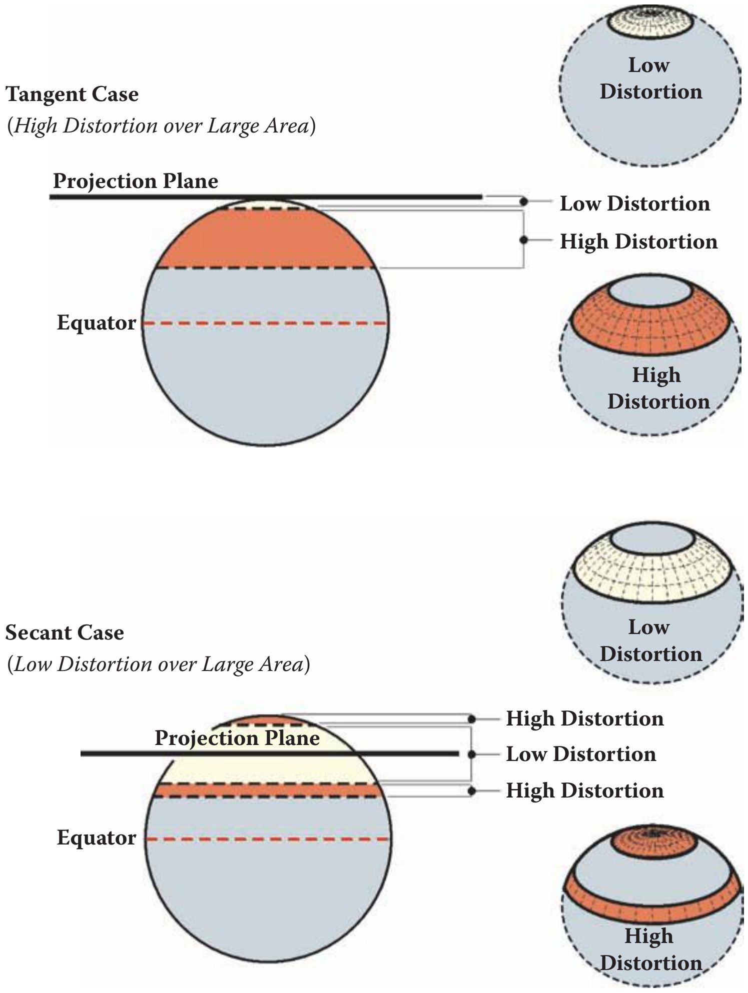


FIGURE 5.10 Limiting Distortion

Choices

Ultimately, the goal is very straightforward relating each position on one surface, the reference ellipsoid, to a corresponding position on another surface as faithfully as possible and then flattening that second surface to accommodate Cartesian coordinates. In fact, the whole procedure is in the service of moving from geographic to Cartesian coordinates and back again. These days the complexities of the mathematics are handled with computers. Of course, that was not always the case.

In the 1932, two engineers in North Carolina's highway department, O.B. Bester and George F. Syme, appealed to the then Coast & Geodetic Survey (C&GS, now NGS) for help. They had found that the stretching and compression inevitable in the representation of the curved earth on a plane was so severe over long route surveys that they could not check into the C&GS geodetic control stations across a

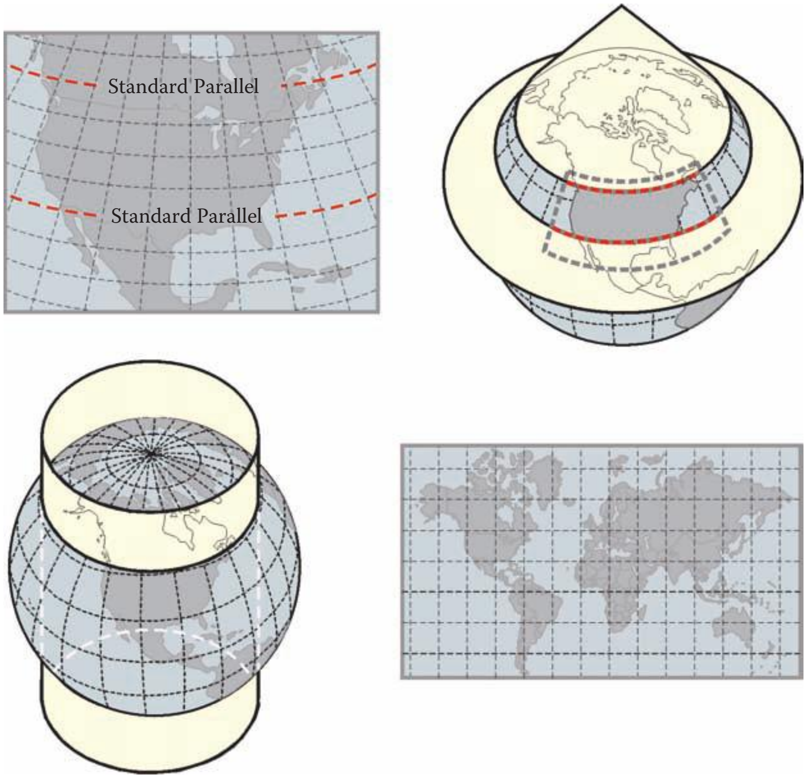


FIGURE 5.11 Two Projections

state within reasonable limits. The engineers suggested that a plane coordinate grid system be developed that was mathematically related to the reference ellipsoid, but could be utilized using plane trigonometry.

Dr. Oscar Adams of the Division of Geodesy, assisted by Charles Claire, designed the first *State Plane Coordinate System* to mediate the problem. It was based on a map projection called the *Lambert Conformal Conic Projection*. Adams realized that it was possible to use this map projection and allow one of the four elements of area, shape, scale, or direction to remain virtually unchanged from its actual value on the earth, but not all four. On a perfect map projection all distances, directions, and areas could be conserved. They would be the same on the ellipsoid and on the map. Unfortunately, it is not possible to satisfy all of these specifications simultaneously, at least not completely. There are inevitable choices. It must be decided which characteristic will be shown the most correctly, but it will be done at the expense of the others. And there is no universal best decision. Still a solution that gives the most satisfactory results for a particular mapping problem is always available.

Adams chose the Lambert Conformal Conic Projection for the North Carolina system. On the Lambert Conformal Conic Projection parallels of latitude are arcs of concentric circles and meridians of longitude are equally spaced straight radial lines, and the meridians and parallels intersect at right angles. The axis of the cone is imagined to be a prolongation of the polar axis. The parallels are not equally spaced because the scale varies as you move north and south along a meridian of longitude. Adams decided to use this map projection in which shape is preserved based on a developable cone.

Map projections in which shape is preserved are known as *conformal* or *orthomorphic*. Orthomorphic means right shape. In a conformal projection the angles between intersecting lines and curves retain their original form on the map. In other words, between short lines, meaning lines under about 10 miles, a 45° angle on the ellipsoid is a 45° angle on the map. It also means that the scale is the same in all directions from a point; in fact, it is this characteristic that preserves the angles. These aspects were certainly a boon for the North Carolina Highway engineers and benefits that all State Plane Coordinate users have enjoyed since. On long lines, angles on the ellipsoid are not exactly the same on the map projection. Nevertheless, the change is small and systematic.

Actually, all three of the projections that were used in the designs of the original State Plane Coordinate Systems were conformal. Each system was based on the North American Datum 1927, NAD27. Along with the Oblique Mercator projection, which was used on the panhandle of Alaska, the two primary projections were the Lambert Conic Conformal Projection and the Transverse Mercator projection. For North Carolina, and other states that are longest east-west, the Lambert Conic projection works best. State Plane Coordinate systems in states that are longest north-south were built on the Transverse Mercator projection. There are exceptions to this general rule. For example, California uses the Lambert Conic projection even though the state could be covered with fewer Transverse Mercator zones. The Lambert Conic projection is a bit simpler to use, which may account for the choice.

The Transverse Mercator projection is based on a cylindrical mapping surface much like that illustrated in Figure 5.11. However, the axis of the cylinder is rotated so that it is perpendicular with the polar axis of the ellipsoid. Unlike the Lambert Conic projection, the Transverse Mercator represents meridians of longitude as curves rather than straight lines on the developed grid. The Transverse Mercator projection is not the same thing as the Universal Transverse Mercator system (UTM). UTM was originally a military system that covers the entire earth and differs significantly from the Transverse Mercator system used in State Plane Coordinates.

In using these projections as the foundation of the State Plane Coordinate systems Adams wanted to have the advantage of conformality and also cover each state with as few *zones* as possible. A zone in this context is a belt across the state that has one Cartesian coordinate grid with one origin and is projected onto one mapping surface. One strategy that played a significant role in achieving that end was Adams's use of *secant* projections in both the Lambert and Transverse Mercator systems.

STATE PLANE COORDINATE SYSTEM (SPCS) MAP PROJECTIONS

For example, using a single secant cone in the Lambert projection and limiting the extent of a zone, or belt, across a state to about 158 miles, approximately 254 km, Adams limited the distortion of the length of lines. Not only were angles preserved on the final product, but also there were no radical differences between the length of a measured line on the earth's surface and the length of the same line on the map projection. In other words, the scale of the distortion was pretty small.

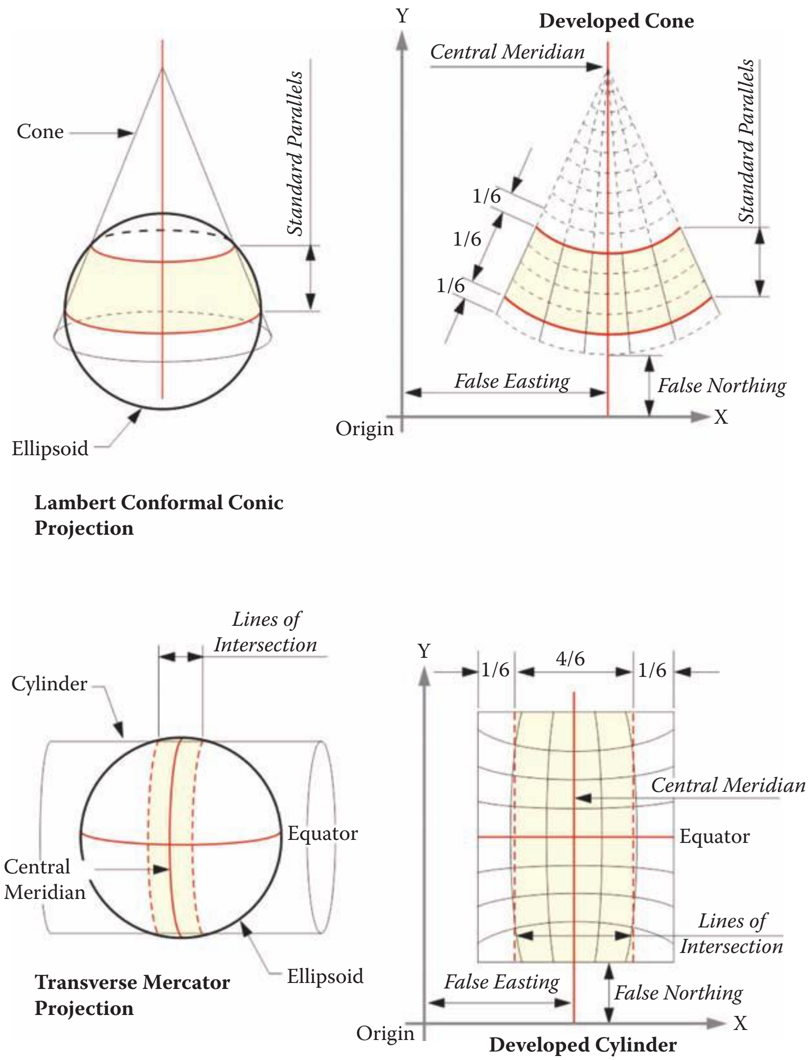
As illustrated in Figure 5.12, he placed four-sixths of the map projection plane between the standard lines, one-sixth outside at each extremity. The distortion was held to 1 part in 10,000. A maximum distortion in the lengths of lines of 1 part in 10,000 means that the difference between the length of a 2-mile line on the ellipsoid and its representation on the map would only be about 1 foot at the most.

State Plane Coordinates were created to be the basis of a method that approximates geodetic accuracy more closely than the then commonly used methods of small-scale plane surveying. Today surveying methods can easily achieve accuracies beyond 1 part in 100,000 and better, but the State Plane Coordinate systems were designed in a time of generally lower accuracy and efficiency in surveying measurement. Today computers easily handle the lengthy and complicated mathematics of geodesy. But the first State Plane Coordinate System was created when such computation required sharp pencils and logarithmic tables. In fact, the original SPCS was so successful in North Carolina similar systems were devised for all the states in the Union within a year or so. The system was successful because, among other things, it overcame some of the limitations of mapping on a horizontal plane while avoiding the imposition of strict geodetic methods and calculations. It managed to keep the distortion of the scale ratio under 1 part in 10,000 and preserved conformality. It did not disturb the familiar system of ordered pairs of Cartesian coordinates and it covered each state with as few zones as possible whose boundaries were constructed to follow county lines. County lines were used so that those relying on State Plane Coordinates could work in one zone throughout a jurisdiction.

SPCS27 TO SPCS83

In several instances the boundaries of State Plane Coordinate Zones today, *SPCS83,* the State Plane Coordinate System based on NAD83 and its reference ellipsoid GRS80, differ from the original zone boundaries. The foundation of the original State Plane Coordinate System, *SPCS27* was NAD27 and its reference ellipsoid Clarke 1866. As mentioned earlier NAD27 geographical coordinates, latitudes and longitudes, differ significantly from those in NAD83. In fact, conversion from geographic coordinates, latitude and longitude, to grid coordinates, x and y, and back is one of the three fundamental conversions in the State Plane Coordinate system. It is important because the whole objective of the SPCS is to allow the user to work in plane coordinates, but still have the option of expressing any of the points under consideration in either latitude and longitude or State Plane Coordinates without significant loss of accuracy. Therefore, when geodetic control was migrated from NAD27 to NAD83, the State Plane Coordinate System had to go along.

When the migration was undertaken in the 1970s, it presented an opportunity for an overhaul of the system. Many options were considered but in the end just a few changes were made. One of the reasons for the conservative approach was the fact that 37 states had passed legislation supporting the use of State Plane Coordinates. Nevertheless some zones got new numbers and some of the zones changed. The



Lambert Conformal Conic Projection

Developed Cone

Transverse Mercator Projection

Developed Cylinder

FIGURE 5.12 State Plane Coordinate Systems

zones are numbered in the SPCS83 system known as *FIPS*. FIPS stands for Federal Information Processing Standard, and each SPCS83 zone has been given a FIPS number. These days the zones are often known as *FIPS zones*. SPCS27 zones did not have these FIPS numbers. As mentioned earlier the original goal was to keep each zone small enough to ensure that the scale distortion was 1 part in 10,000 or less, but when the SPCS83 was designed that scale was not maintained in some states.

Changes in Zones

In five states some SPCS27 zones were eliminated altogether and the areas they had covered consolidation into one zone or added to adjoining zones. In three of those states the result was one single large zone. Those states are South Carolina, Montana, and Nebraska. In SPCS27 South Carolina and Nebraska had two zones, in SPCS83 they have just one, FIPS zone 3900 and FIPS zone 2600, respectively. Montana previously had three zones. It now has one, FIPS zone 2500. Therefore, because the area covered by these single zones has become so large they are not limited by the 1 part in 10,000 standard. California eliminated zone 7 and added that area to FIPS zone 0405, formerly zone 5. Two zones previously covered Puerto Rico and the Virgin Islands. They now have one. It is FIPS zone 5200. In Michigan, three Transverse Mercator zones were entirely eliminated.

False northing and false easting. In both the Transverse Mercator and the Lambert projection the positions of the axes are similar in all SPCS zones. As you can see in Figure 5.12 each zone has a central meridian. These central meridians are true meridians of longitude near the geometric center of the zone. Please note that the central meridian is not the *y*-axis. If it were the *y*-axis negative coordinates would result. To avoid them the actual *y*-axis is moved far to the west of the zone itself. In the old SPCS27 arrangement the *y*-axis was 2,000,000 feet west from the central meridian in the Lambert Conic projection and 500,000 feet in the Transverse Mercator projection. In the SPCS83 design those constants have been changed. The most common values are 600,000 meters for the Lambert Conic and 200,000 meters for the Transverse Mercator. However, there is a good deal of variation in these numbers from state to state and zone to zone. In all cases, however, the y-axis is still far to the west of the zone and there are no negative State Plane Coordinates. No negative coordinates, because the x-axis, also known as the baseline, is far to the south of the zone. Where the *x*-axis and *y*-axis intersect is the origin of the zone, and that is always south and west of the zone itself. This configuration of the axes ensures that all State Plane Coordinates occur in the first quadrant and are, therefore, always positive.

There is sometimes even further detail in the name of particular State Plane Coordinates. As refinements are made to NAD83 the new adjustments are added as a suffix to the SPCS83 label. For example, SPCS83/99 would refer to State Plane Coordinates that were based on a revision to NAD83 from 1999.

It is important to note that the fundamental unit for SPCS27 is the U.S. survey foot and for SPCS it is the meter. The conversion from meters to U.S. survey feet is correctly accomplished by multiplying the measurement in meters by the fraction 3937/1200.

STATE PLANE COORDINATES SCALE AND DISTANCE

Geodetic Lengths to Grid Lengths

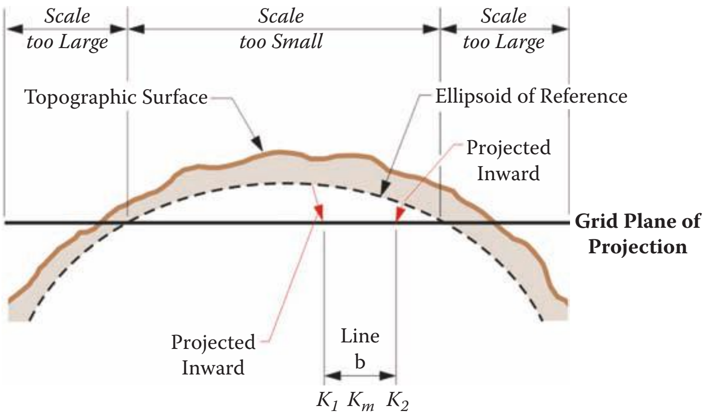
This brings us to the scale factor, also known as the K factor and the projection factor. It was this factor that the original design of the State Plane Coordinate system sought to limit to 1 part in 10,000. As implied by that effort scale factors are ratios that can be used as multipliers to convert ellipsoidal lengths, also known as geodetic distances, to lengths on the map projection surface, also known as grid distances, and vice versa. In other words, the geodetic length of a line, on the ellipsoid, multiplied by the appropriate scale factor will give you the grid length of that line on the map. And the grid length multiplied by the inverse of that same scale factor would bring you back to the geodetic length again.

While referring to Figure 5.12 it is interesting to note that on the projection used most on states that are longest from east to west, that is the Lambert Conic, the scale factor for east-west lines is constant. In other words, the scale factor is the same all along the line. One way to think about this is to recall that the distance between the ellipsoid and the map projection surface never changes east to west in that projection. On the other hand along a north-south line the scale factor is constantly changing on the Lambert Conic. And it is no surprise then to see that the distance between the ellipsoid and the map projection surface is always changing north to south line in that projection. But looking at the Transverse Mercator projection, the projection used most on states longest north to south, the situation is exactly reversed. In that case, the scale factor is the same all along a north-south line, and changes constantly along an east-west line.

Both the Transverse Mercator and the Lambert Conic used a secant projection surface and restricted the width to 158 miles. These were two strategies used to limit scale factors when the State Plane Coordinate systems were designed. Where that was not optimum, the width was sometimes made smaller, which means the distortion was lessened. As the belt of the ellipsoid projected onto the map narrows, the distortion gets smaller. For example, Connecticut is less than 80 miles wide north to south. It has only one zone. Along its northern and southern boundaries, outside of the standard parallels, the scale factor is 1 part in 40,000, a fourfold improvement over 1 part in 10,000. And in the middle of the state the scale factor is 1 part in 79,000, nearly an eightfold increase. On the other hand, the scale factor was allowed to get a little bit smaller that 1 part in 10,000 in Texas. By doing that the state was covered completely with five zones. And among the guiding principles in 1933 was covering the states with as few zones as possible and having zone boundaries follow county lines. Still it requires ten zones and all three projections to cover Alaska.

In Figure 5.13 a typical State Plane Coordinate zone is represented by a grid plane of projection cutting through the ellipsoid of reference. As mentioned earlier between the intersections of the standard lines, the grid is under the ellipsoid. There, a distance from one point to another is longer on the ellipsoid than on the grid. This means that right in the middle of a SPCS zone the scale factor is at its minimum. In the middle a typical minimum SPCS scale factor is not less than 0.9999, though there are exceptions. Outside of the intersections the grid is above the ellipsoid where a distance from one point to another is shorter on the ellipsoid than it is on the grid. There at the edge of the zone a maximum typical SPCS scale factor is generally not more than 1.0001.

When SPCS 27 was current, scale factors were interpolated from tables published for each state. In the tables for states in which the Lambert Conic projection was used scale factors change north–south with the changes in latitude. In the tables for states in which the Transverse Mercator projection was used scale factors change east-west with the changes in x-coordinate. Today scale factors are not interpolated from tables for SPCS83. For both the Transverse Mercator and the Lambert Conic projections they are calculated directly from equations.



$$(Approximate)\ K = \frac{K_1 + K_2}{2}$$

$$(Precise)\ K = \frac{K_1 + K_2 + 4K_m}{6}$$

FIGURE 5.13 The Scale Factor

There are several software applications that can be used to automatically calculate scale factors for particular stations. Perhaps the most convenient is that available free online at http://www.ngs.noaa.gov/PC_PROD/pc_prod.shtml. This is a link to the U.S. National Geodetic Survey (NGS) page where one can locate several programs that provide geodetic computational help. The program *SPCS83* is available for download there. It can be used to convert latitudes and longitudes to State Plane Coordinates. Given the latitude and longitude of the stations under consideration part of the available output from the program are the scale factors for those stations. Scale factors for control stations are also available from NGS geodetic control datasheets.

To illustrate the use of these factors consider line b to have a length on the ellipsoid of 130,210.44 feet, a bit over 24 miles. That would be its geodetic distance. Suppose that the scale factor for that line was 0.9999536, then the grid distance along line b would be:

$$Geodetic\ Distance * Scale\ Factor = Grid\ Distance$$

$$130{,}210.44\,ft. * 0.999953617 = 130{,}204.40\,ft.$$

The difference between the longer geodetic distance and the shorter grid distance here is a little more that 6 feet. That is actually better than 1 part in 20,000; please recall that the 1 part in 10,000 ratio was considered the maximum. Distortion lessens and the scale factor approaches 1 as a line nears a standard parallel.

Please also recall that on the Lambert projection an east–west line, that is a line that follows a parallel of latitude, has the same scale factor at both ends and throughout. However, a line that bears in any other direction will have a different scale factor at each end. A north–south line will have a great difference in the scale factor at its north end compared with the scale factor of its south end. In this vein, note the approximate formula near the bottom of Figure 5.13:

$$K = \frac{K_1 + K_2}{2}$$

Where K is the scale factor for a line, K_1 is the scale factor at one end of the line and K_2 is the scale factor at the other end of the line. Scale factor varies with the latitude in the Lambert projection. For example, suppose the point at the north end of the 24-mile line is called Stormy and has a geographic coordinate of:

37°46'00. 7225"
103°46'35.3195"

and at the south end the point is known as Seven with a geographic coordinate of:

37°30'43.5867"
104°05'26.5420"

The scale factor for point Seven is 0.99996113 and the scale factor for point Stormy is 0.99994609. It happens that point Seven is further south and closer to the standard parallel than is point Stormy, and it therefore follows that the scale factor at Seven is closer to 1. It would be exactly 1 if it were on the standard parallel, which is why the standard parallels are called *lines of exact scale*. The typical scale factor for the line is the average of the scale factors at the two end points:

$$K = \frac{K_1 + K_2}{2}$$

$$0.99995361 = \frac{0.99996113 + 0.99994609}{2}$$

Deriving the scale factor at each end and averaging them is the usual method for calculating the scale factor of a line. The average of the two is sometimes called K_m. In Figure 5.13 is another formula for calculating a more precise scale factor by using a weighted average. In this method K_1 is given a weight of 1, K_m a weight of 4, and K_2 a weight of 1. No matter which method is used the result is still called the scale factor.

But that is not the whole story when it comes to reducing distance to the State Plane Coordinate grid. Measurement of lines must always be done on the topographic surface of the earth, and not on the ellipsoid. Therefore the first step in deriving a grid distance must be moving a measured line from the earth to the ellipsoid. In other words, converting a distance measured on the topographic surface to a geodetic distance on the reference ellipsoid. This is done with another ratio that is also used as a multiplier. Originally, this factor had a rather unfortunate name. It used to be known as the *sea level factor* in SPCS27. It was given that name because as you may recall from Chapters 2 and 3 when NAD27 was established using the Clarke 1866 reference ellipsoid the distance between the ellipsoid and the geoid was declared to be zero at Meades Ranch in Kansas. That meant that in the middle of the country the "sea level" surface, the geoid, and the ellipsoid were coincident by definition. And since the Clarke 1866 ellipsoid fit the United States quite well the separation between the two surfaces, the ellipsoid and geoid, only grew to about 12 meters anywhere in the country. With such a small distance between them many practitioners at the time took the point of view that, for all practical purposes, the ellipsoid and the geoid were in the same place. And that place was called "sea level." Hence reducing a distance measure on the surface of the earth to the ellipsoid was said to be reducing it to "sea level."

Today that idea and that name for the factor is misleading because, of course, the GRS80 ellipsoid on which NAD83 is based is certainly not the same as Mean Sea Level. Now the separation between the geoid and ellipsoid can grow as large as 53 meters. And technology by which lines are measured has improved dramatically. Therefore, in SPCS83 the factor for reducing a measured distance to the ellipsoid is known as the *ellipsoid factor.* In any case, both the old and the new name can be covered under the name the *elevation factor.*

Regardless of the name applied to the factor, it is a ratio. The ratio is the relationship between an approximation of the earth's radius and that same approximation with the mean ellipsoidal height of the measured line added to it. For example, consider station Boulder and station Peak illustrated in Figure 5.14.

Boulder
  39°59'29.1299"
  105°15'39.6758"
Peak
  40°01'19.1582"
  105°30'55.1283"

The distance between these two stations is 72,126.21 feet. This distance is sometimes called the *ground distance,* or the *horizontal distance at mean elevation.* In other words it is not the slope distance but rather the distance between them corrected to an averaged horizontal plane, as is common practice. For practical purposes then this is the distance between the two stations on the topographic surface of the earth. On the way to finding the grid distance Boulder to Peak there is the interim step, calculating the geodetic distance between them, that is the distance on the ellipsoid. We need the elevation factor and here is how it is determined.

The ellipsoidal height of Boulder, $h_{1,}$ is 5437 feet. The ellipsoidal height of Peak, h_2, is 9099 feet. The approximate radius of the earth, traditionally used in this work, is 20,906,000 feet. The elevation factor is calculated:

$$Elevation\ Factor = \frac{R}{R + h_{avg}}$$

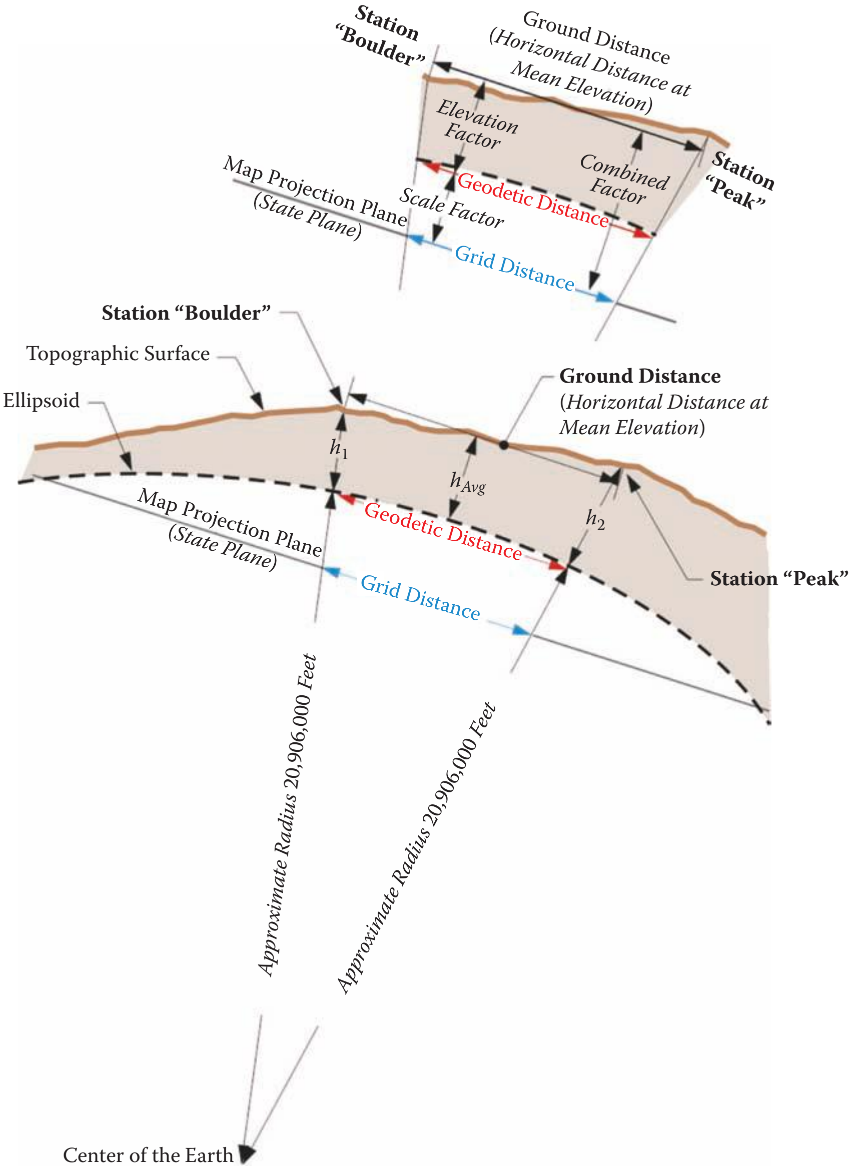


FIGURE 5.14 An Example—Distances

$$Elevation\ Factor = \frac{20,906,000\,ft.}{20,906,000\,ft. + 7268\,ft.}$$

$$Elevation\ Factor = \frac{20,906,000\,ft.}{20,913,268\,ft.}$$

$$Elevation\ Factor = 0.99965247$$

This factor then is the ratio used to move the ground distance down to the ellipsoid, down to the geodetic distance.

$$Ground\ Distance\ Boulder\ to\ Peak = 72{,}126.21\,ft.$$

$$Geodetic\ Distance = Ground\ Distance * Elevation\ Factor$$

$$Geodetic\ Distance = 72{,}126.21 * 0.99965247$$

$$Geodetic\ Distance = 72{,}101.14\,ft.$$

It is possible to refine the calculation of the elevation factor by using an average of the actual radial distances from the center of the ellipsoid to the end points of the line, rather than the approximate 20,906,000 feet. In the area of stations Boulder and Peak the average ellipsoidal radius is actually a bit longer, but it is worth noting that within the continental United States such variation will not cause a calculated geodetic distance to differ significantly. However, it is worthwhile to take care to use the ellipsoidal heights of the stations when calculating the elevation factor, rather than the orthometric heights.

In calculating the elevation factor in SPCS27 no real distinction is made between ellipsoid height and orthometric height. However, in SPCS83 the averages of the ellipsoidal heights at each end of the line are needed for having. If the ellipsoid height is not directly available it can be calculated from the formula

$$h = H + N$$

where:

h = ellipsoid height
H = orthometric height
N = geoid height

As mentioned previously converting a geodetic distance to a grid distance is done with an averaged scale factor:

$$K = \frac{K_1 + K_2}{2}$$

In this instance the scale factor at Boulder is 0.99996703 and at Peak it is 0.99996477.

$$0.99996590 = \frac{0.99996703 \ + 0.99996477}{2}$$

Using the scale factor it is possible to reduce the geodetic distance 72,101.14 feet to a grid distance:

$$Geodetic\ Distance * Scale\ Factor = Grid\ Distance$$

$$72{,}101.14\mathit{ft.} * 0.99996590 = 72{,}098.68\mathit{ft.}$$

There are two steps, first from ground distance to geodetic distance using the elevation factor and second from geodetic distance to grid distance using the scale factor, that can be combined into one. Multiplying the elevation factor and the scale factor produces a single ratio that is usually known as the *combined factor* or the *grid factor*. Using this grid factor the measured line is converted from a ground distance to a grid distance in one jump. Here is how it works. In the example above the elevation factor for the line from Boulder to Peak is 0.99965247 and the scale factor is 0.99996590:

$$Grid\ Factor = Scale\ Factor * Elevation\ Factor$$

$$0.99961838 = 0.99996590 * 0.99965247$$

Then using the grid factor the ground distance is converted to a grid distance.

$$Grid\ Distance = Grid\ Factor * Ground\ Distance$$

$$72{,}098.68\mathit{ft.} = 0.99961838 * 72{,}126.21\mathit{ft.}$$

Also the grid factor can be used to go the other way. If the grid distance is divided by the grid factor it is converted to a ground distance.

$$Ground\ Distance = \frac{Grid\ Distance}{Grid\ Factor}$$

$$72{,}126.21\mathit{ft.} = \frac{72{,}098.68\mathit{ft.}}{0.99961838}$$

UNIVERSAL TRANSVERSE MERCATOR COORDINATES

A plane coordinate system that is convenient for GIS work over large areas is the Universal Transverse Mercator (UTM) system. UTM with the Universal Polar Stereographic system covers the world in one consistent system. It is four times less accurate than typical State Plane Coordinate systems with a scale factor that typically reaches 0.9996. Yet the ease of using UTM and its worldwide coverage makes it very attractive for work that would otherwise have to cross many different SPCS zones. For example nearly all National Geospatial-Intelligence Agency (NGA) topographic maps, U.S. Geological Survey (USGS) quad sheets, and many aeronautical charts show the UTM grid lines.

It is often said that UTM is a military system created by the U.S. Army, but several nations, and the North Atlantic Treaty Organization (NATO), played roles in its creation after World War II. At that time the goal was to design a consistent coordinate system that could promote cooperation between the military organizations of several nations. Before the introduction of UTM allies found that their differing systems hindered the synchronization of military operations.

Conferences were held on the subject from 1945 to 1951 with representatives from Belgium, Portugal, France, and Britain, and the outlines of the present UTM system were developed. By 1951 the U.S. Army introduced a system that was very similar to that currently used.

The UTM projection divides the world into 60 zones that begin at longitude 180º, the International Date Line. Zone 1 is from 180° to the 174° W longitude. The coterminous United States are within UTM zones 10 to 19.

Here is a convenient way to find the zone number for a particular longitude. Consider west longitude negative and east longitude positive, add 180° and divide by 6. Any answer greater than an integer is rounded to the next highest integer and you have the zone. For example, Denver, Colorado is near 105° W. Longitude, –105°.

$$-105° + 180° = 75°$$
$$75\,°/\,6 = 12.50$$
$$\text{Round up to } 13$$

Therefore, Denver, Colorado, is in UTM Zone 13 as shown in Figure 5.15.

All UTM zones have a width of 6° of longitude. From north to south the zones extend from 84° N latitude to 80° S latitude. Originally the northern limit was at 80° N latitude and the southern 80° S latitude. On the south the latitude is a small circle that conveniently traverses the ocean well south of Africa, Australia and South America. However, 80° N latitude was found to exclude parts of Russia and Greenland and was extended to 84° N latitude.

UNIVERSAL TRANSVERSE MERCATOR (UTM) ZONES OF THE WORLD

These zones nearly cover the earth, except the Polar Regions, which are covered by two azimuthal polar zones called the Universal Polar Stereographic (UPS) projection. The foundation of the 60 UTM zones is a secant Transverse Mercator projection very similar to those used in some State Plane Coordinate systems. The central

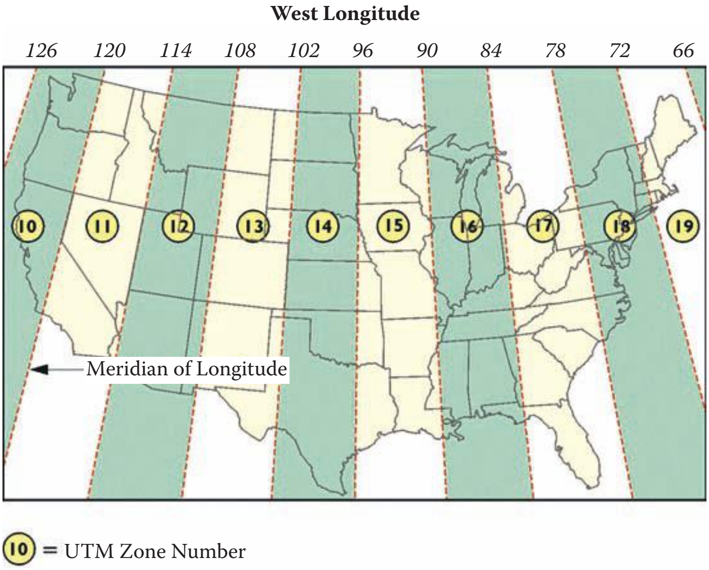


FIGURE 5.15 UTM Zones in Coterminuous United States

meridian of the zones is exactly in the middle. For example, in Zone 1 from 180° W to the 174° W longitude the central meridian is 177° W longitude, so each zone extends 3 degrees east and west from its central meridian.

The UTM secant projection gives approximately 180 kilometers between the lines of exact scale where the cylinder intersects the ellipsoid. The scale factor grows from 0.9996 along the central meridian of a UTM zone to 1.00000 at 180 km to the east and west. Please recall that SPCS zones are usually more limited in width, ~158 miles, and, therefore, have a smaller range of scale factors than do the UTM zones. In state plane coordinates, the scale factor is usually no more than 1 part in 10,000. In UTM coordinates it can be as large as 1 part in 2500.

The reference ellipsoids for UTM coordinates vary among five different figures, but in the United States it is the Clarke 1866 ellipsoid. However, one can obtain 1983 UTM coordinates by referencing the UTM zone constants to the GRS80 ellipsoid of NAD83, then 1983 UTM coordinates will be obtained.

Unlike any of the systems previously discussed every coordinate in a UTM zone occurs twice, once in the Northern Hemisphere and once in the Southern Hemisphere. This is a consequence of the fact that there are two origins in each UTM zone. The origin for the portion of the zone north of the equator is moved 500 km west of the intersection of the zone's central meridian and the equator. This arrangement ensures that all of the coordinates for that zone in the Northern Hemisphere will be positive. The origin for the coordinates in the Southern Hemisphere for the same zone is 500 km west of the central meridian as well. But in the Southern Hemisphere the origin is not on the equator, it is 10,000 km south of it, close to the South Pole. This orientation of the origin guarantees that all of the coordinates in the Southern Hemisphere are in the first quadrant and are positive. In other words, the intersection of each zone's central meridian with the equator defines its origin of coordinates. In the southern hemisphere, each origin is given the coordinates:

$$\text{easting} = X_0 = 500,000 \text{ meters, and northing} = Y_0 = 10{,}000{,}000 \text{ meters}$$

In the northern hemisphere, the values are:

$$\text{easting} = X_0 = 500,000 \text{ meters, and northing} = Y_0 = 0 \text{ meters, at the origin}$$

In fact, in the official version of the UTM system there are actually more divisions in each UTM zone than the north-south demarcation at the equator. As shown in Figure 5.16 each zone is divided into 20 subzones. Each of the subzones covers 8° of latitude and is lettered from C on the south to X on the north. Actually, subzone X is a bit longer than 8°; remember the extension of the system from 80° N latitude to 84° N latitude. That all went into subzone X. It is also interesting that I and O are not included. They resemble one and zero too closely. In any case these subzones are little used outside of military applications of the system.

The developed UTM grid is defined in meters. Each zone is projected onto the cylinder that is oriented in the same way as that used in the Transverse Mercator SPCS described earlier. And the radius of the cylinder is chosen to keep the scale errors within acceptable limits (see Figure 5.17).

A word or two about the polar zones that round out the UTM system. The UPS are azimuthal stereographic projections like those mentioned earlier. The projection has two zones. The North zone covers latitudes 84° N to 90° N. The South zone

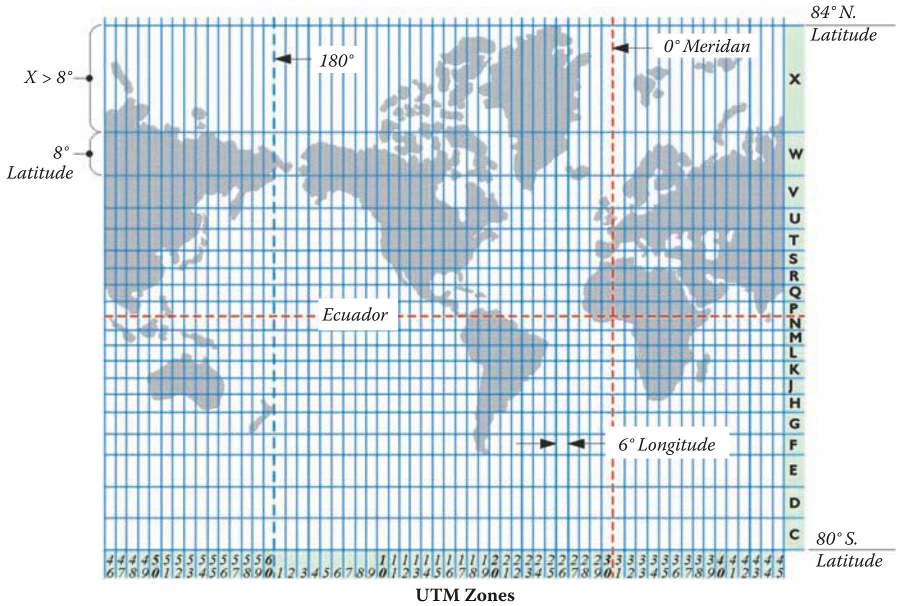


FIGURE 5.16 UTM Zones around the World

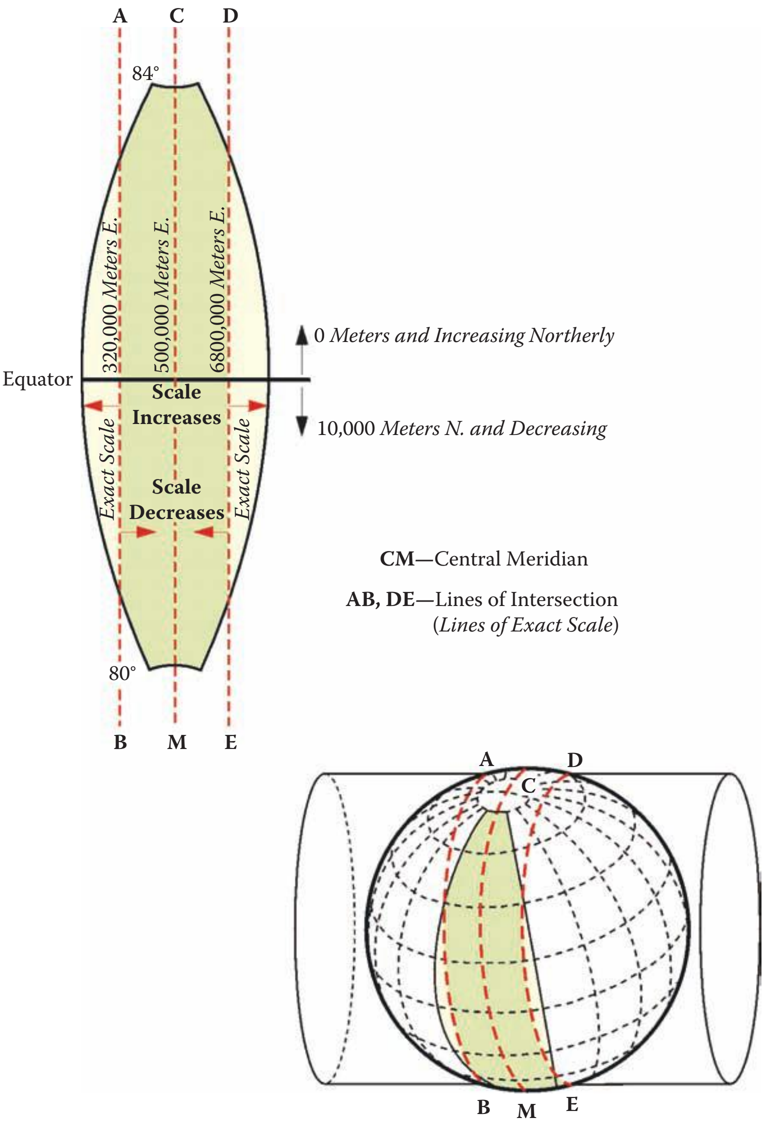


FIGURE 5.17 A UTM Zone

covers latitudes 80° S to 90° S. The scale factor is 0.994 and the false easting and northing are both 2000 km. Their foundation is the International Ellipsoid and its units are meters, just as in the UTM system in general.

There are free utilities online to convert UTM coordinates to Latitude and Longitude at http://www.ngs.noaa.gov/cgi-bin/utm_getgp.prl, and one to convert Latitude and Longitude to UTM coordinates at http://www.ngs.noaa.gov/cgi-bin/utm_getut.prl, both courtesy of NGS.

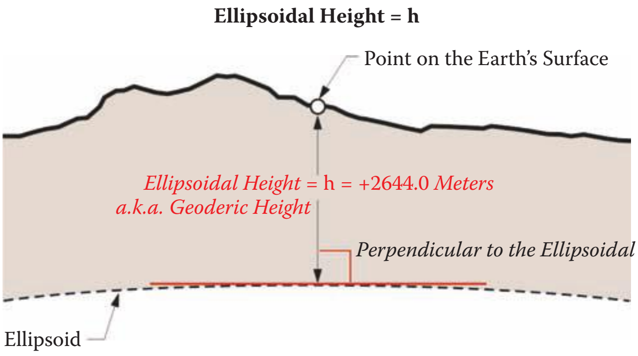


FIGURE 5.18 An Ellipsoidal Height

HEIGHTS

ELLIPSOIDAL HEIGHTS

A point on the earth's surface is not completely defined by its latitude and longitude. In such a context there is, of course, a third element, that of height. Surveyors have traditionally referred to this component of a position as its elevation. One classical method of determining elevations is spirit leveling. A level, correctly oriented at a point on the surface of the earth, defines a line parallel to the geoid at that point. Therefore, the elevations determined by level circuits are orthometric; that is, they are defined by their vertical distance above the geoid as it would be measured along a plumb line. However, orthometric elevations are not directly available from the geocentric position vectors derived from GPS measurements.

As mentioned before, modern geodetic datums rely on the surfaces of geocentric ellipsoids to approximate the surface of the earth. But the actual surface of the earth does not coincide with these nice smooth surfaces, even though that is where the points represented by the coordinate pairs lay. Abstract points may be on the ellipsoid, but the physical features those coordinates intend to represent are on the earth. Though the intention is for the earth and the ellipsoid to have the same center, the surfaces of the two figures are certainly not in the same place. There is a distance between them.

The distance represented by a coordinate pair on the reference ellipsoid to the point on the surface of the earth is measured along a line perpendicular to the ellipsoid. This distance is known by more than one name. It is called the *ellipsoidal height* and it is also called the *geodetic height* and is usually symbolized by *h*.

In Figure 5.18 the ellipsoidal height of a station is illustrated. The concept of an ellipsoidal height is straightforward. A reference ellipsoid may be above or below the surface of the earth at a particular place. If the ellipsoid's surface is below the surface of the earth at the point the ellipsoidal height has a positive sign, if the ellipsoid's surface is above the surface of the earth at the point the ellipsoidal height has a negative sign. But it is important to remember that the measurement of an ellipsoidal height is along a line perpendicular to the ellipsoid, not along a plumb line. Most often they are not the same. And since a reference ellipsoid is a geometric

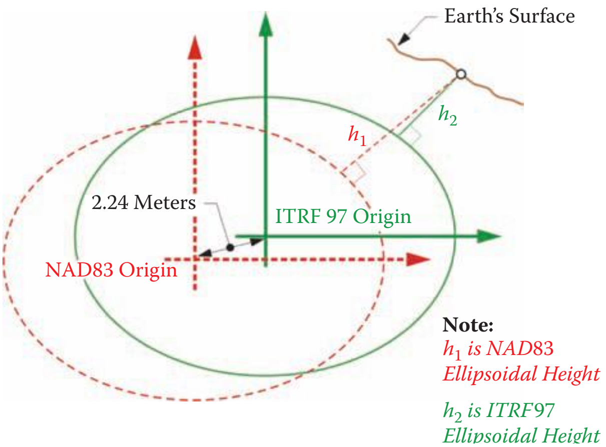


FIGURE 5.19 A Shift

imagining, it is quite impossible to actually set up an instrument on it. That makes it tough to measure ellipsoidal height using surveying instruments. In other words, ellipsoidal height is not what most people think of as an elevation. Said another way, an ellipsoidal height is not measured in the direction of gravity. It is not measured in the conventional sense of down or up.

Nevertheless, the ellipsoidal height of a point is readily determined using a GPS receiver. GPS can be used to discover the distance from the geocenter of the earth to any point on the earth, or above it for that matter. In other words, it has the capability of determining three-dimensional coordinates of a point in a short time. It can provide latitude and longitude, and if the system has the parameters of the reference ellipsoid in its software it can calculate the ellipsoidal height. The relationship between points can be further expressed in the ECEF coordinates, X, Y, and Z, or in a *Local Geodetic Horizon System* (*LHGS*) of north, east, and up. Actually, in a manner of speaking, ellipsoidal heights are new, at least in common usage, since they could not be easily determined until GPS became a practical tool in the 1980s.

However, ellipsoidal heights are not all the same, because reference ellipsoids, or sometimes just their origins can differ. For example, an ellipsoidal height expressed in ITRF97 would be based on an ellipsoid with exactly the same shape as the NAD83 ellipsoid, GRS80. Nevertheless the heights would be different because the origin has a different relationship with the earth's surface (see Figure 5.19).

ORTHOMETRIC HEIGHTS

Spirit Leveling

Long before ellipsoidal heights were so conveniently available, knowing the elevation of a point was critical to the complete definition of a position. In fact there are more than 200 different vertical datums in use in the world today. They were, and still are, determined by spirit leveling.

It is difficult to overstate the amount of effort devoted to differential spirit level work that has carried vertical control across the United States. The transcontinental precision leveling surveys done by the Coast and Geodetic Survey from coast to coast were followed by thousands of miles of spirit leveling work of varying precision. When the 39th parallel survey reached the west coast in 1907 there were approximately 19,700 miles, 31,789 km, of geodetic leveling in the national network. That was more than doubled 22 years later in 1929 to approximately 46,700 miles, 75,159 km. As the quantity of leveling information grew so did the errors and inconsistencies. The foundation of the work was ultimately intended to be *Mean Sea Level* (*MSL*) as measured by *tide station gauges*. Inevitably this growth in leveling information and benchmarks made a new general adjustment of the network necessary to bring the resulting elevations closer to their true values relative to Mean Sea Level.

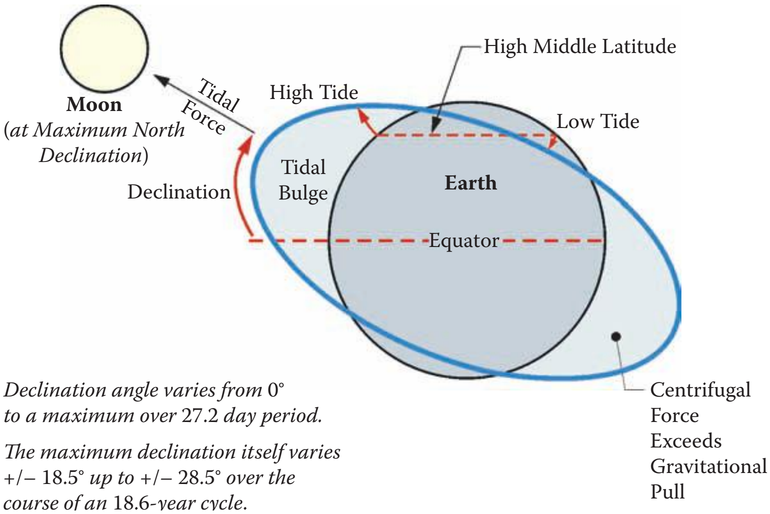
There had already been four previous general adjustments to the vertical network across the United States by 1929. They were done in 1900, 1903, 1907, and 1912. The adjustment in 1900 was based upon elevations held to Mean Sea Level as determined at five tide stations. The adjustments in 1907 and 1912 left the eastern half of the United States fixed as adjusted in 1903. In 1927 there was a special adjustment of the leveling network. This adjustment was not fixed to Mean Sea Level at all tide stations and after it was completed, it became apparent that the Mean Sea Level surface as defined by tidal observations had a tendency to slope upwards to the north along both the Pacific and Atlantic Coasts, with the Pacific being higher than the Atlantic.

But in the adjustment that established the Sea Level Datum of 1929, the determinations of Mean Sea Level at 26 tide stations, 21 in the United States, and 5 in Canada, were held fixed. Sea level was the intended foundation of these adjustments and it might make sense to say a few words about the forces that shape it.

EVOLUTION OF A VERTICAL DATUM

Sea Level

Both the sun and the moon exert tidal forces on the earth, but the moon's force is greater. The sun's tidal force is about half of that exerted on the earth by the moon. The moon makes a complete elliptical orbit around the earth every 27.3 days. There is a gravitational force between the moon and the earth. Each pulls on the other. And at any particular moment the gravitational pull is greatest on the portion of the earth that happens to be closest to the moon. That produces a bulge in the waters on the earth in response to the tidal force. On the side of the earth opposite the bulge centrifugal force exceeds the gravitational force of the earth, and water in this area is forced out away from the surface of the earth, creating another bulge. But the two bulges are not stationary, they move across the surface of the earth. They move because not only is the moon moving slowly relative to the earth as it proceeds along its orbit, but more importantly the earth is rotating in relation to the moon. And the earth's rotation is relatively rapid in comparison with the moon's movement. Therefore, a coastal area in the high middle latitudes may find itself with a high tide early in the day when it is close to the moon and a low tide in the middle of the day when it is has rotated away from it. And this cycle will begin again with another high tide a bit more than 24 hours after the first high tide. A bit more than 24 hours because from the moment the moon reaches a particular meridian to the next time it is there is actually about 24 hours and 50 minutes, a period called a lunar day.



Declination angle varies from 0° *to a maximum over* 27.2 *day period.*

The maximum declination itself varies +/– 18.5° *up to* +/– 28.5° *over the course of an* 18.6-*year cycle.*

FIGURE 5.20 Tides

This sort of tide with one high water and one low water in a lunar day is known as a *diurnal tide*. This characteristic tide would be most likely to occur in the middle latitudes to the high latitudes when the moon is near its maximum declination as you can see from Figure 5.20.

DIURNAL TIDE

The declination of a celestial body is similar to the latitude of a point on the earth. It is an angle measured at the center of the earth from the plane of the equator, positive to the north and negative to the south, to the subject, which is in this case the moon. The moon's declination varies from its minimum of 0° at the equator to its maximum over a 27.2-day period, and that maximum declination oscillates too. It goes from +/–18.5° up to +/–28.5° over the course of an 18.6-year cycle.

Another factor that contributes to the behavior of tides is the elliptical nature of the moon's orbit around the earth. When the moon is closest to the earth, that is its *perigee*, the gravitational force between the earth and the moon is 20% greater than usual. At *apogee*, when the moon is farthest from the earth, the force is 20% less than usual. The variations in the force have exactly the affect you would expect on the tides, making them higher and lower than usual. It is about 27.5 days from perigee to perigee.

To summarize, the moon's orbital period is 27.3 days. It also takes 27.2 days for the moon to move from its maximum declinations back to 0° directly over the equator. And there are 27.5 days from one perigee to the next. You can see that these cycles are almost the same, almost, but not quite. They are just different enough that it takes from 18 to 19 years for the moon to go through the all the possible combinations of its cycles with respect to the sun and the moon. And therefore, if you want to be certain that you have recorded the full range of tidal variation at a place, you must observe and record the tides at that location for 19 years.

This 19-year period, sometimes called the *Metonic cycle*, is the foundation of the definition of Mean Sea Level (MSL). MSL can be defined as the arithmetic mean of hourly heights of the sea at a primary-control tide station observed over a period of 19 years. The mean in Mean Sea Level refers to the average of these observations over time at one place. It is important to note that it does not refer to an average calculation made from measurements at several different places. Therefore, when the Sea Level Datum of 1929 was fixed to MSL at 26 tide stations that meant it was made to fit 26 different and distinct Local Mean Sea Levels. In other words, it was warped to coincide with 26 different elevations.

The topography of the sea changes from place to place and that means, for example, that MSL in Florida is not the same as MSL in California. The fact is Mean Sea Level varies. And the water's temperature, salinity, currents, density, wind, and other physical forces all cause changes in the sea surface's topography. For example, the Atlantic Ocean north of the Gulf Stream's strong current is around 1m lower than it is further south. And the more dense water of the Atlantic is generally about 40 cm lower than the Pacific. At the Panama Canal the actual difference is about 20 cm from the east end to the west end.

A Different Approach

After it was formally established, thousands of miles of leveling were added to the Sea Level Datum of 1929 (SLD29). The Canadian network also contributed data to the Sea Level Datum of 1929, but Canada did not ultimately use what eventually came to be known as the *National Geodetic Vertical Datum of 1929* (*NGVD 29).* The name was changed May 10, 1973, because in the end the final result did not really coincide with Mean Sea Level. It became apparent that the precise leveling done to produce the fundamental data had great internal consistency, but when the network was warped to fit so many tide station determinations of Mean Sea Level that consistency suffered.

By the time the name was changed to NGVD29 in 1973 there were more than 400,000 miles of new leveling work included. There were distortions in the network. Original benchmarks had been disturbed, destroyed, or lost. The NGS thought it time to consider a new adjustment. This time there was a different approach. Instead of fixing the adjustment to tidal stations the new adjustment would be minimally constrained. That means that it would be fixed to only one station, not 26. That station turned out to be Father Point/Rimouski, an *International Great Lakes Datum of 1985* (*IGLD 85*) station near the mouth of the St. Lawrence River and on its southern bank. In other words, for all practical purposes the new adjustment of the huge network was not intended to be a sea level datum at all. It was a change in thinking that was eminently practical.

While is it relatively straightforward to determine Mean Sea Level in coastal areas, carrying that reference reliably to the middle of a continent is quite another matter. Therefore, the new datum would not be subject to the variations in sea surface topography. It was unimportant whether the new adjustment's zero elevation and Mean Sea Level were the same thing or not.

The Zero Point

Precise leveling proceeded from the zero reference established at Pointe-au-Père, Quebec in 1953. The resulting benchmark elevations were originally published in September 1961. The result of this effort was *International Great Lakes Datum 1955*. After nearly 30 years, the work was revised. The revision effort began in 1976 and the result was *IGLD 1985*. It was motivated by several developments including deterioration of the zero reference point gauge location and improved surveying methods. But one of the major reasons for the revision was the movement of previously established benchmarks due to *isostatic rebound*. This effect is literally the earth's crust rising slowly, rebounding, from the removal of the weight and subsurface fluids caused by the retreat of the glaciers from the last ice age.

The choice of the tide gauge at Pointe-au-Père, Quebec as the zero reference for IGLD was logical in 1955. It was reliable. It had already been connected to the network with precise leveling. It was at the outlet of the Great Lakes. But by 1984 the wharf at Pointe-au-Père had deteriorated and the gauge was moved. And it was subsequently moved about 3 miles to Rimouski, Quebec and precise levels were run between the two. It was there that the zero reference for IGLD 1985 and what became a new adjustment called *North American Vertical Datum 1988, NAVD88* was established.

The re-adjustment, known as NAVD88, was begun in the 1970s. It addressed the elevations of benchmarks all across the nation. The effort also included field work. Destroyed and disturbed benchmarks were replaced and over 50,000 miles of leveling were actually redone before NAVD88 was ready in June of 1991. The differences between elevations of benchmarks determined in NGVD29 compared with the elevations of the same benchmarks in NAVD88 vary from approximately –1.3 feet in the east to approximately +4.9 feet in the west in the 48 coterminous states of the United States. The larger differences tend to be on the coasts, as one would expect since NGVD29 was forced to fit Mean Sea Level at many tidal stations and NAVD88 was held to just one.

THE GEOID

Any object in the earth's gravitational field has *potential energy* derived from being pulled toward the Earth. Quantifying this potential energy is one way to talk about height, because the amount of potential energy an object derives from the force of gravity is related to its height. There are an infinite number where the potential of gravity is always the same. They are known as *equipotential surfaces*.

Mean Sea Level itself is not an equipotential surface at all, of course. Forces other than gravity affect it, forces such as temperature, salinity, currents, wind, and so forth. The geoid, on the other hand, is defined by gravity alone. The geoid is the particular equipotential surface arranged to fit Mean Sea Level as well as possible, in a least squares sense (see Figure 5.21).

So while there is a relationship between Mean Sea Level and the geoid, they are not the same. They could be the same if the oceans of the world could be utterly still, completely free of currents, tides, friction, variations in temperature, and all

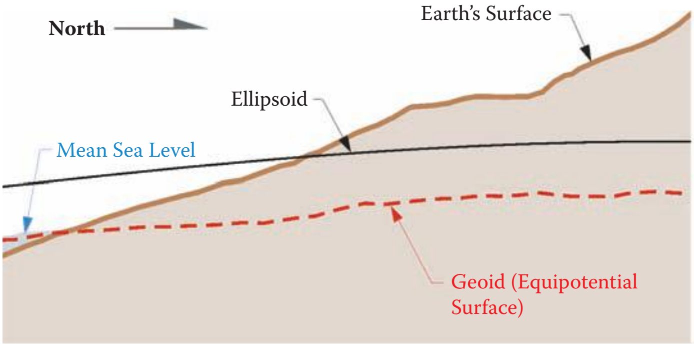


FIGURE 5.21 Ellipsoid–Geoid–Mean Sea Level

other physical forces, except gravity. Reacting to gravity alone, these unattainable calm waters would coincide with the geoid. If the water was then directed by small frictionless channels or tubes and allowed to migrate across the land, the water would then, theoretically, define the same geoidal surface across the continents, too. Of course, the 70% of the earth covered by oceans is not so cooperative, and the physical forces cannot really be eliminated. These unavoidable forces actually cause Mean Sea Level to deviate up to 1, even 2, meters from the geoid.

Since the geoid is completely defined by gravity it is not smooth and continuous. It is lumpy because gravity is not consistent across the surface of the earth. At every point gravity has a magnitude and a direction. Anywhere on the earth, a vector can describe gravity, but these vectors do not all have the *same* direction or magnitude. Some parts of the earth are denser than others. Where the earth is denser, there is more gravity, and the fact that the earth is not a sphere also affects gravity. It follows then that defining the geoid precisely involves actually measuring the direction and magnitude of gravity at many places.

The geoid undulates with the uneven distribution of the mass of the earth. It has all the irregularity that the attendant variation in gravity implies. In fact, the separation between the lumpy surface of the geoid and the smooth GRS80 ellipsoid worldwide varies from about +85 meters west of Ireland to about -106 meters, the latter in the area south of India near Ceylon.

In the coterminous United States, sometimes-abbreviated *CONUS*, the distances between the geoid and the GRS80 ellipsoid are less. They vary from about –8 meters to about –53 meters. But even these smaller distances are larger than those under the old NAD27. Please recall that its orientation at Meades Ranch, Kansas was arranged so that the distance between the Clarke 1866 ellipsoid and the geoid was zero. And across the United States the difference between them never grew to more than 12 meters. In fact, for all practical purposes the ellipsoid and the geoid were often assumed to coincide in that system. However, in NAD83 based on GRS80 *geoid heights* are larger.

A geoid height is the distance measured along a line perpendicular to the ellipsoid of reference to the geoid. Also, as you can see, these geoid heights are negative. They are usually symbolized, *N*. If the geoid is above the ellipsoid, N is positive; if the geoid is below the ellipsoid, N is negative. It is negative here because the geoid is underneath the ellipsoid throughout the coterminous United States. In Alaska it is the other way around; the ellipsoid is underneath the geoid and N is positive.

Please recall that an ellipsoid height is symbolized, h. The ellipsoid height is also measured along a line perpendicular to the ellipsoid of reference, but to a point on the surface of the earth. However, an orthometric height, symbolized, *H,* is measured along a plumb line from the geoid to a point on the surface of the earth.
In either case by using the formula,

$$H = h - N$$

one can convert an ellipsoidal height, h, derived say from a GPS observation, into an orthometric height, H, by knowing the extent of geoid-ellipsoid separation, also known as the geoidal height, N, at that point.

As you can see from Figure 5.22 the ellipsoid height of a particular point is actually smaller than the orthometric height throughout the coterminous United States.

The formula H = h – N does not account for the fact that the plumb line along which an orthometric height is measured is curved as you see in Figure 5.22. Curved because it is perpendicular with each and every equipotential surface through which it passes. And since equipotential surfaces are not parallel with each other, the plumb line must curve to maintain perpendicularity with them. This deviation of a plumb line from the perpendicular to the ellipsoid reaches about 1 minute of arc in only the most extreme cases. Therefore, any height difference that is caused by the curvature is negligible. It would take a height of over 6 miles for the curvature to amount to even 1 mm of difference in height.

Geoid Models

Major improvements have been made over the past quarter century or so in the mapping the geoid on both national and global scales. And because there are large complex variations in the geoid related to both the density and relief of the earth, geoid models and interpolation software have been developed to support the conversion of GPS elevations to orthometric elevations. For example, in early 1991 NGS presented a program known as GEOID90. This program allowed a user to find *N*, the geoidal height, in meters for any NAD83 latitude and longitude in the United States.

The GEOID90 model was computed at the end of 1990, using over a million gravity observations. It was followed by the GEOID93 model. It was computed at the beginning of 1993 using more than 5 times the number of gravity values used to create GEOID90. Both provided a grid of geoid height values in a 3 minutes of latitude by 3 minutes of longitude grid with an accuracy of about 10 cm. Next the GEOID96 model resulted in a gravimetric geoid height grid in a 2 minutes of latitude by 2 minutes of longitude grid.

GEOID99 covered the coterminous United States, and it includes U.S. Virgin Islands, Puerto Rico, Hawaii, and Alaska. The grid is 1 degree of latitude by 1 degree of longitude and it is the first to combine gravity values with GPS ellipsoid heights on previously leveled benchmarks. According to NGS, “When comparing the GEOID99 model with GPS ellipsoid heights in the NAD 83 reference frame

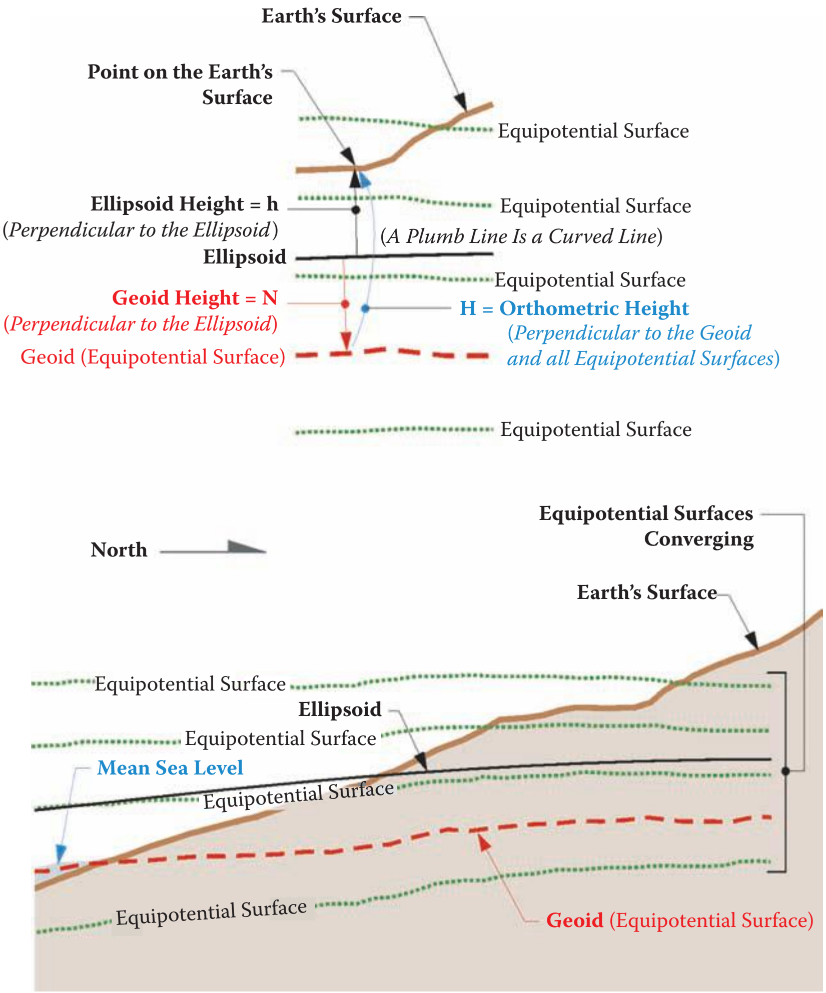


FIGURE 5.22 Height Conversion

and leveling in the NAVD 88 datum, it is seen that GEOID99 has roughly a 4.6 cm absolute accuracy (one sigma) in the region of GPS on benchmark coverage. In those states with sparse (150 km+) GPS on benchmark coverage, less point accuracy may be evident" (http://www.ngs.noaa.gov/GEOID/GEOID99/g1999.txt, the GEOID99 README FILE Version, January 12, 2000, the GEOID99 Model). GEOID03 has superseded the previous models for the continental United States and improvements will undoubtedly continue.

EXERCISES

1. What is the datum used for the GPS Navigation message?

   (a) NAD83 (CORS96)
   (b) NAD27
   (c) GRS80
   (d) WGS84 (G1150)

2. What technology contributed the least number of positions to the original least-squares adjustment of NAD83?

   (a) EDM baselines
   (b) TRANSIT Doppler positions
   (c) Conventional optical surveying
   (d) GPS

3. What is the highest minimum geometric accuracy specified by the FGCC standards for Order AA, Order A, and Order B GPS surveys at the 95%, or 2σ, confidence level?

   (a) 1mm ± 0.01 ppm
   (b) 3mm ± 0.01 ppm
   (c) 5mm ± 0.1 ppm
   (d) 8mm ± 1 ppm

4. In the State Plane Coordinate systems in the United States, which mapping projection listed below is not used?

   (a) the transverse Mercator projection
   (b) the oblique Mercator projection
   (c) the Lambert conformal conic projection
   (d) the universal transverse Mercator projection

5. What information is necessary to convert an ellipsoidal height to an orthometric height?

   (a) the geoid height
   (b) the State Plane coordinate
   (c) the semimajor axis of the ellipsoid
   (d) the GPS Time

6. Which statement about the geoid is correct?

   (a) The geoid's surface is always perpendicular to gravity.
   (b) The geoid's surface is the same as mean sea level.
   (c) The geoid's surface is always parallel with an ellipsoid.
   (d) The geoid's surface is the same as the topographic surface.

7. What acronym is used to describe the cooperative ventures that took place between NGS and the states to provide extremely accurate, vehicle-accessible, regularly spaced control points with good overhead visibility for GPS?

   (a) ITRS
   (b) HARN
   (c) CORS
   (d) NAVD

8. Which UTM zones cover the coterminous United States?

   (a) Zones 10 North to 18 North
   (b) Zones 1 North to 12 North
   (c) Zones 6 North to 30 North
   (d) Zones 20 North to 30 North

9. Which of the following organizations currently maintains the International Terrestrial Reference System (*ITRS*)?

   (a) NGS
   (b) IERS
   (c) BIH
   (d) C&GS

10. The combined factor is used in SPCS conversion. How is the combined factor calculated?

    (a) The scale factor is multiplied by the elevation factor.
    (b) The scale factor is divided by the grid factor.
    (c) The grid factor added to the elevation factor and the sum is divided by two.
    (d) The scale factor is multiplied by the grid factor.

ANSWERS AND EXPLANATIONS

1. Answer is (c)

Explanation: With very slight changes, GRS80 is the reference ellipsoid for the coordinate system, known as the *World Geodetic System 1984* (*WGS84*). This datum has been used by the U.S. Military since January 21, 1987, as the basis for the GPS Navigation message computations. Therefore, coordinates provided directly by GPS receivers are based in WGS84. The newest incarnation of WGS84 is WGS84 (G1150). It was implemented by GPS *Operational Control Segment* January 20, 2002.

2. Answer is (d)

Explanation: It took more than 10 years to readjust and redefine the horizontal coordinate system of North America into what is now NAD83. More than 1.7 million positions derived from classical surveying techniques throughout the Western Hemisphere were involved in the least-squares adjustment. They were supplemented by approximately 30,000 EDM measured baselines, 5,000 astronomic azimuths, and 650 Doppler stations positioned by the TRANSIT satellite system. Over 100 Very Long Baseline Interferometry (VLBI) vectors were also included. But GPS, in its infancy, contributed only five points.

3. Answer is (b)

Explanation: The FGCC is an organization chartered in 1968 and composed of representatives from 11 agencies of the federal government. It revises and updates surveying standards for geodetic control networks, among other duties. Its provisional standards and specifications for GPS work include Orders AA, A, and B, which are defined as having minimum geometric accuracies of 3mm “ 0.01 ppm, 5mm “ 0.1 ppm and 8mm “ 1 ppm, respectively, at the 95%, or 2σ, confidence level.

4. Answer is (d)

Explanation: In the United States, state plane systems are based on the transverse Mercator projection, an oblique Mercator projection, and the Lambert conic map projection, grid. Every state, Puerto Rico, and the U.S. Virgin Islands has its own plane rectangular coordinate system.

5. Answer is (a)

Explanation: The geoid undulates with the uneven distribution of the mass of the earth and has all the irregularity that implies. In fact, the separation between the bumpy surface of the geoid and the smooth GRS80 ellipsoid varies from 0 up to 100 meters. Therefore, the only way a surveyor can convert an ellipsoidal height from a GPS observation on a particular station into a useable orthometric elevation is to know the extent of geoid-ellipsoid separation, also known as the *geoid height*, at that point.

Toward that end, major improvements have been made over the past quarter century or so in the mapping the geoid on both national and global scales. This work has gone a long way toward the accurate determination of the geoid-ellipsoid separation, or geoid height, known as N. The formula for transforming ellipsoidal heights, h, into orthometric elevations, H, is $H = h - N$.

6. Answer is (a)

Explanation: The geoid is a representation of the earth's gravity field. It is an equipotential surface that is everywhere perpendicular to the direction of gravity. In other words, it is perpendicular to a plumb line at every point.

Mean sea level is the average height of the surface of the sea for all stages of the tide. It was and sometimes still is used as a reference for elevations. However, it is not the same as the geoid. Mean sea level departs from the surface of the geoid; these displacements are known as the sea surface topography. Neither is the ellipsoid, a smooth mathematically defined surface, always parallel to the bumpy geoid. Finally, the geoid is certainly not coincident with the topographic surface of the earth.

7. Answer is (b)

Explanation: The creation of *High Accuracy Reference Networks* (*HARN*) were cooperative ventures between NGS and the states, and often include other organizations as well.

With heavy reliance on GPS observations, these networks are intended to provide extremely accurate, vehicle-accessible, regularly spaced control points with good overhead visibility. To ensure coherence, when the GPS measurements are complete, they were submitted to NGS for inclusion in a statewide readjustment of the existing NGRS covered by the state. Coordinate shifts of 0.3 to 1.0 m from NAD83 values have been typical in these readjustments.

8. Answer is (a)

Explanation: The UTM projection divides the world into 60 zones that begin at λ 180°, each with a width of 6° of longitude, extending from 84° Nφ and 80° Sφ. Its coverage is completed by the addition of two polar zones. The coterminous United States are within UTM zones 10 to 18.

The UTM grid is defined in meters. Each zone is projected onto a cylinder that is oriented in the same way as that used in the transverse Mercator state plane coordinates described above. The radius of the cylinder is chosen to keep the scale errors within acceptable limits. Coordinates of points from the reference ellipsoid within a particular zone are projected onto the UTM grid.

The intersection of each zone's central meridian with the equator defines its origin of coordinates. In the southern hemisphere, each origin is given the coordinates: easting = X_0 = 500,000 meters, and northing = Y_0 = 10,000,000 meters, to ensure that all points have positive coordinates. In the northern hemisphere, the values are: easting = X_0 = 500,000 meters, and northing = Y_0 = 0 meters, at the origin. The scale factor grows from 0.9996 along the central meridian of a UTM zone to 1.00000 at 180,000 meters to the east and west.

9. Answer is (b)

Explanation: The best geocentric reference frame currently available is the International Terrestrial Reference Frame, *ITRF*. Its origin is at the center of mass of the whole earth including the oceans and atmosphere. The unit of length is the meter. The orientation of its axes was established as consistent with that of the IERS's predecessor, Bureau International de l'Heure, BIH, at the beginning of 1984.

Today, the ITRF is maintained by the International Earth Rotation Service, IERS, which monitors Earth Orientation Parameters (EOP) for the scientific community through a global network of observing stations. This is done with GPS, Very Long Baseline Interferometry, (VLBI), Lunar Laser Ranging (LLR), Satellite Laser Ranging (SLR), Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS), and the positions of the observing stations are now considered to be accurate to the centimeter level.

The International Terrestrial Reference Frame (ITRF) is actually a series of realizations. In other words, it is revised and published on a regular basis. And today NAD83 can be realizably defined in terms of a best-fit transformation from ITRF96.

10. Answer is (d)

Explanation: The grid factor changes with the ellipsoidal height of the line. It also changes with its location in relation to the standard lines of its SPCS zone. The grid factor is derived by multiplying the scale factor by the elevation factor. The product is nearly 1 and is known as either the grid factor or the combination factor. There is a different combination factor for every line in the correct application of SPCS.

6 GPS Surveying Techniques

STATIC GPS SURVEYING

If a static GPS control survey is carefully planned, it usually progresses smoothly. The technology has virtually conquered two stumbling blocks that have defeated the plans of conventional surveyors for generations. Inclement weather does not disrupt GPS observations, and a lack of intervisibility between stations is of no concern whatsoever, at least in postprocessed GPS. Still, GPS is far from so independent of conditions in the sky and on the ground that the process of designing a survey can now be reduced to points-per-day formulas, as some would like. Even with falling costs, the initial investment in GPS remains large by most surveyors' standards. However, there is seldom anything more expensive in a GPS project than a surprise.

PLANNING

New Standards

The Federal Geodetic Control Committee (FGCC) has written provisional accuracy standards for GPS relative positioning techniques. The older standards of first, second, and third order are classified under the group C in the new scheme. In the past, the cost of achieving first-order accuracy was considered beyond the reach of most conventional surveyors. Besides, surveyors often said that such results were far in excess of their needs anyway. The burden of the equipment, techniques, and planning that is required to reach its 2σ relative error ratio of 1 part in 100,000 was something most surveyors were happy to leave to government agencies. But the FGCC's proposed new standards of B, A, and AA are respectively 10, 100, and 1000 times more accurate than the old first-order. The attainment of these accuracies does not require corresponding 10-, 100-, and 1000-fold increases in equipment, training, personnel, or effort. They are now well within the reach of private GPS surveyors both economically and technically.

New Design Criteria

These upgrades in accuracy standards not only accommodate control by static GPS; they also have cast survey design into a new light for many surveyors. Nevertheless, it is not correct to say that every job suddenly requires the highest achievable accuracy, nor is it correct to say that every GPS survey now demands an elaborate design. In some situations, a crew of two, or even one surveyor on-site may carry a GPS survey from start to finish with no more planning than minute-to-minute decisions can provide even though the basis and the content of those decisions may be quite different from those made in a conventional survey.

In areas that are not heavily treed and generally free of overhead obstructions, the now-lower C group of accuracy may be possible without a prior design of any significance. But while it is certainly unlikely that a survey of photocontrol or work on a cleared construction site would present overhead obstructions problems comparable with a static GPS control survey in the Rocky Mountains, even such open work may demand preliminary attention. For example, just the location of appropriate vertical and horizontal control stations or obtaining permits for access across privately owned property or government installations can be critical to the success of the work.

The Lay of the Land

An initial visit to the site of the survey is not always possible. Today, online mapping browsers are making virtual site evaluation possible as well. Topography as it affects the line of sight between stations is of no concern on a static GPS project, but its influence on transportation from station to station is a primary consideration. Perhaps some areas are only accessible by helicopter or other special vehicle. Initial inquiries can be made. Roads may be excellent in one area of the project and poor in another. The general density of vegetation, buildings, or fences may open general questions of overhead obstruction or multipath. The pattern of land ownership relative to the location of project points may raise or lower the level of concern about obtaining permission to cross property.

Maps

Maps, both digital and hard-copy, are particularly valuable resources for preparing a static GPS survey design. Local government and private sources can sometimes provide appropriate mapping, or it maybe available online. Other mapping that may be helpful is available from various government agencies: for example, the U.S. Forest Service in the Department of Agriculture; the Department of Interior's Bureau of Land Management, Bureau of Reclamation, and National Park Service; the U.S. Fish and Wildlife Service in the Department of Commerce; and the Federal Highway Administration in the Department of Transportation are just a few of them. Even county and city maps should be considered since they can sometimes provide the most timely information available.

Depending on the scope of the survey, various scales and types of maps can be useful. For example, a GPS survey plan may begin with the plotting of all potential control and project points on a map of the area. However, one vital element of the design is not available from any of these maps: the *National Spatial Reference System* (*NSRS*) stations.

NATIONAL GEODETIC SURVEY (NGS) CONTROL

NGS Control Data Sheets

It is important to understand the information available on NGS data sheets. A rectangular search based upon the range of latitudes and longitudes can now be performed on the NGS Internet site. It is also possible to do a radial search, defining the region of the survey with one center position and a radius. You may also retrieve individual data sheets by the Permanent Identifier (PID) control point name, which is known as the *designation*, survey project identifier or USGS quad. It is best to ask for the desired horizontal and vertical information within a region that is somewhat larger than that which is contained by the boundaries of the survey. The Internet address for NGS Data Sheets is http://www.ngs.noaa.gov/cgi-bin/datasheet.prl. There is a huge amount of information about survey monuments on each individual sheet. NGS also provides a very convenient GIS map interface called the *NGS Survey Control Map* from data sheets may be retrieved from http://www.ngs.noaa.gov/ims/NgsMap2/viewer.htm.

The information available from an NGS control sheet is valuable at the earliest stage of a GPS survey (see Figure 6.1). In addition to the latitude and longitude, the published data include the state plane coordinates in the appropriate zones. The coordinates facilitate the plotting of the station's position on the project map.

The first line of each data sheet includes the retrieval date. Then the station's category is indicated. There are several, and among them are Continuously Operating Reference Station, Federal Base Network Control Station, and Cooperative Base Network Control Station.

This is followed by the station's designation, which is its name, and its PID. Either of these may be used to search for the station in the NGS database. The PID is also found all along the left side of each data sheet record and is always two upper case letters followed by four numbers.

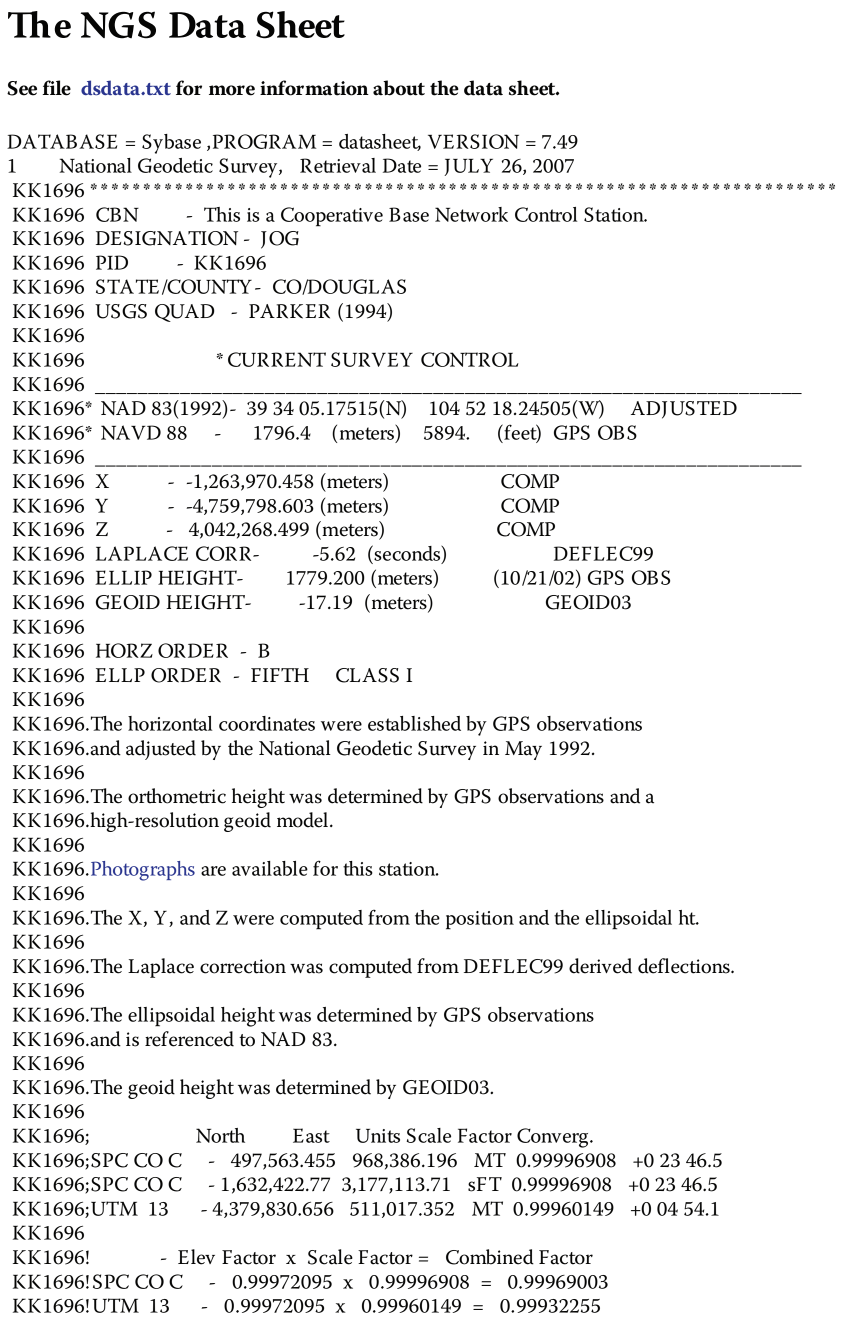
The state, county, and U.S. Geological Survey (USGS)-7.5 minute quad name follows. Even though the station is located in the area covered by the quad sheet, it may not actually appear in the map.

Under the heading, "Current Survey Control," you will find the latitude and longitude of the station in NAD83 and its height in NAVD88. Adjustments to NAD27 and NGVD29 datums are a thing of the past. However, these old values may be shown under *Superseded Survey Control*. Horizontal values may be either *Scaled*, if the station is a benchmark or *Adjusted*, if the station is indeed a horizontal control point.

When a date is shown in parentheses after NAD83 in the data sheet it means that the position has been readjusted since. Often these new adjustments are due to the station's inclusion in a state High Accuracy Reference Network (HARN) effort. There is more information on these cooperative projects in Chapter 5.

There are 13 sources of vertical control values shown on NGS data sheets. Here are a few of the categories. There is *Adjusted,* which are given to 3 decimal places and are derived from least squares adjustment of precise leveling. Another category is *Posted,* which indicates that the station was adjusted after the general NAVD adjustment in 1991. When a station's elevation has been found by precise leveling but non-rigorous adjustment, it is called *Computed*.

Stations' vertical values are given to 1 decimal place if they are from GPS observation (*Obs*) or vertical angle measurements (*Vert Ang*). And they have no decimal places if they were scaled from topographic map, *Scaled,* or found by conversion from NGVD29 values using the program known as *VERTCON*.



The NGS Data Sheet

See file dsdata.txt for more information about the data sheet.

```
DATABASE = Sybase ,PROGRAM = datasheet, VERSION = 7.49
1      National Geodetic Survey,   Retrieval Date = JULY 26, 2007
 KK1696 ***********************************************************************
 KK1696  CBN          -  This is a Cooperative Base Network Control Station.
 KK1696  DESIGNATION -  JOG
 KK1696  PID         -  KK1696
 KK1696  STATE/COUNTY-  CO/DOUGLAS
 KK1696  USGS QUAD   -  PARKER (1994)
 KK1696
 KK1696                        *CURRENT SURVEY CONTROL
 KK1696  ___________________________________________________________________
 KK1696*  NAD 83(1992)-  39 34 05.17515(N)    104 52 18.24505(W)     ADJUSTED
 KK1696*  NAVD 88     -      1796.4    (meters)     5894.     (feet)  GPS OBS
 KK1696  ___________________________________________________________________
 KK1696  X           -  -1,263,970.458 (meters)                 COMP
 KK1696  Y           -  -4,759,798.603 (meters)                 COMP
 KK1696  Z           -   4,042,268.499 (meters)                 COMP
 KK1696  LAPLACE CORR-           -5.62  (seconds)                    DEFLEC99
 KK1696  ELLIP HEIGHT-        1779.200 (meters)          (10/21/02) GPS OBS
 KK1696  GEOID HEIGHT-          -17.19  (meters)                    GEOID03
 KK1696
 KK1696  HORZ ORDER  -  B
 KK1696  ELLP ORDER  -  FIFTH     CLASS I
 KK1696
 KK1696.The horizontal coordinates were established by GPS observations
 KK1696.and adjusted by the National Geodetic Survey in May 1992.
 KK1696
 KK1696.The orthometric height was determined by GPS observations and a
 KK1696.high-resolution geoid model.
 KK1696
 KK1696.Photographs are available for this station.
 KK1696
 KK1696.The X, Y, and Z were computed from the position and the ellipsoidal ht.
 KK1696
 KK1696.The Laplace correction was computed from DEFLEC99 derived deflections.
 KK1696
 KK1696.The ellipsoidal height was determined by GPS observations
 KK1696.and is referenced to NAD 83.
 KK1696
 KK1696.The geoid height was determined by GEOID03.
 KK1696
 KK1696;                    North         East     Units Scale Factor Converg.
 KK1696;SPC CO C    -   497,563.455   968,386.196   MT  0.99996908   +0 23 46.5
 KK1696;SPC CO C    - 1,632,422.77  3,177,113.71   sFT  0.99996908   +0 23 46.5
 KK1696;UTM  13     - 4,379,830.656   511,017.352   MT  0.99960149   +0 04 54.1
 KK1696
 KK1696!            -  Elev Factor  x  Scale Factor =    Combined Factor
 KK1696!SPC CO C    -   0.99972095  x   0.99996908  =   0.99969003
 KK1696!UTM  13     -   0.99972095  x   0.99960149  =   0.99932255
```

FIGURE 6.1 An NGS Control Data Sheet

When they are available, earth-centered earth-fixed (ECEF) coordinates are shown. These are right-handed system, 3-D Cartesian coordinates. They are the same type of X, Y, and Z coordinates presented in Chapter 5. These values are followed by the quantity which, when added to an astronomic azimuth, yields a geodetic azimuth, is known as the *Laplace correction*.

It is important to note that NGS uses a clockwise rotation regarding the Laplace correction. The ellipsoid height per the NAD83 ellipsoid is shown followed by the geoid height where the position is covered by NGS's GEOID program. Please see Chapter 5 for a more complete discussion of these values.

Survey Order and Class

Here the new accuracy standards mentioned earlier come into play. On NGS data sheets each adjusted control station will be assigned a horizontal, vertical (orthometric), and vertical (ellipsoid) order and class, where they apply.

Regarding horizontal control stations, first-, second-, and third-order continue to be published under group C. However, these designations are now augmented by AA-, A-, and B-order stations as well. Horizontal AA-order stations have a relative accuracy of 3 mm +/- 1:100,000,000 relative to other AA-order stations. Horizontal A-order stations have a relative accuracy of 5 mm +/- 1:10,000,000 relative to other A-order stations. Horizontal B-order stations have a relative accuracy of 8 mm +/- 1: 1,000,000 relative to other A- and B-order stations.

Order and class continue to be published in first-, second-, and third-order for orthometric vertical control stations. Under the orders, class 0 is sometimes used. First-order, class 0 is used for stations whose tolerance is 2.0 mm or less. Second-order, class 0 is used for stations whose tolerance is 8.4 mm or less. Third-order, class 0 is used for stations whose tolerance is 12.0 mm or less. Posted benchmarks are given a distribution rate code from a to f, respectively, to indicate their reliability from 0mm per km to 8 mm or more per km. Ellipsoid vertical control stations are also given order categories by NGS from first to fifth and each with a class 1 and 2, but the idea has not yet been adopted by the Federal Geodetic Control Committee (FGCC).

Photographs of the station may also be available in some cases. When the data sheet is retrieved online one can use the link provided to bring them up. Also, the geoidal model used is noted.

Coordinates

NGS data sheets also provide State Plane and Universal Transverse Mercator (UTM) coordinates, the latter only for horizontal control stations. State Plane Coordinates are given in either U.S. Survey Feet or International Feet and UTM coordinates are given in meters. Azimuths to the primary azimuth mark are clockwise from north and scale factors for conversion from ellipsoidal distances to grid distances. This information may be followed by distances to reference objects. Coordinates are not given for azimuth marks or reference objects on the data sheet.

The Station Mark

Along with mark setting information, the type of monument and the history of mark recovery, the NGS data sheets provide a valuable *to-reach* description. It begins with the general location of the station. Then starting at a well-known location, the route is described with right and left turns, directions, road names, and the distance traveled along each leg in kilometers. When the mark is reached the monument is described and horizontal and vertical ties are shown. Finally, there may be notes about obstructions to GPS visibility, and so forth.

Significance of the Information

The value of the description of the monument's location and the route used to reach it is directly proportional to the date it was prepared and the remoteness of its location. The conditions around older stations often change dramatically when the area has become accessible to the public. If the age and location of a station increases the probability that it has been disturbed or destroyed then reference monuments can be noted as alternatives worthy of on-site investigation. However, special care ought to be taken to ensure that the reference monuments are not confused with the station marks themselves.

Control from Continuously Operating Networks

The requirement to occupy physical geodetic monuments in the field can be obviated by downloading the tracking data available online from appropriate *continuously operating reference stations* (*CORS*) where their density is sufficient. These stations, also known as *Active Stations,* comprise fiducial networks that support a variety of GPS applications. While they are frequently administered by governmental organizations, some are managed by public–private organizations and some are commercial ventures. The most straightforward benefit of CORS is the user's ability to do relative positioning without operating his own base station by depending on that role being fulfilled by the network's reference stations.

While CORS can be configured to support differential GPS (DGPS) and real-time kinematic (RTK) applications, as in *Real-Time Networks*, most networks constantly collect GPS tracking data from known positions and archive the observations for subsequent download by users from the Internet.

In many instances the original impetus of a network of CORS was geodynamic monitoring as illustrated by the *GEONET* established by the *Geographical Survey Institute* (*GSI*) in Japan after the Kobe earthquake. Networks that support the monitoring of the International Terrestrial Reference System (ITRS) have been created around the world by the *International GNSS Service* (*IGS*), which is a service of the International Association of Geodesy and the Federation of Astronomical and Geophysical Data Analysis Services originally established in 1993. And the *Southern California Integrated GPS Network* (*SCIGN*) is a network run by a government–university partnership.

Despite the original motivation for the establishment of a CORS network, the result has been a boon for high-accuracy GPS positioning. The data collected by these networks is quite valuable to GPS surveyors around the world. Surveyors in the United States can take advantage of the CORS network administered by the National Geodetic Survey, NGS. The continental NGS system is has two components, the Cooperative CORS and the National CORS. Together they comprise a network of hundreds of stations which constantly log dual-frequency GPS data and make the data available in the *receiver independent exchange format* (*RINEX*) format.

NGS Continuously Operating Reference Stations

NGS manages the National CORS system to support postprocessing GPS data. Information is available online at http://www.ngs.noaa.gov/CORS/. Both code and carrier phase GPS data from receivers at these stations throughout the United States and its territories are archived in Silver Springs, Maryland and Boulder, Colorado. That data can then be conveniently downloaded in its original form from the Internet free of charge for up to 30 days after its collection. It is also available later, but after it has been decimated to a 30-second format.

The Cooperative CORS system supplements the National CORS system. The NGS does not directly provide the data from the cooperative system of stations. Its stations are managed by participating university, public, and private organizations that operate the sites. The partners are listed at this address http://www.ngs.noaa.gov/CORS/Organizations/Organizations.html. NGS provides links to that data from their Web page.

Nearly all coordinates provided by NGS for the CORS sites are available in NAD83 (CORS96) epoch 2002.0 and the international reference frame ITRF00. The epoch means that the published NAD 83 coordinate represents the station's position on January 1, 2002. To compute its location at another date one would need to apply the station's velocity, which NGS provides.

Some CORS stations coordinates are not in NAD83 (CORS96). Since the islands in the Pacific move at a rate of centimeters per year relative to the North American tectonic plate to which NAD83 (CORS96) is tied, the coordinates of the CORS there are presented in NAD 83 (PACP00) or NAD 83 (MARP00). Another exception occurs in Alaska where the coordinates of the CORS are available in NAD83 (CORS96) but epoch 2003.00 rather than 2002.0.

The coordinates of CORS stations are also published in ITRF00 and as mentioned in Chapter 5. WGS84 (G1150) is the same as ITRF00. However, these positions differ from NAD83. The ITRF00 coordinates are also accompanied by velocities since they are moving with respect to NAD83. NGS uses the epoch 1997.0, that is January 1, 1997, for its ITRF00 positions. Again, provided velocities can be used to calculate the stations position at a different date using the *Horizontal Time Dependent Positioning* (*HTDP*) utility available at http://www.ngs.noaa.gov/TOOLS/Htdp/Htdp.html.

NGS CORS Reference Points

At a CORS site NGS provides the coordinates of the L1 phase center and the Antenna Reference Point (ARP). Generally speaking it is best to adopt the position that can be physically measured. The coordinates given for the ARP are those that are part of the antenna from which the phase center offsets are calculated that is usually the bottom mount.

As mentioned in Chapter 4, the phase centers of antennas are not immovable points. They actually change slightly, mostly as the elevation of the satellite's signals change. In any case, the phase centers for L1 and L2 differ from the position of the ARP both vertically and horizontally; please see http://www.ngs.noaa.gov/ANTCAL. NGS provides the position of the phase center on average at a particular CORS site. As most postprocessing software will, given the ARP, provide the correction for the phase center of an antenna, based on antenna type, the ARP is the most convenient coordinate value to use.

NGS CORS Precise Orbits

A significant improvement in positioning is available by using the post computed precise orbital data that can be downloaded from the *User Friendly CORS, UFCORS,* portion of the NGS site http://www.ngs.noaa.gov/CORS/. This service will provide the best orbital information available at the time of the download.

The orbits preferred on the NGS CORS site are produced in cooperation with the IGS. The most accurate is known as the precise orbit, which is usually available in ~12 days. Only after a full GPS week's worth of data is available can the precise orbit be completed. There are also rapid orbits, which are available within ~24 hours. With a 5 cm orbital integrity and one-tenth of a nanosecond clock accuracy it is only slightly less reliable than the precise orbit data itself. Ultrarapid data are available within ~6 hours. These are a bit less reliable than the precise orbits.

International Global Navigation Satellite System (GNSS) Service (IGS)

Like NGS, IGS also provides CORS data. However, it has a global scope illustrated by the organizations online map at http://igscb.jpl.nasa.gov/network/maps/allmaps.html. The information on the individual stations can be accessed by clicking on the map. There are variable upload rates for the IGS CORS sites. While http://itrf.ensg.ign.fr/ITRF_solutions/2000/sol.php provides the ITRF00 Cartesian coordinates and velocities for the IGS sites, not all the sites are available on all IGS servers. IGS data organized by GPS week is available at ftp://igscb.jpl.nasa.gov/igscb/product/ and further explanation of IGS data products and formats can be found at http://igscb.jpl.nasa.gov/index.html.

The *Scripps Orbit and Permanent Array Center* (*SOPAC*) is a convenient access point for IGS data. A virtual map of all GPS networks available there can be found at http://sopac.ucsd.edu/maps/. The data archive is available at http://sopac.ucsd.edu/dataArchive/.

PROJECT DESIGN

Horizontal Control

When geodetic surveying was more dependent on optics than electronic signals from space, horizontal control stations were set with station intervisibility in mind, not ease of access. Therefore it is not surprising that they are frequently difficult to reach. Not only are they found on the tops of buildings and mountains, they are also in woods, beside transmission towers, near fences, and generally obstructed from GPS signals. The geodetic surveyors that established them could hardly have foreseen a time when a clear view of the sky above their heads would be crucial to high-quality control.

In fact, it is only recently that most private surveyors have had any routine use for NGS stations. Many station marks have not been occupied for quite a long time. Since the primary monuments are often found deteriorated, overgrown, unstable, or destroyed, it is important that surveyors be well acquainted with the underground marks, R.M.'s, and other methods used to perpetuate control stations.

Obviously, it is a good idea to propose reconnaissance of several more than the absolute minimum of three horizontal control stations. Fewer than three makes any check of their positions virtually impossible. Many more are usually required in a GPS route survey. In general, in GPS networks the more well-chosen horizontal control stations that are available, the better. Some stations will almost certainly prove unsuitable unless they have been used previously in GPS work or are part of a HARN.

Station Location

The location of the stations, relative to the GPS project itself, is also an important consideration in choosing horizontal control. For work other than route surveys, a handy rule of thumb is to divide the project into four quadrants and to choose at least one horizontal control station in each. The actual survey should have at least one horizontal control station in three of the four quadrants. Each of them ought to be as near as possible to the project boundary. Supplementary control in the interior of the network can then be used to add more stability to the network (Figure 6.2).

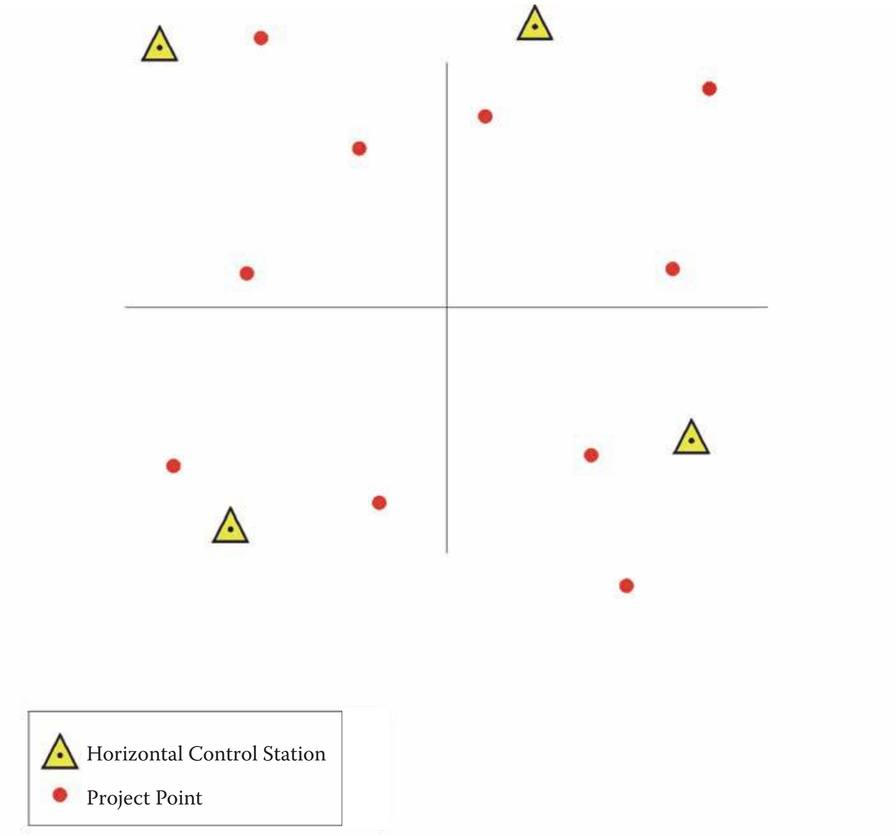
At a minimum route surveys require horizontal control at the beginning, the end, and the middle. Long routes should be bridged with control on both sides of the line at appropriate intervals. The standard symbol for indicating horizontal control on the project map is a triangle.

Vertical Control

Those stations with a published accuracy high enough for consideration as vertical control are symbolized by an open square or circle on the map. Those stations that are sufficient for both horizontal and vertical control are particularly helpful and are designated by a combination of the triangle and square (or circle).

A minimum of four vertical control stations are needed to anchor a GPS network. A large project should have more. In general, the more high-order benchmarks available the better. Vertical control is best located at the four corners of a project.

Orthometric elevations are best transferred by means of classic spirit leveling. When vertical control is too far removed from the project or when the benchmarks are obstructed, if project efficiency is not drastically impaired, such work should be built into the project plan. When the distances involved are too long, two independent GPS measurements may suffice to connect a benchmark to the project. However, it is important to recall the difference between the ellipsoidal heights available from a GPS observation and the orthometric elevations yielded by a level circuit.



FGCC GEOMETRIC CONTROL STANDARDS	GROUP	AA	A	B	C
	ORDER	AA	A	B	1,2-I & II,3
	PPM	0.01	0.1	1.0	10,20,50,100
	BASE (CM)	0.03	0.5	0.8	1 2 3 5
Minimum number of NGRS Horizontal Control Stations		4	3	3	2

FIGURE 6.2 A Survey Design I

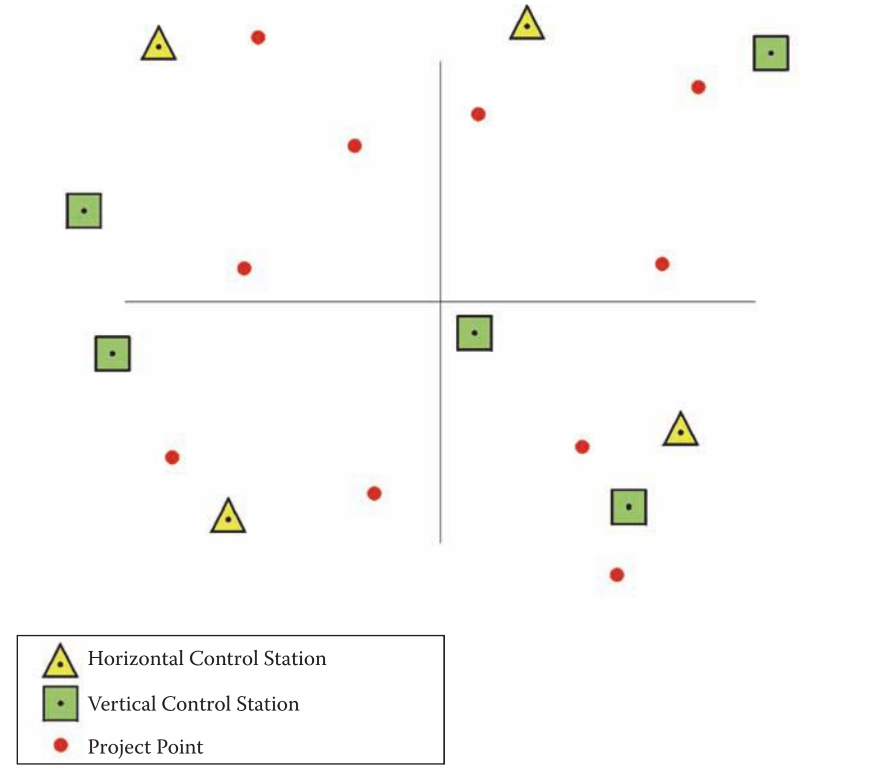
Further, third-order level work is not improved by beginning at a first-order benchmark. When spirit levels are planned to provide vertical control positions, special care may be necessary to ensure that the precision of the conventional work is as consistent as possible with the rest of the GPS survey (Figure 6.3).

Route surveys require vertical control at the beginning and the end. They should be bridged with benchmarks on both sides of the line at intervals from 5 to 10 km.

PREPARATION

Plotting Project Points

A solid dot is the standard symbol used to indicate the position of project points. Some variation is used when a distinction must be drawn between those points that



FGCC GEOMETRIC CONTROL STANDARDS	GROUP	AA	A	B	C
	ORDER	AA	A	B	1,2-I & II,3
	PPM	0.01	0.1	1.0	10,20,50,100
	BASE (CM)	0.03	0.5	0.8	1 2 3 5
Minimum number of NGRS Horizontal Control Stations		4	3	3	2
Minimum number of NGRS Vertical Control Stations		5	5	5	4

FIGURE 6.3 A Survey Design II

are in place and those that must be set. When its location is appropriate, it is always a good idea to have a vertical or horizontal control station serve double duty as a project point. While the precision of their plotting may vary, it is important that project points be located as precisely as possible even at this preliminary stage.

The subsequent observation schedule will depend to some degree on the arrangement of the baselines drawn on the map to connect the plotted points. Also, the preliminary evaluation of access, obstructions, and other information depends on the position of the project point relative to these features.

Evaluating Access

When all potential control and project positions have been plotted and given a unique identifier, some aspects of the survey can be addressed a bit more specifically. If good roads are favorably located, if open areas are indicated around the stations, and if no station falls in an area where special permission will be required for its occupation, then the preliminary plan of the survey ought to be remarkably trouble-free. However, it is likely that one or more of these conditions will not be so fortunately arranged.

The speed and efficiency of transportation from station to station can be assessed to some degree from the project map. It is also wise to remember that while inclement weather does not disturb GPS observations whatsoever, without sufficient preparation it can play havoc with surveyors' ability to reach points over difficult roads or by aircraft.

Planning Offsets

If control stations or project points are located in areas where the map indicates that topography or vegetation will obstruct the satellite's signals, alternatives may be considered. A shift of the position of a project point into a clear area may be possible where the change does not have a significant effect on the overall network. A control station may also be the basis for a less obstructed position, transferred with a short level circuit or traverse. Of course, such a transfer requires availability of conventional surveying equipment on the project. In situations where such movement is not possible, careful consideration of the actual paths of the satellites at the station itself during on-site reconnaissance may reveal enough windows in the gaps between obstructions to collect sufficient data by strictly defining the observation sessions.

Planning Azimuth Marks

Azimuth marks are a common requirement in GPS projects. They are almost always a necessary accompaniment to static GPS stations when a client intends to use them to control subsequent conventional surveying work. Of course, the line between the station and the azimuth mark should be as long as convenience and the preservation of line-of-sight allows.

It is wise to take care that short baselines do not degrade the overall integrity of the project. Occupations of the station and its azimuth mark should be simultaneous for a direct measurement of the baseline between them. Both should also be tied to the larger network as independent stations. There should be two or more occupations of each station when the distance between them is less than 2 km.

While an alternative approach may be to derive the azimuth between a GPS station and its azimuth mark with an astronomic observation, it is important to remember that a small error, attributable to the deflection of the vertical, will be present in such an observation. The small angle between the plumb line and a normal to the ellipsoid at the station can either be ignored or removed with a Laplace correction.

Obtaining Permissions

Another aspect of access can be considered when the project map finally shows all the pertinent points. Nothing can bring a well-planned survey to a halt faster than a locked gate, an irate landowner, or a government official that is convinced he should have been consulted previously. To the extent that it is possible from the available mapping, affected private landowners and government jurisdictions should be identified and contacted. Taking this precaution at the earliest stage of the survey planning can increase the chance that the sometimes long process of obtaining permissions, gate keys, badges, or other credentials has a better chance of completion before the survey begins.

Any aspect of a GPS survey plan derived from examining mapping, virtual or hardcopy, must be considered preliminary. Most features change with time, and even those that are relatively constant cannot be portrayed on a map with complete exactitude. Nevertheless, steps toward a coherent workable design can be taken using the information they provide.

SOME GPS SURVEY DESIGN FACTS

Though much of the preliminary work in producing the plan of a GPS survey is a matter of estimation, some hard facts must be considered, too. For example, the number of GPS receivers available for the work and the number of satellites above the observer's horizon at a given time in a given place are two ingredients that can be determined with some certainty.

Software Assistance

Most GPS software packages provide users with routines that help them determine the satellite *windows*, the periods of time when the largest numbers of satellites are simultaneously available. Now that the GPS system is operational and a full constellation of satellites are on orbit, observers are virtually assured of 24-hour coverage. This assurance is a welcome relief from the forced downtime in the early days of GPS. However, the mere presence of adequate satellites above an observer's horizon does not guarantee collection of sufficient data. Therefore, despite the virtual certainty that at least 4 satellites will be available, evaluation of their configuration as expressed in the position dilution of precision (PDOP) is still crucial in planning a GPS survey.

Position Dilution of Precision (PDOP)

In GPS, the receiver's position is derived from the simultaneous solution of vectors between it and at least 4 satellites. The quality of that solution depends, in large part, on the distribution of the vectors. For example, any position determined when the satellites are crowded together in one part of the sky will be unreliable, because all the vectors will have virtually the same direction. Given the ephemeris of each satellite, the approximate position of the receiver, and the time of the planned observation, a computer can predict such an unfavorable configuration and indicate the problem by giving the PDOP a large number. The GPS survey planner, on notice that the PDOP is large for a particular period of time, should consider an alternate observation plan.

On the other hand, when one satellite is directly above the receiver and three others are near the horizon and 120° in azimuth from one another, the arrangement is nearly ideal for a 4-satellite constellation. The planner of the survey would be likely to consider such a window. However, more satellites would improve the resulting position even more, as long as they are well distributed in the sky above the receiver. In general, the more satellites, the better. For example, if the planner finds 8 satellites will be above the horizon in the region where the work is to be done and the PDOP is below 2, that window would be a likely candidate for observation.

There are other important considerations. The satellites are constantly moving in relation to the receiver and to each other. Satellites rise and set and the PDOP is constantly changing. Within all this movement, the GPS survey designer must have some way of correlating the longest and most important baselines with the longest windows, the most satellites, and the lowest PDOP. Most GPS software packages, given a particular location and period of time, can provide illustrations of the satellite configuration.

Polar Plot

One such diagram is a plot of the satellite's tracks drawn on a graphical representation of the upper half of the celestial sphere with the observer's zenith at the center and perimeter circle as the horizon. The azimuths and elevations of the satellites above the specified mask angle are connected into arcs that represent the paths of all available satellites. The utility of this sort of drawing has lessened with the completion of the GPS constellation. In fact, there are so many satellites available that the picture can become quite crowded and difficult to decipher.

Another printout is a tabular list of the elevation and azimuth of each satellite at time intervals selected by the user.

An Example

The position of point Morant in Table 6.1 needed expression to the nearest minute only, a sufficient approximation for the purpose. The ephemeris data were 5 days old when the chart was generated by the computer, but the data were still an adequate representation of the satellite's movements to use in planning. The mask angle was specified at 15°, so the program would consider a satellite set when it moved below that elevation angle. The zone time was Pacific Daylight Time, 7 hours behind Coordinated Universal Time, UTC. The full constellation provided 24 healthy satellites, and the sampling rate indicated that the azimuth and elevation of those above the mask angle would be shown every 10 minutes.

At 0:00 hour satellite pseudorandom noise (PRN) 2 could be found on an azimuth of 219° and an elevation of 16° above the horizon by an observer at 36°45'Nφ and 121°45'Wλ. The table indicates that PRN 2 was rising, and got continually higher in the sky for the 2 hours and 10 minutes covered by the chart. The satellite PRN 16 was also rising at 0:00 but reached its maximum altitude at about 1:10 and began to set. Unlike PRN 2, PRN 16 was not tabulated in the same row throughout the chart. It was supplanted when PRN 7 rose above the mask angle and PRN 16 shifted one column to the right. The same may be said of PRN 18 and PRN 19. Both of these satellites began high in the sky, unlike PRN 28 and PRN 29. They were just above 15° and setting when the table began and set after approximately 1 hour of availability. They would not have been seen again at this location for about 12 hours.

TABLE 6.1
Satellite Azimuth and Elevation Table I

Point: *Morant* **Lat** *36:45:0 N* **Lon** *121:45:0W* **Ephemeris:** *9/24/93*

Date: *Wed., Sept. 29, 1993* **Mask Angle:** *15 (deg)* **Zone:** *Time Pacific Day (-7)*

24 Satellites: *1 2 3 7 9 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 31*

Sampling Rate: *10 minutes*

Time	El	Az	El	Az	El	Az	El	Az	El	Az	El	Az	El	Az	El	Az	PDOP
							Constellation of 8 SVs										
SV	2		16		18		19		27		28		29		31		
0:00	16	219	15	317	77	121	66	330	41	287	23	65	36	129	30	109	1.7
0:10	20	221	18	314	73	131	67	341	44	292	22	60	32	132	33	104	1.8
0:20	24	223	20	310	68	137	68	353	47	297	21	56	28	135	35	99	1.8
0:30	28	226	22	306	64	142	68	5	50	302	20	51	24	138	36	93	1.9
0:40	32	229	23	302	59	146	67	17	52	308	18	48	20	140	37	88	1.8
0:50	36	232	24	297	54	148	66	28	55	314	16	44	16	142	38	82	1.8
							Constellation of 6 SVs										
SV	2		16		18		19		27		31						
1:00	40	235	24	293	49	151	65	39	58	320	38	76					3.0
1:10	43	239	24	288	44	153	63	49	61	328	37	70					3.0
1:20	47	244	24	283	40	155	61	57	64	336	36	64					2.8
1:30	51	249	23	278	35	156	59	65	66	345	34	60					2.6
							Constellation of 7 SVs										
SV	2		7		16		18		19		27		31				
1:40	54	254	16	186	22	273	30	157	56	73	68	356	32	55			2.3
1:50	57	260	21	186	20	269	26	158	53	79	70	9	29	51			2.2
2:00	60	268	25	186	19	264	22	159	50	85	71	23	26	48			2.0
							Constellation of 8 SVs										
SV	2		7		16		18		19		26		27		31		
2:10	66	276	30	185	16	260	17	160	47	91	15	319	71	38	23	45	1.7

This chart indicated changes in the available constellation from eight space vehicles (SVs), between 0:00 and 0:50, six between 1:00 and 1:30, seven from 1:40 to 2:00, and back to eight at 2:10. The constellation never dipped below the minimum of 4 satellites, and the PDOP was good throughout. The PDOP varied between a low of 1.7 and a high of 3.0. Over the interval covered by the table, the PDOP never reached the unsatisfactory level of 5 or 6 which is when a planner should avoid observation.

Choosing the Window

Using this chart, the GPS survey designer might well have concluded that the best available window was the first. There was nearly an hour of 8-satellite data with a PDOP below 2. However, the data indicated that good observations could be made at any time covered here, except for one thing: it was the middle of the night. When a small number of satellites were available in the early days of GPS, the discomfort of such observations were ignored from necessity. With a full constellation, the loss of sleep can be avoided, and the designer may look at a more convenient time of day to begin the field work.

Ionospheric Delay

It is worth noting that the ionospheric error is usually smaller after sundown. In fact, the FGCC specifies two-frequency receivers for daylight observations that hope to meet AA-, A-, and B-order accuracy standards, due, in part, to the increased ionospheric delay during those hours. There are provisions for compensation by modeling the error with two-frequency data from other sources where only single-frequency receivers are available. However, the specification illustrates the importance of considering atmospheric error sources.

An Example

Table 6.2, shows later in the day, covers a period of two hours when a constellation of 5 and 6 satellites was always available. However, through the first hour, from 6:30 to 7:30, the PDOP hovered around 5 and 6. For the first half of that hour, four of the satellites—PRN 9, PRN 12, PRN 13, and PRN 24—were all near the same elevation. During the same period, PRN 9 and PRN 12 were only approximately 50° apart in azimuth as well. Even though a sufficient constellation of satellites was constantly available, the survey designer may well have considered only the last 30 to 50 minutes of the time covered by this chart as suitable for observation.

There is one caution, however. Azimuth-elevation tables are a convenient tool in the division of the observing day into sessions, but it should not be taken for granted that every satellite listed is healthy and in service. For the actual availability of satellites and an update on atmospheric conditions, it is always wise to call the recorded message on the U.S. Coast Guard hotline at (703) 313-5907 or online you can check *GPS Status Message* at http://www.navcen.uscg.gov/ftp/GPS/status.txt before and after a project. In the planning stage, the call can prevent creation of a design dependent on satellites that prove unavailable. Similarly, after the field work is completed, it can prevent inclusion of unhealthy data in the postprocessing.

TABLE 6.2
Satellite Azimuth and Elevation Table II

Point: *Morant* Lat *36:45:0 N* Lon *121:45:0W* Ephemeris: *9/24/93*
Date: *Wed., Sept. 29, 1993* Mask Angle: *15 (deg)* Zone: *Time Pacific Day (-7)*
24 Satellites: *1 2 3 7 9 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 31*
Sampling Rate: *10 minutes*

Time	El Az	El Az	El Az	El Az	El Az	El Az	El Az	El Az	PDOP
				Constellation of 5 SVs					
SV	7	9	12	13	24				
6:30	28 54	60 271	61 319	62 15	48 177				6.3
6:40	24 57	60 261	66 314	57 19	53 176				6.0
6:50	21 60	59 252	70 305	53 22	58 175				5.3
7:00	18 62	57 243	73 292	49 25	63 172				4.6
				Constellation of 5 SVs					
SV	9	12	13	20	24				
7:10	54 235	74 274	44 28	16 308	68 169				4.8
7:20	51 229	74 255	40 32	20 310	72 163				5.7
7:30	47 224	72 238	37 35	23 311	77 153				5.1
7:40	43 219	68 226	33 38	27 313	80 134				4.0
				Constellation of 6 SVs					
SV	9	12	13	16	20	24			
7:50	39 215	64 218	29 41	16 149	31 314	81 102			2.1
8:00	35 212	59 213	26 45	19 146	36 314	80 73			2.3
8:10	31 209	54 209	23 48	23 143	40 315	76 57			2.4
8:20	27 207	49 206	19 52	27 140	44 314	72 49			2.5
8:30	23 204	44 204	16 55	30 137	48 314	67 45			2.5

Supposing that the period from 7:40 to 8:30 was found to be a good window, the planner may have regarded it as a single 50-minute session, or divided it into shorter sessions. One aspect of that decision was probably the length of the baseline in question. In static GPS, a long line of 30 km may require 50 minutes of 6-satellite data, but a short line of 3 km may not. If the planned survey was not done by static GPS, but instead with rapid-static, a 10-minute session may have been sufficient. Therefore, another aspect of the decision as to how the window was divided probably depended on the anticipated GPS surveying technique. A third consideration was probably the approximation of the time necessary to move from one station to another.

Naming the Variables

The next step in the GPS survey design is drawing the preliminary plan of the baselines on the project map. Once some idea of the configuration of the baselines has been established, an observation schedule can be organized. Toward that end, the FGCC has developed a set of formulas provided in appendix F of their provisional *Geometric Geodetic Accuracy Standards and Specifications for Using GPS Relative Positioning Techniques*. Those formulas will be used here.

For illustration, suppose that the project map (Figure 6.4) includes horizontal control, vertical control, and project points for a planned GPS network. They will be symbolized by m. There are four dual-frequency GPS receivers available for this project. They will be symbolized by r. There will be five observation sessions each day during the project. They will be symbolized by d. To summarize:

$$m = \text{total number of stations (existing and new)} = 14$$

$$d = \text{number of possible observing sessions per observing day} = 5$$

$$r = \text{number of receivers} = 4 \text{ dual frequency}$$

The design developed from this map must be preliminary. The session for each day of observation will depend on the success of the work the day before. Please recall that the plan must be provisional until the baseline lengths, the obstructions at the observation sites, the transportation difficulties, the ionospheric disturbances, and the satellite geometry are actually known. Those questions can only be answered during the reconnaissance and the observations that follow. Even though these equivocations apply, the next step is to draw the baseline's measurement plan.

DRAWING THE BASELINES

Horizontal Control

A good rule of thumb is to verify the integrity of the horizontal control by observing baselines between these stations first. The vectors can be used to both corroborate the accuracy of the published coordinates and later to resolve the scale, shift, and rotation parameters between the control positions and the new network that will be determined by GPS.

These baselines are frequently the longest in the project, and there is an added benefit to measuring them first. By processing a portion of the data collected on the

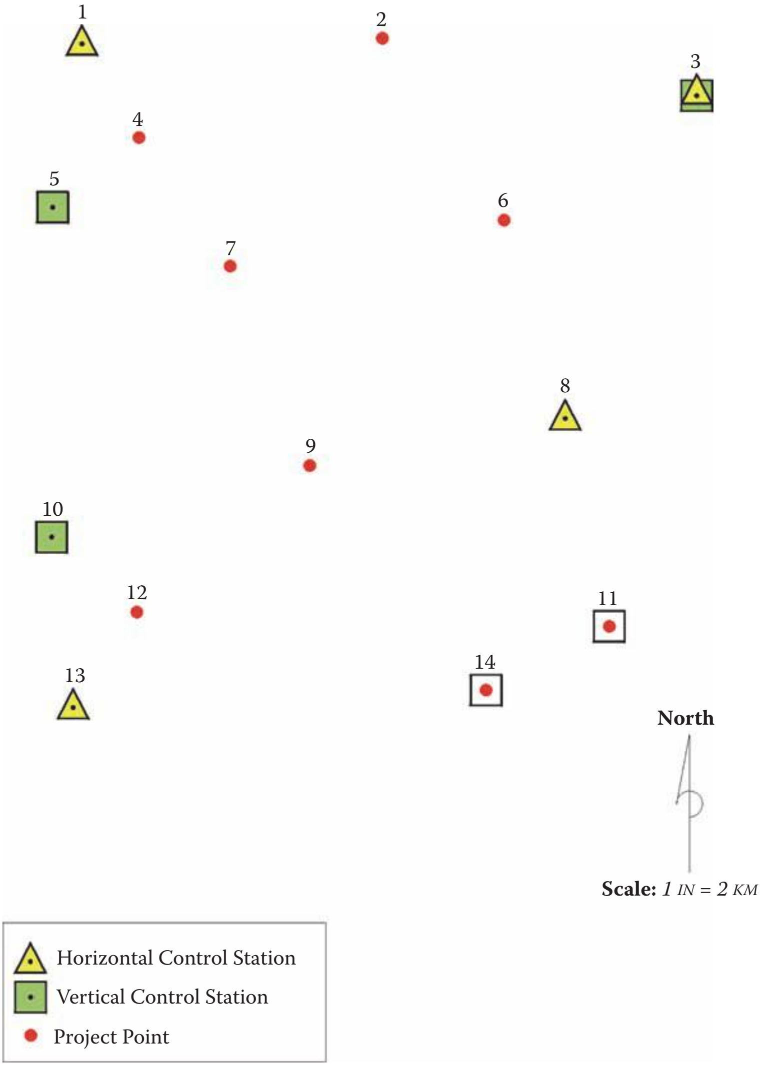
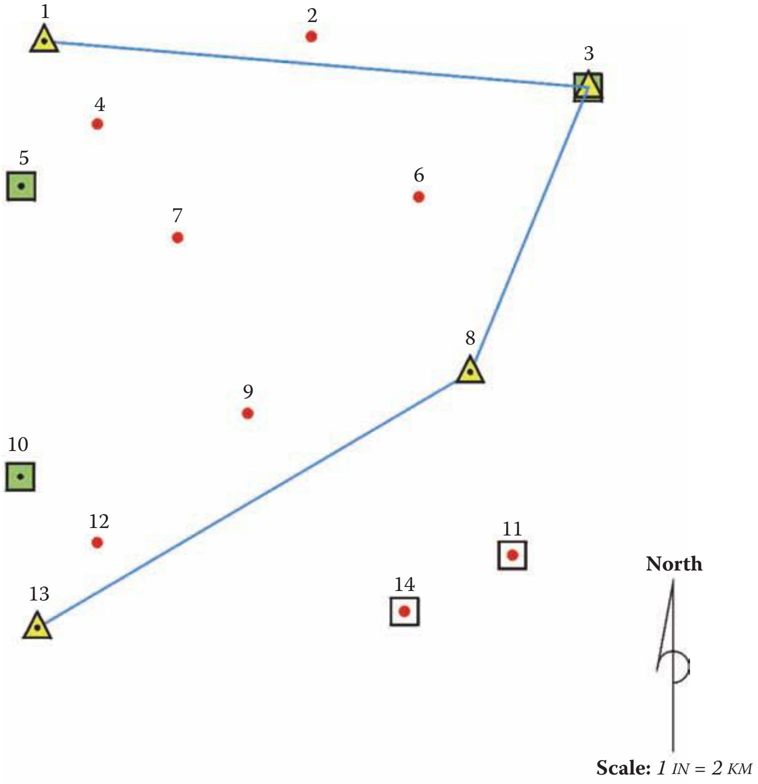


FIGURE 6.4 A Survey Design III

longest baselines early in the project, the degree that the sessions could have been shortened without degrading the quality of the measurement can be found. This test may allow improvement in the productivity on the job without erosion of the final positions (Figure 6.5).

Julian Day in Naming Sessions

The table at the bottom of Figure 6.5 indicates that the name of the first session connecting the horizontal control is 49-1. The date of the planned session is given in the Julian system. Taken most literally, Julian dates are counted from January 1, 4713 B.C. However, most practitioners of GPS use the term to mean the day of



Session	Receivers				Independent Lines	Trivial Lines
	A	B	C	D		
49-1	1	3	8	13	1-3 3-8 8-13	13-1 1-8 3-13

Session	Receivers				Independent Lines	Trivial Lines
	A	B	C	D		

FIGURE 6.5 Drawing the Baselines

the current year measured consecutively from January 1. Under this construction, since there are 31 days in January, Julian day 49 is February 18 of the current year. The designation 49-1 means that this is to be the first session on that day. Some prefer to use letters to distinguish the session. In that case, the label would be 49-A.

Independent Lines

This project will be done with four receivers. The table shows that receiver A will occupy point 1; receiver B, point 3; receiver C, point 8; and receiver D, point 13 in the first session. However, the illustration shows only three of the possible six base lines that will be produced by this arrangement. Only the *independent*, also known as *nontrivial*, lines are shown on the map. The three lines that are not drawn are called *trivial*, and are also known as *dependent lines*. This idea is based on restricting the use of the lines created in each observing session to the absolute minimum needed to produce a unique solution.

Whenever four receivers are used, six lines are created. However, any three of those lines will fully define the position of each occupied station in relation to the others in the session.

Therefore, the user can consider any three of the six lines independent. But once the decision is made only those three baselines are included in the network. The remaining baselines are then considered trivial and discarded. In practice, the three shortest lines in a four-receiver session are almost always deemed the independent vectors, and the three longest lines are eliminated as trivial, or dependent. That is the case with the session illustrated.

Where *r* is the number of receivers, every session yields *r*-1 independent baselines. For example, four receivers used in 10 sessions would produce 30 independent baselines. It cannot be said that the shortest lines are always chosen to be the independent lines. Sometimes there are reasons to reject one of the shorter vectors due to incomplete data, cycle slips, multipath, or some other weakness in the measurements. Before such decisions can be made, each session will require analysis after the data has actually been collected. In the planning stage, it is best to consider the shortest vectors as the independent lines.

Another aspect of the distinction between independent and trivial lines involves the concept of error of closure, or loop closure. Loop closure is a procedure by which the internal consistency of a GPS network is discovered. A series of baseline vector components from more than one GPS session, forming a loop or closed figure, is added together. The closure error is the ratio of the length of the line representing the combined errors of all the vector's components to the length of the perimeter of the figure. Any loop closures that only use baselines derived from a single common GPS session will yield an apparent error of zero, because they are derived from the same simultaneous observations. For example, all the baselines between the four receivers in session 49-1 of the illustrated project will be based on ranges to the same GPS satellites over the same period of time. Therefore, the trivial lines of 13-1, 1-8, and 3-13 will be derived from the same information used to determine the independent lines of 1-3, 3-8, and 8-13. It follows that, if the fourth line from station

13 to station 1 were included to close the figure of the illustrated session, the error of closure would be zero. The same may be said of the inclusion of any of the trivial lines. Their addition cannot add any redundancy or any geometric strength to the lines of the session, because they are all derived from the same data. If redundancy cannot be added to a GPS session by including any more than the minimum number of independent lines, how can the baselines be checked? Where does redundancy in GPS work come from?

Redundancy

If only two receivers were used to complete the illustrated project, there would be no trivial lines and it might seem there would be no redundancy at all. But to connect every station with its closest neighbor, each station would be have to be occupied at least twice, and each time during a different session. For example, with receiver A on station 1 and receiver B on station 2, the first session could establish the baseline between them. The second session could then be used to measure the baseline between station 1 and station 4. It would certainly be possible to simply move receiver B to station 4 and leave receiver A undisturbed on station 1. However, some redundancy could be added to the work if receiver A were reset. If it were recentered, replumbed, and its H.I. remeasured, some check on both of its occupations on station 1 would be possible when the network was completed. Under this scheme, a loop closure at the end of the project would have some meaning.

If one were to use such a scheme on the illustrated project and connect into one loop all of the 14 baselines determined by the 14 two-receiver sessions, the resulting error of closure would be useful. It could be used to detect blunders in the work, such as mismeasured H.I.s. Such a loop would include many different sessions. The ranges between the satellites and the receivers defining the baselines in such a circuit would be from different constellations at different times. On the other hand, if it were possible to occupy all 14 stations in the illustrated project with 14 different receivers simultaneously and do the entire survey in one session, a loop closure would be absolutely meaningless.

In the real world, such a project is not usually done with 14 receivers nor with 2 receivers, but with 3, 4, or 5. The achievement of redundancy takes a middle road. The number of independent occupations is still an important source of redundancy. In the two-receiver arrangement every line can be independent, but that is not the case when a project is done with any larger number of receivers. As soon as three or more receivers are considered, the discussion of redundant measurement must be restricted to independent baselines, excluding trivial lines.

Redundancy is then partly defined by the number of independent baselines that are measured more than once, as well as by the percentage of stations that are occupied more than once. While it is not possible to repeat a baseline without reoccupying its endpoints, it is possible to reoccupy a large percentage of the stations in a project without repeating a single baseline. These two aspects of redundancy in GPS—the repetition of independent baselines and the reoccupation of stations—are somewhat separate.

Federal Geodetic Control Committee (FGCC) Standards for Redundancy

To meet order AA geometric accuracy standards, the FGCC requires three or more occupations on 80% of the stations in a project. Three or more occupations are necessary on 40%, 20%, and 10% of the stations for A, B, and C standards, respectively. When the distance between a station and its azimuth mark is less than 2 km, both points must be occupied at least twice to meet any standard above second-order. All vertical control stations must be occupied at least twice for all orders of accuracy. Two or more occupations are required for all horizontal control stations in order AA—the percentage requirements for repeat occupations on horizontal control stations drops to 75%, 50%, and 25% for A, B, and C, respectively. For new project points, reoccupation is mandated on 80%, 50%, and 10% of the stations in the project for A, B, and C, respectively.

The standards for repeat measurements of independent baselines in the FGCC provisional specifications note that an equal number of N–S and E–W vectors should be remeasured in a network. Of the independent baselines, 25% should be repeated in a project to meet order AA geometric accuracy standards. The standards require 15%, 5%, and 5% for orders A, B, and C, respectively.

Unless a project is to be *blue-booked*, that is, submitted to the NGS for inclusion in the national network, or there is a contractual obligation, there is usually no need to meet the letter of the specifications listed above. They are offered here as an indication of the level of redundancy that is necessary for high-accuracy GPS work.

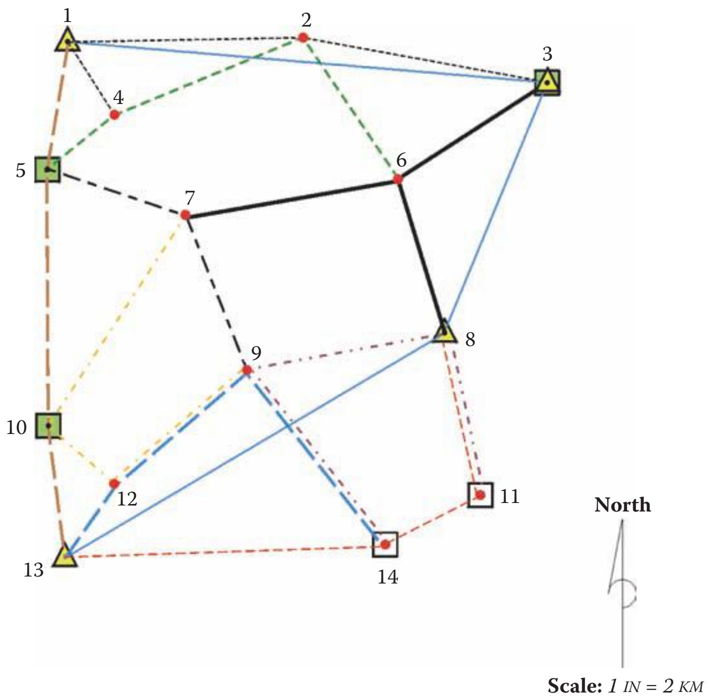
Figure 6.6 shows one of the many possible approaches to setting up the baselines for this particular GPS project. The survey design calls for the horizontal control to be occupied in session 49-1. It is to be followed by measurements between two control stations and the nearest adjacent project points in session 49-2. As shown in the table at the bottom of Figure 6.6, there will be redundant occupations on stations 1 and 3. Even though the same receivers will occupy those points, their operators will be instructed to reset them at a different H.I.s for the new session. A better, but probably less efficient, plan would be to occupy these stations with different receivers than were used in the first session.

Forming Loops

As the baselines are drawn on the project map for a static GPS survey, or any GPS work where accuracy is the primary consideration, the designer should remember that part of their effectiveness depends on the formation of complete geometric figures. When the project is completed, these independent vectors should be capable of formation into closed loops that incorporate baselines from two to four different sessions. In the illustrated baseline plan, no loop contains more than 10 vectors, no loop is more than 100 km long, and every observed baseline will have a place in a closed loop.

Finding the Number of Sessions

The illustrated survey design calls for 10 sessions, but the calculation does not include human error, equipment breakdown, and other unforeseeable difficulties. It would be



Session	Receivers				Independent Lines	Trivial Lines
	A	B	C	D		
49-1	1	3	8	13	1-3 3-8 8-13	13-1 1-8 3-13
49-2	1	3	2	4	1-2 2-3 1-4	1-3 3-4 2-4
49-3	5	6	2	4	5-4 4-2 2-6	5-6 2-5 4-6
49-4	5	6	7	9	5-7 7-9 6-7	5-6 6-9 5-9
49-5	5	1	10	13	1-5 5-10 10-13	1-10 1-13 5-13

Session	Receivers				Independent Lines	Trivial Lines
	A	B	C	D		
50-1	10	12	7	9	7-10 10-12 9-12	7-12 9-10 7-9
50-2	14	12	13	9	9-12 12-13 9-14	13-14 12-14 9-13
50-3	14	8	13	11	13-14 14-11 11-8	13-8 8-14 13-11
50-4	14	8	11	9	14-9 8-9 8-11	9-11 8-14 14-11
50-5	7	6	3	8	6-8 6-7 3-6	3-8 3-7 7-8

FIGURE 6.6 A Static Survey Plan

impractical to presume a completely trouble-free project. The FGCC proposes the following formula for arriving at a more realistic estimate:

$$s = \frac{(m \cdot n)}{r} + \frac{(m \cdot n)(p-1)}{r} + k \cdot m$$

where

s = the number of observing sessions,
r = the number of receivers,
m = the total number of stations involved

But *n*, *p*, and *k* require a bit more explanation. The variable *n* is a representation of the level of redundancy that has been built into the network, based on the number of occupations on each station. The illustrated survey design includes more than two occupations on all but 4 of the 14 stations in the network. In fact, 10 of the 14 positions will be visited three or four times in the course of the survey. There are a total of 40 occupations by the 4 receivers in the 10 planned sessions. By dividing 40 occupations by 14 stations, it can be found that each station will be visited an average of 2.857 times. Therefore in the FGCC formula, the planned redundancy represented by factor *n* is equal to 2.857 in this project.

The experience of a firm is symbolized by the variable *p* in the formula. The division of the final number of actual sessions required to complete past projects by the initial estimation yields a ratio that can be used to improve future predictions. That ratio is the production factor, *p*. A typical production factor is 1.1.

A safety factor of 0.1, known as *k*, is recommended for GPS projects within 100 km of a company's home base. Beyond that radius, an increase to 0.2 is advised.

The substitution of the appropriate quantities for the illustrated project increases the prediction of the number of observation sessions required for its completion:

$$s = \frac{(mn)}{r} + \frac{(mn)(p-1)}{r} + km$$

$$s = \frac{(14)(2.857)}{4} + \frac{(14)(2.857)(1.1-1)}{4} + (0.2)(14)$$

$$s = \frac{40}{4} + \frac{40}{4} + 2.8$$

$$s = 10 + 1 + 2.8$$

s = 14 sessions (rounded to the nearest integer)

In other words, the 2-day, 10-session schedule is a minimum period for the baseline plan drawn on the project map. A more realistic estimate of the observation schedule includes 14 sessions. It is also important to keep in mind that the observation schedule does not include time for on-site reconnaissance.

Ties to the Vertical Control

The ties from the vertical control stations to the overall network are usually not handled by the same methods used with the horizontal control. The first session of the illustrated project was devoted to occupation of all the horizontal control stations. There is no similar method with the vertical control stations. First, the geoidal undulation would be indistinguishable from baseline measurement error. Second, the primary objective in vertical control is for each station to be adequately tied to its closest neighbor in the network.

If a benchmark can serve as a project point, it is nearly always advisable to use it, as was done with stations 11 and 14 in the illustrated project. A conventional level circuit can often be used to transfer a reliable orthometric elevation from vertical control station to a nearby project point.

REAL-TIME KINEMATIC (RTK) AND DIFFERENTIAL GPS (DGPS)

Most, not all, GPS surveying relies on the idea of differential positioning. The mode of a base or reference receiver at a known location logging data at the same time as a receiver at an unknown location together provide the fundamental information for the determination of accurate coordinates. While this basic approach remains today, the majority of GPS surveying is not done in the static postprocessed mode just described. Postprocessing is most often applied to control work. Now, the most commonly used methods utilize receivers on reference stations that provide correction signals to the end user via a data link sometimes over the Internet, radio signal, or cell phone and often in real-time.

In this category of GPS surveying work there is sometimes a distinction made between code-based, DGPS, and carrier based, RTK, solutions. In fact, most systems use a combination of code and carrier measurements so the distinction is more a matter of emphasis rather than an absolute difference.

THE GENERAL IDEA

Errors in satellite clocks, imperfect orbits, the trip through the layers of the atmosphere, and many other sources contribute inaccuracies to GPS signals by the time they reach a receiver.

These errors are variable, so the best to way to correct them is to monitor them as they happen. A good way to do this is to set up a GPS receiver on a station whose position is known exactly, a base station. This base station receiver's computer can calculate its position from satellite data, compare that position with its actual known position, and find the difference and presto, error corrections. It works well, but the errors are constantly changing so a base station has to monitor them all the time, at least all the time the rover receiver or receivers are working. While this is happening the rovers move from place to place collecting the points whose positions you want to know relative to the base station, which is the real objective after all. Then all you have to do is get those base station corrections and the rover's data together somehow. That combination can be done over a data link in real-time, or applied later in postprocessing.

Both RTK and DGPS have been built on the foundation of the idea that, with the important exceptions of multipath and receiver noise, GPS error sources are correlated. In other words, the closer the rover is to the base the more the errors at the ends of the baseline match. The shorter the baseline, the more the errors are correlated. The longer the baseline, the less the errors are correlated.

RADIAL GPS

In both RTK and DGPS surveying radial GPS has become the typical surveying style. There are advantages to the approach (Figure 6.7). The advantage is a large number of positions can be established in a short amount of time with little or no planning. The disadvantage is that there is little or no redundancy in positions derived from this approach since all the baselines originate from the same control station. Redundancy can be incorporated, but it requires repetition of the observations. One way to do it is to occupy the project points, the unknown positions, successively with more than one rover. It is best if these successive occupations are separated by at least 4 hours and not more than 8 hours so the satellite constellation can reach a significantly different configuration.

Another more convenient but less desirable approach is to do a second occupation almost immediately after the first. The roving receiver's antenna is blocked or tilted until the lock on the satellites is interrupted. It is then reoriented on the unknown position a second time for the repeat solution. This does offer a second solution, but from virtually the same constellation.

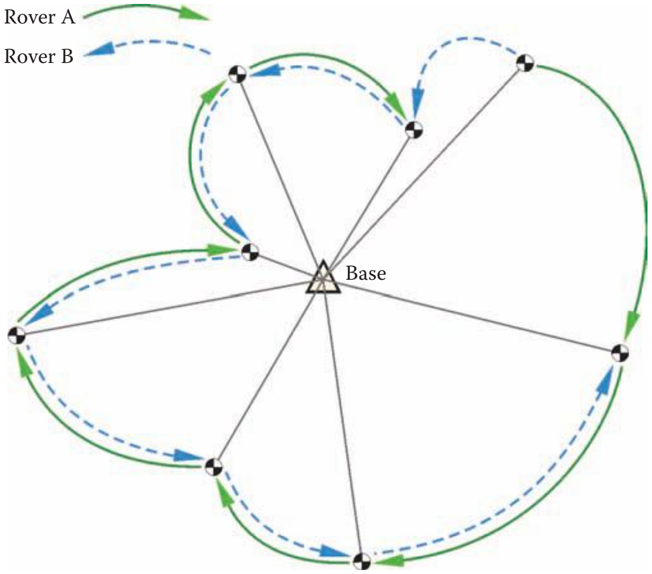


FIGURE 6.7 Radial GPS

A third way to achieve redundancy is to occupy each point with the same rover but utilizing a different base station. This approach allows a solution to be available from two separate control stations. Obviously, this can be done with reoccupation of the project points after one base station has been moved to a new control point, or a two base stations can be up and running from the very outset and throughout the work, as would be the case using two CORS stations. It is best if there are both two occupations on each point and each of the two utilize different base stations.

More efficiency can be achieved by adding additional roving receivers. However, as the number of receivers rises, the logistics become more complicated, and a survey plan becomes necessary. Also, project points that are simultaneously near one another but far from the control station should be directly connected with a baseline to maintain the integrity of the survey. Finally, if the base receiver loses lock and it goes unnoticed, it will completely defeat the radial survey for the time it is down.

The Correction Signal

The agreed upon format first designed for communication between the base station and rovers used in marine navigation is known as *Radio Technical Commission for Maritime Services* (*RTCM*)**.** In 1985, *the, RTCM Special Committee* (*SC-104*) created a standard that is still more used than any proprietary formats that have come along since. It was originally designed to accommodate a slow data rate with a configuration somewhat similar to the navigation message. It is important to note that RTCM is *open*; in other words, it is a general purpose format and is not restricted to a particular receiver type. It works across platforms and the pseudorange correction is made up of a sequence of different message types.

DGPS

The term *DGPS* is most often used to refer to differential GPS that is based on pseudoranges, aka code phase (Figure 6.8). Even though the accuracy of code phase applications was given a boost with the elimination of Selective Availability (SA) in May 2000 consistent accuracy better than 5 meters or so still requires reduction of the effect of correlated ephemeris and atmospheric errors by differential corrections. Though the corrections could be applied in postmission processing services that supply these corrections, most often operate in real-time.

REAL-TIME

Usually, pseudorange corrections, rather than coordinate corrections, are broadcast from the base to the rover or rovers for each satellite in the visible constellation. Rovers with an appropriate *input/output* (*I/O*) port that can receive the correction signal and calculate coordinates. The real-time signal comes to the receiver over a data link. It can originate at a project specific base station or it can come to the user through a service of which there are various categories. Some are open to all users and some are by subscription only. Coverage depends on the spacing of the beacons, aka transmitting base stations, their power, interference, and so forth. Some systems require two-way, some one-way, communication with the base stations. Radio

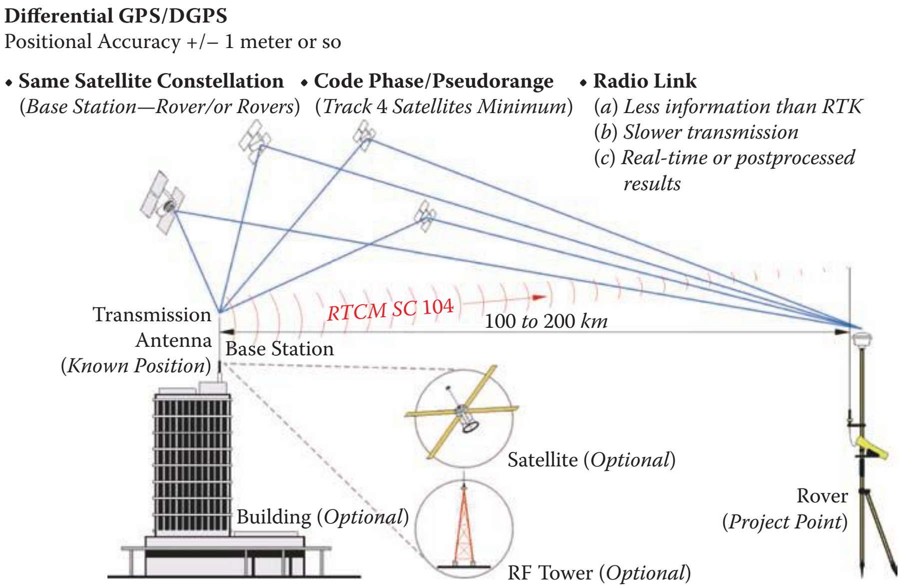


FIGURE 6.8 DGPS

systems, geostationary satellites, low-earth-orbiting satellites and cellular phones are some of the options available for two-way data communication. In any case, most of the wide variety of DGPS services were not originally set up to augment surveying and mapping applications of GPS; they were established to aid GPS navigation.

LOCAL AND WIDE AREA DPGS

As mentioned earlier the correlation between most of the GPS biases becomes weaker as the rover gets farther from the base. The term *Local Area Differential GPS* (*LADGPS*), is used when the baselines from a single base station to the roving receivers using the service are less than a couple of hundred kilometers.

The term *Wide Area Differential GPS* (*WADGPS*) is used when the service uses a network of base stations and distributes correction over a larger area, an area that may even be continental in scope. Many bases operating together provide a means by which the information from several of them can be combined to send a normalized or averaged correction tailored to the rover's geographical position within the system. Some use satellites to provide the data link between the service provider and the subscribers. Such a system depends on the network of base stations receiving signals from the GPS satellites and then streaming that data to a central computer at a control center. There the corrections are calculated and uploaded to a geo-stationary communication satellite. Then the communication satellite broadcasts the corrections to the service's subscribers.

In all cases, the base stations are at known locations and their corrections are broadcast to all rovers that are equipped to receive their particular radio message carrying real-time corrections in the RTCM format. An example of such a DGPS service originated as an augmentation for marine navigation.

Maritime DGPS

Both the *U.S. Coast Guard* (*USCG*) and the *Canadian Coast Guard* (*CCG*) instituted DGPS services to facilitate harbor entrances, ocean mapping, and marine traffic control as well as navigation in inland waterways. Their system base stations beacons broadcast GPS corrections along major rivers, major lakes, the east coast, and the west coast. The sites use marine beacon frequencies of 255 to 325 kHz, which has the advantage of long range propagation that can be several hundreds of kilometers. Access to the broadcasts is free and over recent years the service has become very popular outside of its maritime applications, particularly among farmers engaged in GPS aided precision agriculture. Therefore, the system has been extended beyond waterways across the continental U.S. and is now known as the *Nationwide DGPS (NDGPS*). There are currently 86 base stations. Of these 39 are USCG sites, 38 are *Department of Transportation* (*DOT*) sites, and 9 are *Corps of Engineers* sites.

Wide Area Augmentation System (WAAS)

Another U.S. DGPS service initiated in 1994 cooperatively by the *Department of Transportation* and the *Federal Aviation Administration* (*FAA*) is known as WAAS (Figure 6.9). It is available to users with GPS receivers equipped to receive it. The signal is free and its reliability is excellent. While the official horizontal accuracy is 7.6 m, its capability to actually deliver approximately 1 m horizontally makes it a possible alternative to Wide Area DGPS. It utilizes both satellite-based, also known as *SBAS,* and ground-based augmentations and was initially designed to assist aerial navigation from takeoff, en route through landing. Reference stations on the ground send their data via processing sites to two Ground Earth Stations that upload differential corrections and time to geostationary satellites, *Inmarsat III's*, devoted to transmission of GPS differential corrections to users.

Another WAAS is planned for the *European Geostationary Navigation Overlay Service* (*EGNOS*). This system will augment GPS and GLONASS using three geostationary and a network of ground stations. The Japanese are planning a similar WAAS known by the acronym *MSAS*.

LATENCY

It takes some time for the base station to calculate corrections, and it takes some time for it to put the data into packets in the correct format and transmit them. Then the data makes its way from the base station to the rover over the data link. It is decoded and must go through the rover's software. The time this takes is called the *latency* of the communication between the base station and the rover**.** It can be as little as a quarter of a second or as long as a couple of seconds. And since the base stations corrections are only accurate for the moment they were created, the base station must send a range rate correction along with them. Using this rate correction, the rover can *back date* the correction to match the moment it made that same observation.

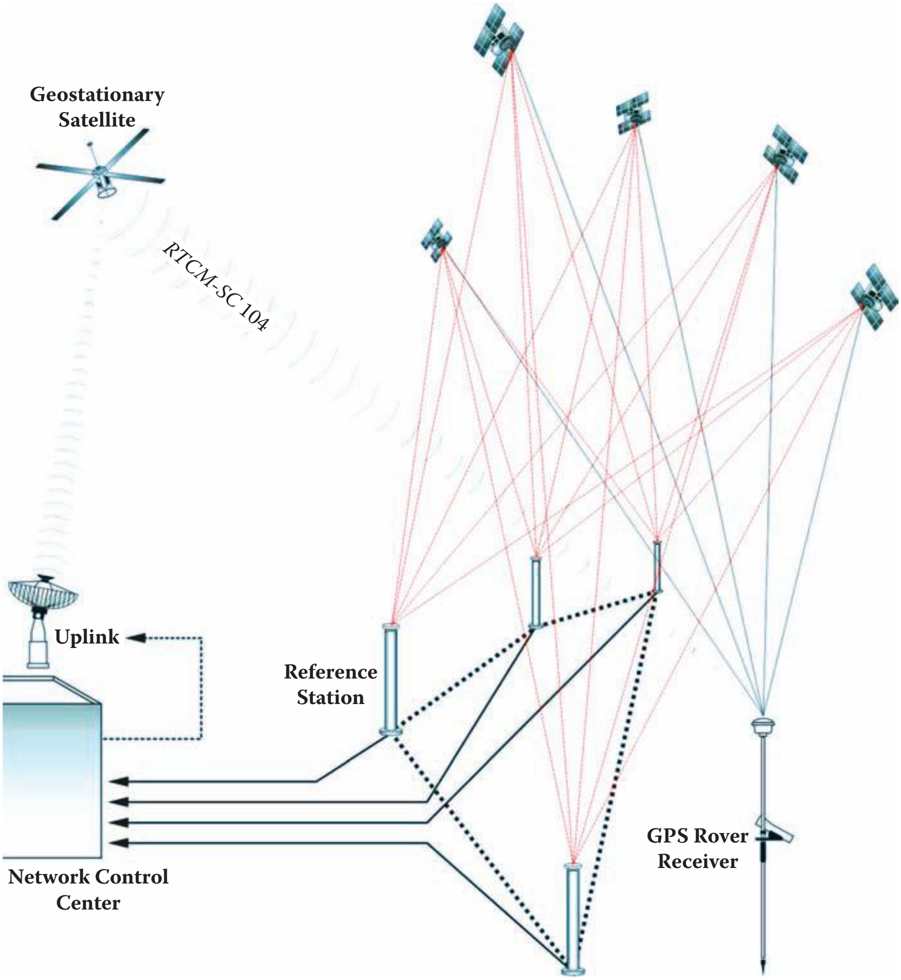


FIGURE 6.9 Wide Area Differential GPS

IDENTICAL CONSTELLATION

DGPS requires that all receivers collect pseudoranges from the same constellation of satellites. It is vital that the errors corrected by the base station are common to the rovers. The rover must share its selection of satellites with the base station; otherwise it would be necessary to create differential corrections for all the combinations of all the available satellites. That could get unmanageable in a hurry; for example, with just 4 satellites above the observer's horizon there can be more than 80 such combinations.

GEOGRAPHIC INFORMATION SYSTEMS (GIS) APPLICATIONS FOR DGPS

Aerial navigation, marine navigation, agriculture, vehicle tracking, and construction are all now using DGPS. DGPS is also useful in land and hydrographic surveying, but perhaps the fastest growing application for DGPS is in data collection, data updating, and even in-field mapping for Geographic Information Systems (GIS).

GIS data has long been captured from paper records such as digitizing and scanning paper maps. Photogrammetry, remote sensing, and conventional surveying has also been data sources for GIS. More recently, data collected in the field with DGPS has become significant in GIS.

GIS data collection with DGPS requires the integration of the position of features of interest and relevant attribute information about those features. Whether a handheld data logger, an electronic notebook, or a pen computer are used, the attributes to be collected are defined by the data dictionary designed for the particular GIS.

In GIS it is frequently important to return to a particular site or feature to perform inspections or maintenance. DGPS with real-time correction makes it convenient to load the position or positions of features into a data logger, and navigate back to the vicinity. But to make such applications feasible, a GIS must be kept current. It must be maintained. A receiver configuration including real-time DGPS, sufficient data storage, and graphic display allows easy verification and updating of existing information.

DGPS allows the immediate attribution and validation in the field with accurate and efficient recording of position. In the past, many GIS mapping efforts have often relied on ties to street centerlines, curb-lines, railroads, and so forth. Such dependencies can be destroyed by demolition or new construction. But, meter level positional accuracy even in obstructed environments such as urban areas, amid high-rise buildings, is possible with DGPS. In other words, with DGPS the control points are the satellites themselves. Therefore it can provide reliable positioning even if the landscape has changed. And its data can be integrated with other technologies, such as laser range-finders, and so forth, in environments where DGPS is not ideally suited to the situation.

Finally, loading GPS data into a GIS platform does not require manual intervention. GPS data processing can be automated; the results are digital and can pass into a GIS format without redundant effort, reducing the chance for errors.

RTK

Kinematic surveying, also known as stop-and-go kinematic surveying, is not new. The original kinematic GPS innovator, Dr. Benjamin Remondi, developed the idea in the mid-1980s. RTK is a method that can offer positional accuracy nearly as good as static carrier phase positioning, but RTK does it in real-time (Figure 6.10). Today, RTK has become routine in development and engineering surveys where the distance between the base and roving receivers can most often be measured in thousands of feet. RTK is capable of delivering accuracy within a few centimeters. RTK is a GPS method that definitely uses carrier phase observations corrected in real-time, and therefore it depends on the fixing of the integer cycle ambiguity.

Fixing the Integer Ambiguity in RTK

The earliest processing software was capable of making C/A-code pseudorange, L1 carrier phase, and usually half wavelength L2 carrier phase measurements. These

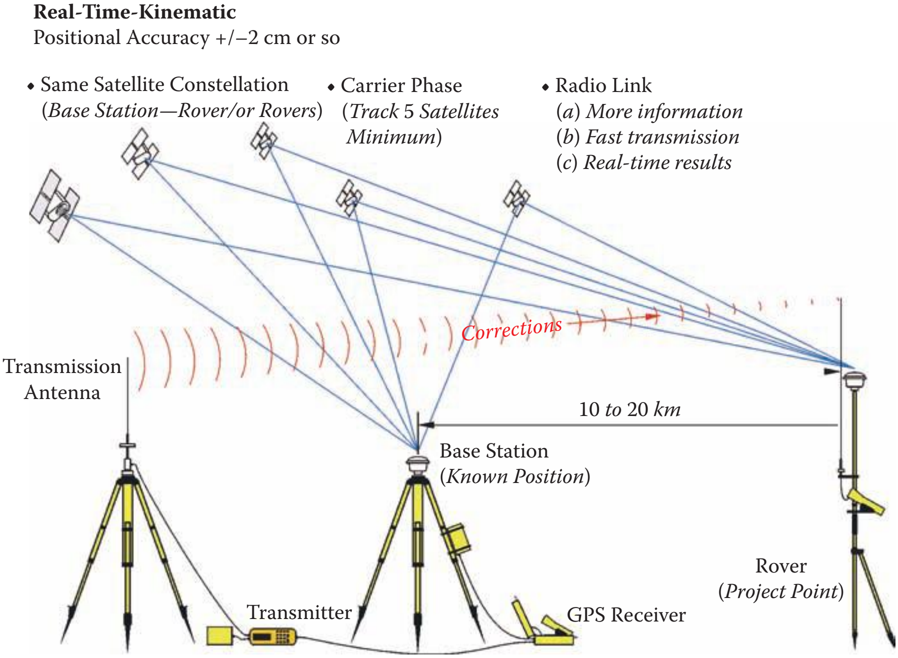


FIGURE 6.10 Real-Time Kinematic GPS

techniques tended to require long static occupations up to an hour or more. With that amount of data the software could estimate the integer ambiguity rather well and then round the results to the nearest integer. In this era kinematic GPS was not used often. Later both receiver hardware and data processing algorithms improved and by the 1990s both rapid static positioning and RTK positioning systems began using "on-the-fly" (OTF) integer ambiguity resolution.

Many RTK systems resolve the integer ambiguity, *on-the-fly*. *On-the-fly* refers to a method of resolving the carrier phase ambiguity very quickly. The method requires a dual-frequency GPS receiver capable of making both carrier phase and precise pseudorange measurements. The receiver is not required to remain stationary.

Here is one way it can be done. A search area is defined in the volume of the possible solutions, but that group is narrowed down quite a bit by using pseudoranges. If the number of integer combinations to be tested is greatly reduced with precise pseudoranges, the search can be very fast. The possible solutions in that volume are tested statistically, according to a minimal variance criterion, and the best one is found. This candidate is verified, that is, compared with the second best candidate. The process can take less than 10 seconds under the best circumstances where the receivers are tracking a large constellation of satellites, the PDOP is small, the receivers are dual-frequency, there is no multipath, and the receiver noise is low. This technique relies on dual-frequency information. Observations on L1 and L2 are combined into a wide lane, which has an ambiguity of about 86 cm, and the integer ambiguity is solved in a first pass. This information is used to determine the kinematic solution on L1. Therefore, it is a good idea to restrict RTK to situation, where there is good correlation of atmospheric biases at both ends of the baseline. In other words, RTK is best used when the distance between the base and rover is less than 20 km; this is usually not a problem.

Today the development of GPS receivers with virtually instantaneous carrier phase-based positioning has become feasible on a routine basis. Not only that, but these techniques of integer ambiguity resolution, validation, and quality control are being further improved to apply to GPS, GLONASS, and Galileo data processing.

RADIO LICENSE

RTK often requires a radio link between the base station and the rover, and the modems at each end must be tuned to the same frequency. The usual configuration operates at 4800 baud or faster. The units communicate with each other along a direct line-of-sight. Most radios connected to RTK GPS surveying equipment operate between UHF 400-475 MHz or VHF 170-220 MHz, and emergency voice communications also tend to operate in this same range, which can present problems from time to time.

The transmitter at the base station is usually the larger and more powerful of the two radios. However, the highest wattage radios, 35 Watts or so, cannot be legally operated in some countries. Lower power radios, from ½ W to 2 W, are sometimes used in such circumstances. The radio at the rover has usually lower power and is smaller. The *Federal Communications Commission* (*FCC*) is concerned with some RTK GPS operations interfering with other radio signals, particularly voice communications. It is important for GPS surveyors to know that voice communications have priority over data communications.

The FCC requires cooperation among licensees that share frequencies. Interference should be minimized. For example, it is wise to avoid the most typical community voice repeater frequencies. They usually occur between 455_460 MHz and 465_470 MHz. Part 90 of the Code of Federal Regulations, 47 CFR 90, contains the complete text of the FCC Rules including the requirements for licensure of radio spectrum for private land mobile use. The FCC does require application be made for licensing a radio transmitter. Fortunately, when the transmitter and rover receivers required for RTK operations are bought simultaneously radio licensing and frequency selection are often arranged by the GPS selling agent. Nevertheless, it is important that surveyors do not operate a transmitter without a proper license. Please remember that the FCC can levee fines for several thousand dollars for each day of illegal operation. More can be learned by consulting the FCC Wireless Fee Filing Guide http://wireless.fcc.gov/feesforms/feeguide/ and http://wireless.fcc.gov/feesforms/feeguide/services/landmobile.pdf.

There are also other international and national bodies that govern frequencies and authorize the use of signals elsewhere in the world. In some areas certain bands are designated for public use, and no special permission is required. For example, in Europe, it is possible to use the 2.4 GHz band for spread spectrum communication without special authorization with certain power limitations. Here in the United States the band for spread spectrum communication is 900 MHz.

CELL PHONES

There is an alternative to the radio link method of RTK; the corrections can be carried to the rover using a cell phone. The cell phone connection does tend to ameliorate the signal interruptions that can occur over the radio link, and it offers a somewhat wider effective range in some circumstances. The use of cell phones in this regard is frequently a characteristic of *Real-Time Network* (RTN) solutions, more about that later in this chapter.

TYPICAL RTK

A typical RTK set-up includes a base station and rover or rovers. They can be single- or dual-frequency receivers with GPS antennas, but dual-frequency receivers are usual. The radio receiving antennas for the rovers will either be built into the GPS antenna or separate units. It is usual that the radio antenna for the data transmitter and the rover are omnidirectional whip antennas, however at the base it is usually on a separate mast and has a higher gain than those at the rovers.

The position of the transmitting antenna affects the performance of the system significantly. It is usually best to place the transmitter antenna as high as is practical for maximum coverage, and the longer the antenna, the better its transmission characteristics. It is also best if the base station occupies a control station that has no overhead obstructions, is unlikely to be affected by multipath, and is somewhat away from the action if the work is on a construction site. It is also best if the base station is within line of sight of the rovers. If line of sight is not practical, as little obstruction as possible along the radio link is best.

The data radio transmitter consists of an antenna, a radio modulator, and an amplifier. The modulator converts the correction data into a radio signal. The amplifier increases the signal's power, which determines how far the information can travel. Well, not entirely; the terrain and the height of the antenna have something to do with it too. RTK work requires a great deal of information be successfully communicated from the base station to the receivers. The base station transmitter ought to be VHF, UHF, or spread spectrum-frequency hopping to have sufficient capacity to handle the load. UHF spread spectrum radio modems are the most popular for DGPS and RTK applications. The typical gain on the antenna at the base is 6 dB. But while DGPS operations may need no more than 200 bits-per-second (bps), updated every 10 seconds or so, RTK requires at least 2400 bps updated about every ½ second or less. Like the power of the transmission, the speed of the link between the base and rover, the *data rate,* can also be a limiting factor in RTK performance.

As mentioned earlier, RTK is at its best when the distance between the base station and the rovers is less than 20 km, under most circumstances, but even before that limit is reached the radio data link can be troublesome. In areas with high radio traffic it can be difficult to find an open channel. It is remarkable how often the interference emanates from other surveyors in the area doing RTK as well. That is why most radio data transmitters used in RTK allow the user several frequency options within the legal range.

It is vital, of course, that the rover and the base station are tuned to the same frequency for successful communication. The receiver also has an antenna and a demodulator. The demodulator converts the signal back to an intelligible form for the rover's receiver. The data signal from the base station can be weakened or lost at the rover from reflection, refraction, atmospheric anomalies, or even being too close. A rover that is too close to the transmitter may be overloaded and not receive the signal properly, and, of course, even under the best circumstances the signal will fade as the distance between the transmitter and the rover grows too large.

THE VERTICAL COMPONENT IN RTK

The output of RTK can appear to be somewhat similar to that of optical surveying with an electronic distance measuring (EDM) and a level. The results can be immediate and with similar relative accuracy. Nevertheless, it is not a good idea to consider the methods equivalent. RTK offers some advantages and some disadvantages when compared with more conventional methods. For example, RTK can be much more productive since it is available 24 hours a day and is not really affected by weather conditions. However, when it comes to the vertical component of surveying RTK and the level are certainly not equal.

GPS can be used to measure the differences in ellipsoidal height between points with good accuracy. However, unlike a level—unaided GPS cannot be used to measure differences in orthometric height. Or, as stated in Chapter 5, "orthometric elevations are not directly available from the geocentric position vectors derived from GPS measurements." The accuracy of orthometric heights in GPS is dependent on the veracity of the geoidal model used and the care with which it is applied.

Fortunately, ever improving geoid models have been, and still are, available from NGS. Since geoidal heights can be derived from these models, and ellipsoidal heights are available from GPS it is certainly feasible to calculate orthometric heights. In the past these calculations were done exclusively in postprocessed network adjustments. Today, more and more manufacturers are finding ways to include a geoid model in their RTK systems. However, it is important to remember that without a geoid model RTK will only provide differences in ellipsoid heights between the base station and the rovers.

It is not a good idea to presume that the surface of the ellipsoid is sufficiently parallel to the surface of the geoid and ignore the deviation between the two. They may depart from one another as much as a meter in 4 or 5 kilometers.

SOME PRACTICAL RTK SUGGESTIONS

Typical Satellite Constellations

In RTK, generally speaking, the more satellites that are available the faster the integer ambiguities will be resolved. In the United States, there are usually 6 satellites or so above an observer's horizon most of the time. And there are likely to be approximately 8 satellites above an observer's horizon about a third of the time, more only seldom. For baselines under 10 km an 8-satellite constellation should be quite adequate for good work under most circumstances.

Dual-Frequency Receiver

A dual-frequency receiver is a real benefit in doing RTK. Using a dual-frequency receiver instead of a single-frequency receiver is almost as if there were one and a half more satellites available to the observer.

Setting up a Base Station

Set up the base station over a known position first, before configuring the rover. After the tripod and tribrach are level and over the point, attach the GPS antenna to the tribrach and, if possible, check the centering again.

Set up the base station transmitter in a sheltered location at least 10 feet from the GPS antenna, and close to the radio transmitter's antenna. It is best if the airflow of the base station transmitter's cooling fan is not restricted. The radio transmitter's antenna is often mounted on a range pole attached to a tripod. Set the radio transmitter's antenna as far as possible from obstructions and as high as stability will allow.

The base station transmitter's power is usually provided by a deep-cycle battery. Even though the cable is usually equipped with a fuse, it is best to be careful to not reverse the polarity when connecting it to the battery. It is also best to have the base station transmitter properly grounded, and avoid bending or kinking any cables.

After connecting the base station receiver to the GPS antenna, to the battery and the data collector, if necessary, carefully measure the GPS antenna height. This measurement is often the source of avoidable error, both at the base station and the rovers. Many surveyors measure the height of the GPS antenna to more than one place on the antenna, and it is often measured in both meters and feet for additional assurance.

Select a channel on the base station transmitter that is not in use, and be sure to note the channel used so that it may be set correctly on the rovers as well.

After the RTK Survey

When the RTK work is done it is best to review the collected data from the data logger. Whether or not fixed height rods have been used it is a good idea to check the antenna heights. Incorrect antenna heights are a very common mistake. Another bulwark against blunders is the comparison of different observations of the same stations. If large discrepancies arise there is an obvious difficulty. Along the same line it is worthwhile to check for discrepancies in the base station coordinates. Clearly if the base coordinate is wrong the work created from that base is also wrong. Finally, look at the residuals of the final coordinates to be sure they are within reasonable limits. Remember that multipath and signal attenuation can pass by the observer without notice during the observations, but will likely affect the residuals of the positions where they occur.

COMPARING RTK AND DGPS

Multipath in RTK and DGPS

While most other errors in GPS discussed earlier can be mediated, or cancelled, due to the relatively short distance between receivers, the same cannot be said for multipath.

Environments that have the highest incidence of multipath, such as downtown city streets, are the places where RTK and DGPS called upon most often. Yet in these techniques the position is computed over a very few epochs; there is little if any of the averaging of errors from epoch to epoch available in other applications. Therefore, in RTK the effect of multipath is rather direct, and even if it distorts the results only slightly, it may be too much in some situations. In principle, multipath affects carrier observations the least. It is a bit worse for P-code pseudoranges. And in the observable used in most DGPS, C/A-code pseudoranges, multipath affects the results the most. But receiver manufacturers have created ingenious technologies to minimize the multipath effects in DGPS pseudoranges. Therefore, it is entirely possible to find that it is in the RTK carrier phase application that multipath contributes the largest bias to the error budget.

Initialization

As mentioned, RTK relies on the carrier phase and the integer ambiguity must be solved. In other words, the method requires some time for initialization, usually at the start of day. Initialization is also required any time after which the continuous tracking of all available satellite signals stops for even the briefest length of time. After tracking of the same GPS satellites begins at the base station and the roving receiver or receivers, there is usually a short wait required for initialization to be accomplished. With many dual-frequency RTK receivers capable of reinitializing, on-the-fly initialization can happen very quickly. Some receivers might require a return to a known coordinated position for reinitialization. On the other hand, DGPS relies on pseudoranges; its immediate initialization is a clear advantage over some RTK systems. But DGPS can only provide meter accuracy, whereas RTK offers centimeter-level positional accuracies.

Base Station

Concerning base stations, in both RTK and DGPS, a base station on a coordinated control position must be available. Its observations must be simultaneous with those at the roving receivers and it must observe the same satellites.

It is certainly possible to perform a differential survey in which the position of the base station is either unknown or based on an assumed coordinate at the time of the survey. However, unless only relative coordinates are desired, the position of the base station must be known. In other words, the base station must occupy a control position, even if that control is established later.

Utilization of the DGPS techniques requires a minimum 4 satellites for three-dimensional positioning. RTK ought to have at least 5 satellites for initialization. Tracking 5 satellites is a bit of insurance against losing one abruptly; also it adds considerable strength to the results. While cycle slips are always a problem it is imperative in RTK that every epoch contains a minimum of 4 satellite data without cycle slips. This is another reason to always track at least 5 satellites when doing RTK. Both methods most often rely on real-time communication between the base station and roving receivers. But RTK base station corrections are generally more complex than those required in DGPS.

Radio Technical Commission for Maritime (RTCM) Services Version 3

For DGPS there are more sources of real-time correction signals all the time. There are commercial providers and earthbound systems. Originally sources offered code-phase corrections in the RTCM SC-104 format appropriate to DGPS. However, when it became clear in 1994 that including carrier phase information in the message could improve the accuracy of the system, RTCM Special Committee 104 added four new message types to Version 2.1 to fulfill the needs of RTK. Still proprietary message formats were more widely used in RTK work than in DGPS so further improvements were made along the same line in Version 2.2, which became available in 1998. And the changes continue: in 2007 the Radio Technical Commission for Maritime Services Special Committee 104 published its Version 3 for Differential Global Navigation Satellite System (GNSS) services. It is called RTCM 10403.1; documents concerning the details of this standard are available at www.rtcm.org.

The GPS constellation along with the Russian GLONASS system and the European Galileo system are currently known together as the *Global Navigation Satellite Systems, GNSS*. It is likely that more systems will become included under the GNSS concept in the future. It is also likely that more accuracy of autonomous positions will be available from GNSS than GPS alone. However, in GNSS, as with GPS, even better accuracies can be achieved by broadcasting corrections from reference stations at precisely known locations. And by utilizing RTCM 10403.1 it is not only possible to use receivers from different manufacturers together, but also to incorporate signals from satellites other than GPS. Perhaps the term DGPS will be expanded to become *DGNSS*.

REAL-TIME NETWORK SERVICES

There is no question that RTK dominates the GPS surveying applications. It is applicable to much of engineering, surveying, air-navigation, mineral exploration, machine control, hydrography, and a myriad of other areas that require centimeter-level accuracy in real-time. However, the requirements of setting up a GPS reference station on a known position, the establishment of an radio frequency (RF) transmitter and all attendant components before a single measurement can be made are both awkward and expensive. This, along with the baseline limitation of short baselines, 10 to 20 kilometers for centimeter level work, has made RTK both more cumbersome and less flexible than most surveyors prefer.

In an effort to alleviate these difficulties, services have arisen around the world to provide RTCM real-time corrections to surveyors by a different means. The pace of the development of these Real-Time Networks (RTN) (see Figure 6.11), both by governments and commercial interests, is accelerating. The services are sometimes free and sometimes require the arrangement of a subscription or the payment of a fee before the surveyor can access the broadcast corrections over a data link via a modem such as a cell phone or some other device. Nevertheless, there are definite advantages including the elimination of individual base station preparation and the measurement of longer baselines without rapid degradation of the results. These benefits are accomplished by the services gleaning corrections from a whole network of continuously

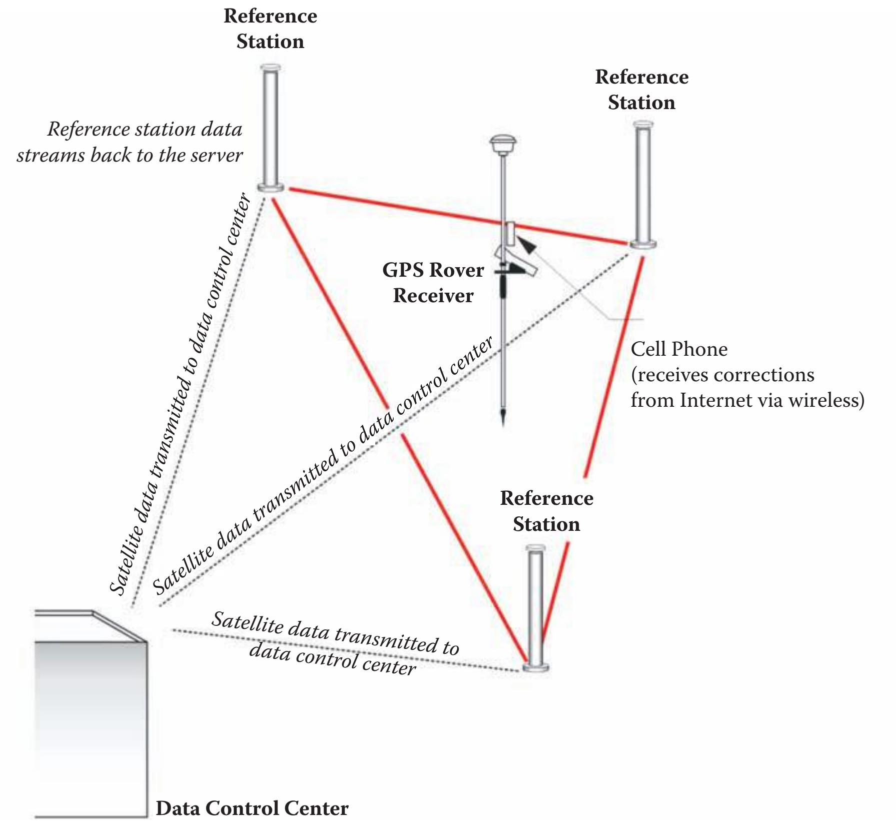


FIGURE 6.11 Real-Time Network GPS

operating reference stations, CORS, rather than just a single base. In this way, quality control is facilitated by the ability to check corrections from one CORS with those generated from another, and should a CORS go off-line or give incorrect values other CORS in the network can take up the slack little accuracy loss.

The central idea underlying RTN differential corrections is the combination of observations from several CORS at known positions used to derive a model of an entire region. So rather than being considered as isolated beacons with each covering its own segregated area—the CORS are united into a network. The data from the network can then be used to produce a virtual model of the area of interest, and from this model *distance-dependent* biases such as ionospheric, tropospheric, and orbit errors can be calculated. Once the roving receiver's place within that network is established it is possible to predict the errors at that position with a high degree of certainty. And not only can the CORS network be used to model errors in a region more correctly, but the multi-base solution also can improve redundancy. Solving several baselines that converge on a project point simultaneously rather than relying on just one from a typical RTK set up adds more certainty to resulting coordinate.

Implementing an RTN requires data management and communication. The information from the CORS must be communicated to the central master control station where all the calculations are done. Their raw measurement data, orbits, and so forth, must be managed as they are received in real-time from each of the CORS that make up the network. Along with the modeling of the distance-dependent errors all the integer ambiguities must be fixed for each CORS in real-time. This is probably the most significant data processing difficulty required of an RTN, especially considering that there are usually large distances between the CORS. To accomplish it, postcomputed ephemerides, antenna phase center corrections, and all other available information are brought to bear on the solution such as tropospheric modeling and ionospheric modeling.

Modeling is subject to variation in both space and time. For example, ionospheric and orbit biases are satellite specific, whereas tropospheric corrections can be estimated station by station. But the ionospheric, dispersive, biases change more rapidly than the tropospheric and orbit biases, which are nondispersive. Therefore, ionospheric corrections must be updated more frequently than orbit and tropospheric corrections. And while it is best to keep the modeling for ionosphere within the limited area around three or so CORS, when it comes to tropospheric and orbital modeling the more stations used the better.

Finally, the pseudorange and/or carrier phase residuals must be determined for the L1 and/or L2, by using one of many techniques to interpolate the actual distance-dependent corrections for the surveyor's particular position within the network. Then the subsequent corrections must be communicated to the surveyor in the field, which typically requires the transmission of a large amount of data. There is more than one way the appropriate correction can be determined for a particular position within an RTN. So far there is no clear best method. One approach is the creation of a *Virtual Reference Station* (*VRS*) and the attendant corrections. This approach requires a two-way communication link. Users must send their approximate positions to the master control center, usually via an National Marine Electronics Association (NMEA) string. The master control center returns corrections for an individual VRS, via RTCM and then the baseline processing software inside the rover calculates its position using the VRS, which seems to the receiver to be a single nearby reference station. Another method involves sending basic RTK type corrections. Or the system may broadcast raw data for all the reference stations.

PRECISE POINT POSITIONING

Differential GPS, whether real-time or postprocessed, is the foundation of nearly all surveying applications in the field. However, when precise GPS orbital information and clock data became broadly available, a new technique arose that is not differential in the usual sense.

As mentioned earlier, errors in satellite clocks and imperfect orbital information contribute substantial inaccuracies to GPS work, especially in autonomous positioning. And the corrections available from the Navigation Message are not adequate to reduce those biases sufficiently, but positions several orders of magnitude better than the broadcast corrections can support can be achieved with *Precise Point Positioning* (*PPP*). The technique often utilizes the precise GPS orbit and clock products available from international sources to correct data collected by dual-frequency receivers to ameliorate the ionospheric delay. It is important to note that PPP is not a code-only solution but rather relies on code and phase observations. The sources for the precise orbit and clocks products include the *International GNSS Service* (*IGS*)*, the Jet Propulsion Laboratory* (*JPL*)*,* and *Natural Resources Canada* (*NRCan*) among others. In fact, NRCan offers an online service known as *CSRS-Precise Point Positioning* (*CSRS-PPP*), which permits users to submit their collected GPS data over the Internet for PPP processing. GPS observations with the results can be returned in NAD83 or ITRF -http://www.geod.nrcan.gc.ca/ppp_e.php. On the downside, there can be a bit of lag time between the collection of the data and the availability of precise ephemeris and clock information from free sources and PPP can require a rather long initialization period, perhaps a half-hour. And since the method utilizes indifferenced observations, PPP cannot rely on the concept that the carrier phase ambiguity, N, is an integer. The lack of certain knowledge that N must be an integer complicates the solution.

However, there are several advantages to PPP. A single GPS receiver is all that is required and there is no need to run a base station during the work. While an RTN also alleviates a surveyor from the responsibility of operating a second receiver, PPP removes the necessity to stay within the area covered by a network, a base station, or a CORS. PPP can provide more consistency over large areas, generally speaking and offers a valuable solution where a control network is simply unavailable.

SUMMARY

So here are the fundamental ideas that underlie some GPS surveying techniques. Most often, a fixed point with previously established coordinates is occupied by a one receiver, a base station. The first receiver, the base, provides the data to compute the differences between its known position and the unknown positions measured at the second, or roving receiver. Static GPS requires postprocessing and is used for establishing control by GPS. With DGPS the corrections can either be applied in postprocessing or in real-time. And while correlated systematic errors can be virtually eliminated with differential correction, the biases such as multipath and receiver channel noise are certainly not. These errors have been much reduced in modern GPS receivers, but not completely defeated. Biases such as high PDOP can only be resolved with both good receiver design and care as to when and where the surveying is done. Real-Time Kinematic is the method of choice for much of GPS work, but the technique is being supported more and more by Real-Time Networks as opposed to dedicated individual base stations. And Precise Point Positioning (PPP) offers a centimeter or decimeter solution that does not rely on differential correction as it is usually defined, but rather precise ephemeris and clock data. Finally, all of these techniques will be improved as the Global Navigation Satellite System, GNSS, matures.

EXERCISES

1. What is the meaning of latency as applied to DGPS and RTK?

   (a) The baud-rate of a radio modem in real-time GPS.
   (b) The time taken for a system to compute corrections and transmit them to users in real-time GPS.
   (c) The frequency of the RTCM SC104 correction signal in real-time GPS.
   (d) The range rate broadcast with the corrections from the base station.

2. Which of the following errors are not reduced by using DGPS or RTK methods?

   (a) atmospheric errors
   (b) satellite clock bias
   (c) ephemeris bias
   (d) multipath

3. Which of the following statements about RTCM SC104 is not true?

   (a) RTCM SC104 is a format for communication used in real-time GPS.
   (b) RTCM SC104 was originally designed in1985.
   (c) RTCM SC104 can be used in both DGPS and RTK applications.
   (d) RTCM SC104 was revised in 1994 and 1996.

4. Who was the original kinematic GPS innovator?

   (a) Ann Bailey
   (b) Benjamin Remondi
   (c) George F. Syme
   (d) R.E. Kalman

5. Which of the following statements about rules governing RTK radio communication is correct?

   (a) According to the FCC regulations, voice communications have priority over data communications.
   (b) Code of Federal Regulations, 48 CFR 91, contains the complete text of the FCC Rules.
   (c) Most typical community voice repeater frequencies are about 900 MHz.
   (d) In Europe it is possible to use the 4.2 GHz band for spread spectrum communication without special authorization.

6. Which of the following data rates is closest to that required for RTK?

   (a) 200 bps updated every 10 seconds
   (b) 1,400 bps updated about every 5 seconds
   (c) 2,400 bps updated about every ½ of a second
   (d) 3,500 bps updated about every 0.05 of a second

7. How many satellites is the minimum for the best initialization with RTK?

   (a) 4 satellites
   (b) 5 satellites
   (c) 6 satellites
   (d) 8 satellites

8. What is a typical range of systematic error that can currently be expected in an autonomous point position?

   (a) ±100 m to ±120 m
   (b) ±75 m to ±95 m
   (c) ±45 m to ±65 m
   (d) ±20 m to ±40 m

9. Which of the following is not a part of a typical data radio transmitter used in RTK?

   (a) a radio modulator
   (b) an amplifier
   (c) a demodulator
   (d) an antenna

10. Which of the following does not correctly describe a difference between RTK and DGPS?

    (a) RTK relies on the carrier phase and the integer ambiguity must be solved, but with DGPS it does not.
    (b) The RTK base station corrections are more complex than those required in DGPS.
    (c) RTK offers centimeter-level positional accuracies; DGPS provides meter-level positional accuracies.
    (d) RTK is affected by multipath, but DGPS is not.

11. Which of the following information is not available from the NGS data sheet for a particular station?

    (a) the type of monument
    (b) the State Plane coordinates
    (c) the latitude and longitude of the primary azimuth mark
    (d) the Permanent Identifier

12. Which of the following FGCC geometric accuracy standards would be correctly included under C?

    (a) 5 mm +/– 1:10,000,000
    (b) 1 cm +/– 1: 1000,000
    (c) 8 mm +/– 1: 1,000,000
    (d) 3 mm +/– 1:100,000,000

13. Which of the following is a quantity which, when added to an astronomic azimuth, yields a geodetic azimuth?

    (a) the primary azimuth
    (b) the mapping angle, also known as the convergence
    (c) the Laplace correction
    (d) the second term

14. How many nontrivial, or independent, and how many trivial, or dependent baselines are created in 1 GPS session using 4 receivers?

    (a) 1 independent baseline and 5 dependent baselines.
    (b) 3 independent baselines and 1 dependent baseline.
    (c) 2 independent baselines and 3 dependent baselines.
    (d) 3 independent baselines and 3 dependent baselines.

15. Which of the following statements concerning loop closure in a GPS network is not correct?

    (a) A loop closure that uses baselines from one GPS session will appear to have no error at all.
    (b) Loop closure is a procedure by which the internal consistency of a GPS network is discovered.
    (c) No baselines should be excluded from a loop closure analysis.
    (d) At the completion of a GPS control survey the independent vectors should be capable of formation into closed loops that incorporate baselines from two to four different sessions.

16. How many observation sessions will be required for a GPS control survey involving 20 stations that is to be done with 4 GPS receivers, a planned redundancy of 2, a production factor of 1.1, and a safety factor is 0.2?

    (a) 21 observation sessions
    (b) 17 observation sessions
    (c) 15 observation sessions
    (d) 12 observation sessions

17. Why is it advisable that successive occupations in a GPS survey be separated by at least a quarter of an hour and less than a full day when possible?

    (a) To eliminate multipath when it corrupted the first occupation of the station.
    (b) To allow the satellite constellation can reach a significantly different configuration than it had during the first occupation of the station.
    (c) To overcome an overhead obstruction during the first occupation of the station.
    (d) To eliminate receiver clock errors.

18. What of the four following options describes a best configuration for the horizontal control of a GPS static control survey?

    (a) Divide the project into four quadrants and choose at least one horizontal control station near the project boundary in each quadrant.
    (b) Divide the project into three parts and choose at least one horizontal control station near the center of the project in each.
    (c) The primary base station should be located at the center of the project.
    (d) Draw a north-south line through the center of the project, and choose at least three horizontal control stations near that line.

19. Which of the following data sheet searches is not possible on the NGS Internet site?

    (a) A rectangular search based upon the range of latitudes and longitudes.
    (b) A search using a particular station's geoid height.
    (c) A search using a particular station's Permanent Identifier (PID).
    (d) A radial search that is defining the region of the survey with one center position and a radius.

20. Which of the following is the standard symbol used in this book for project points in a GPS control survey?

    (a) A solid dot
    (b) A triangle
    (c) A square
    (d) A circle

ANSWERS AND EXPLANATIONS

1. Answer is (b)

Explanation: In DGPS and RTK it takes some time for the base station to calculate corrections and it takes some time for it to transmit them. The base station's data is put into packets in the correct format. The data makes its way from the base station to the rover over the data link. It must then be decoded and go through the rover's software. The time it takes for all of this to happen is called the latency of the communication between the base station and the rover. It can be as little as a quarter of a second or as long as a couple of seconds.

2. Answer is (d)

Explanation: In autonomous point positioning all the biases discussed earlier affect the GPS position. For example, in autonomous point positioning the satellite clock bias normally contributes about 3 m to the final GPS positional error, ephemeris bias contributes about 5 m and solar radiation, and so forth, contribute about another 1.5 m. But with DGPS and RTK, when errors are well correlated all these biases are virtually zero, as are the atmospheric errors, unlike autonomous point positioning where the ionospheric delay contributes about 5 m and the troposphere about 1.5 m. While most other errors in GPS discussed earlier can be mediated, or canceled, due to the relatively short distance between receivers, the same cannot be said for multipath.

3. Answer is (d)

Explanation: In DGPS there is an agreed upon format for the communication between the base station and rovers. It was first designed for marine navigation. In 1985, *the Radio Technical Commission for Maritime Services* (*RTCM*) *Study Committee* (*SC-104*) created a standard format for DGPS corrections that is still used more than any proprietary formats that have come along since. Both the United States and the Canadian Coast Guards use this format in their coastal beacons. When it became clear in 1994 that including carrier phase information in the message could improve the accuracy of the system RTCM Special Committee 104 added four new message types to Version 2.1 to fulfill the needs of RTK. Further improvements were made along the same line in Version 2.2, which became available in 1998.

4. Answer is (b)

Explanation: Kinematic surveying, also known as stop-and-go kinematic surveying, is not new. The original kinematic GPS innovator, Dr. Benjamin Remondi, developed the idea in the mid-1980s.

5. Answer is (a)

Explanation: In the United Stated most transmitters connected to RTK GPS surveying equipment operate between 450_470 MHz, and voice communications also operate in this same range. It is important for GPS surveyors to know that voice communications have priority over data communications. The FCC requires cooperation among licensees that share frequencies. Interference should be minimized. For example, it is wise to avoid the most typical community voice repeater frequencies. They usually occur between 455_460 MHz and 465_470 MHz. Part 90 of the Code of Federal Regulations, 47 CFR 90, contains the complete text of the FCC Rules.

There are also other international and national bodies that govern frequencies and authorize the use of signals elsewhere in the world. In some areas certain bands are designated for public use, and no special permission is required. For example, in Europe it is possible to use the 2.4 GHz band for spread spectrum communication without special authorization with certain power limitations. Here in the United States the band for spread spectrum communication is 900 MHz.

6. Answer is (b)

Explanation: DGPS operations may need no more than 200 bps, bits-per-second, updated every 10 seconds or so. RTK requires at least 2,400 bps updated about every ½ second or less.

7. Answer is (b)

Explanation: Utilization of the DGPS techniques requires a minimum 4 satellites for three-dimensional positioning. RTK ought to have at least 5 satellites for initialization. Tracking 5 satellites is a bit of insurance against losing one abruptly; also it adds considerable strength to the results. While cycle slips are always a problem it is imperative in RTK that every epoch contains a minimum of 4 satellite data without cycle slips. This is another reason to always track at least 5 satellites when doing RTK. However, most receivers allow the number to drop to 4 after initialization.

8. Answer is (d)

Explanation: An autonomous point position in GPS with SA turned off will usually have about ±20m to ±40m of systematic bias in a horizontal position, whereas a DGPS position may be subject to a tenth of that, or even less.

9. Answer is (c)

Explanation: A data radio transmitter consists of a radio modulator, an amplifier, and an antenna. The modulator converts the correction data into a radio signal. The amplifier increases the signal's power. And then the information is transmitted via the antenna.

10. Answer is (d)

Explanation: Both methods most often rely on real-time communication between the base station and roving receivers. But RTK base station corrections are more complex than those required in DGPS. In surveying applications, since RTK must work in a somewhat limited area it usually requires establishment of a project specific base station. RTK relies on the carrier phase and the integer ambiguity must be solved. On the other hand, DGPS relies mostly on pseudoranges; its immediate initialization is a clear advantage over some RTK systems. But DGPS can only provide meter accuracy, whereas RTK offers centimeter-level positional accuracies. Both techniques are affected by multipath.

11. Answer is (c)

Explanation: NGS data sheets provide State Plane and UTM coordinates, the latter only for horizontal control stations. State Plane Coordinates are given in either U.S. Survey Feet or International Feet and UTM coordinates are given in meters. They also provide mark setting information, the type of monument, and the history of mark recovery. The data sheet certainly shows the station's designation, which is its name and its Permanent Identifier (PID). Either of these may be used to search for the station in the NGS database. The PID is also found all along left side of each data sheet record and is always two upper case letters followed by four numbers. However, it does not show the latitude and longitude of azimuth marks. That information may sometimes be found on a data sheet devoted to the particular azimuth mark.

12. Answer is (b)

Explanation: The geometric accuracy of 1 cm +/- 1: 1000,000 is a familiar standard for first-order horizontal control. This designation and second- and third-order are now augmented by AA-, A-, and B- order stations as well. Horizontal AA-order stations have a relative accuracy of 3 mm +/- 1:100,000,000 relative to other AA-order stations. Horizontal A-order stations have a relative accuracy of 5 mm +/- 1:10,000,000 relative to other A-order stations. Horizontal B-order stations have a relative accuracy of 8 mm +/- 1: 1,000,000 relative to other A- and B-order stations.

13. Answer is (c)

Explanation: The Laplace correction is a quantity which, when added to an astronomically observed azimuth, yields a geodetic azimuth. It is important to note that NGS uses a clockwise rotation regarding the Laplace correction. It can contribute several seconds of arc.

14. Answer is (c)

Explanation: Where *r* is the number of receivers, every session yields *r*-1 independent baselines. Four receivers used in 1 session would produce 6 baselines. Of these 3 would be independent or nontrivial baselines, and 3 would be dependent or trivial baselines. It cannot be said that the shortest lines are always chosen to be the independent lines. Sometimes there are reasons to reject one of the shorter vectors due to incomplete data, cycle slips, multipath, or some other weakness in the measurements. Before such decisions can be made, each session will require analysis after the data has actually been collected. In the planning stage, it is best to consider the shortest vectors as the independent lines, but once the decision is made only those 3 baselines are included in the network.

15. Answer is (c)

Explanation: Any loop closures that only use baselines derived from a single common GPS session will yield an apparent error of zero, because they are derived from the same simultaneous observations. When the project is completed, the independent vectors in the network, excluding the dependent, or trivial baselines, should be capable of formation into closed loops that incorporate baselines from two to four different sessions.

In GPS the error of closure is valid for orders A and B when three or more independent baselines from three or more GPS sessions are included in the loop. Order AA requires four independent observations be included. Loop closures for first-order and lower should include a minimum of two observing sessions.

16. Answer is (c)

Explanation: The appropriate formula is

$$s = \frac{(m \cdot n)}{r} + \frac{(m \cdot n)(p-1)}{r} + k \cdot m$$

where

s = the number of observing sessions
r = the number of receivers
m = the total number of stations involved
n = the planned redundancy
p = the experience, or production factor
k = the safety factor

The variable n is a representation of the level of redundancy that has been built into the network, based on the number of occupations on each station. The experience of a firm is symbolized by the variable *p* in the formula. A typical production factor is 1.1. A safety factor of 0.1, known as *k*, is recommended for GPS projects within 100 km of a company's home base. Beyond that radius, an increase to 0.2 is advised.

The substitution of the appropriate quantities for the illustrated project increases the prediction of the number of observation sessions required for its completion:

$$s = \frac{(mn)}{r} + \frac{(mn)(p-1)}{r} + km$$

$$s = \frac{(20)(2)}{4} + \frac{(20)(2)(1.1-1)}{4} + (0.2)(20)$$

$$s = \frac{40}{4} + \frac{4}{4} + 4$$

$$s = 10 + 1 + 4$$

$$s = 15 \text{ sessions}$$

17. Answer is (b)

Explanation: Successive occupations ought to be separated by at least a quarter of an hour and less than a full day so the satellite constellation can reach a significantly different configuration than that which it had during the first occupation. Please recall that GPS measurements are not actually made between the occupied stations, but directly to the satellites themselves.

18. Answer is (c)

Explanation: For work other than route surveys, a handy rule of thumb is to divide the project into four quadrants and to choose at least one horizontal control station in each quadrant. The actual survey should have at least one horizontal control station in three of the four quadrants. Each of them ought to be as near as possible to the project boundary. Supplementary control in the interior of the network can then be used to add more stability to the network.

19. Answer is (b)

Explanation: It is quite important that to have the most up-to-date control information from NGS. A rectangular search based upon the range of latitudes and longitudes can now be performed on the NGS internet site. It is also possible to do a radial search—defining the region of the survey with one center position and a radius. You may also retrieve individual data sheets by the Permanent Identifier (PID) control point name, which is known as the *designation*, survey project identifier, or USGS quad. However, you cannot retrieve a data sheet for a station based only on its geoid height.

20. Answer is (a)

Explanation: A solid dot is the standard symbol used to indicate the position of project points. Some variation is used when a distinction must be drawn between those points that are in place and those that must be set. When its location is appropriate, it is always a good idea to have a vertical or horizontal control station serve double duty as a project point. While the precision of their plotting may vary, it is important that project points be located as precisely as possible even at this preliminary stage. A horizontal control point is shown as a triangle, and, vertical control point is shown as a square.

7 Observing and Processing

STATIC GPS CONTROL OBSERVATIONS

The prospects for the success of a GPS project are directly proportional to the quality and training of the people doing it. The handling of the equipment, the on-site reconnaissance, the creation of field logs, and the inevitable last-minute adjustments to the survey design all depend on the training of the personnel involved for their success. There are those who say the operation of GPS receivers no longer requires highly qualified survey personnel. That might be true if effective GPS surveying needed only the pushing of the appropriate buttons at the appropriate time. In fact, when all goes as planned, it may appear to the uninitiated that GPS has made experienced field surveyors obsolete. But when the unavoidable breakdowns in planning or equipment occur, the capable people, who seemed so superfluous moments before, suddenly become indispensable.

EQUIPMENT

Conventional Equipment

Most GPS projects require conventional surveying equipment for spirit-leveling circuits, offsetting horizontal control stations, and monumenting project points, among other things. It is perhaps a bit ironic that this most advanced surveying method also frequently has need of the most basic equipment. The use of brush hooks, machetes, axes, and so forth, can sometimes salvage an otherwise unusable position by removing overhead obstacles. Another strategy for overcoming such hindrances has been developed using various types of survey masts to elevate a separate GPS antenna above the obstructing canopy.

Flagging, paint, and the various techniques of marking that surveyors have developed over the years are still a necessity in GPS work. The pressure of working in unfamiliar terrain is often combined with urgency. Even though there is usually not a moment to spare in moving from station to station, a GPS surveyor frequently does not have the benefit of having visited the particular points before. In such situations, the clear marking of both the route and the station during reconnaissance is vital.

Despite the best route marking, a surveyor may not be able to reach the planned station, or, having arrived, finds some new obstacle or unanticipated problem that can only be solved by marking and occupying an impromptu offset position for a session. A hammer, nails, shiners, paint, and so forth, are essential in such situations.

In short, the full range of conventional surveying equipment and expertise have a place in GPS. For some, their role may be more abbreviated than it was formerly, but one element that can never be outdated is good judgment.

Safety Equipment

The high-visibility vests, cones, lights, flagmen, and signs needed for traffic control cannot be neglected in GPS work. Unlike conventional surveying operations, GPS observations are not deterred by harsh weather. Occupying a control station in a highway is dangerous enough under the best of conditions, but in the midst of a rainstorm, fog, or blizzard, it can be absolute folly without the proper precautions. And any time and trouble taken to avoid infraction of the local regulations regarding traffic management will be compensated by an uninterrupted observation schedule.

Weather conditions also affect travel between the stations of the survey, both in vehicles and on foot. Equipment and plans to deal with emergencies should be part of any GPS project. First aid kits, fire extinguishers, and the usual safety equipment are necessary. Training in safety procedures can be an extraordinary benefit, but perhaps the most important capability in an emergency is communication.

Communications

Whether the equipment is handheld or vehicle mounted, two-way radios and cell phones are used in most GPS operations. However, the line of sight that is no longer necessary for the surveying measurements in GPS is sorely missed in the effort to maintain clear radio contact between the receiver operators. A radio link between surveyors can increase the efficiency and safety of a GPS project, but it is particularly valuable when last-minute changes in the observation schedule are necessary. When an observer is unable to reach a station or a receiver suddenly becomes inoperable, unless adjustments to the schedule can be made quickly, each end of all of the lines into the missed station will require reobservation.

The success of static GPS hinges on all receivers collecting their data simultaneously. However, it is more and more difficult to ensure reliable communication between receiver operators in geodetic surveys, especially as their lines grow longer.

One alternative to contact between surveyors is reliance on the preprogramming feature available on most GPS receivers today. This attribute usually allows the start/stop time, sampling rate, bandwidth, satellites to track, mask angle, data file name, and start position to be preset so that the operator need only set up the receiver at the appropriate station before the session begins. The receiver is expected to handle the rest automatically. In theory, this approach eliminates the chance for an operator error ruining an observation session by missing the time to power-up or improperly entering the other information. Theory falls short of practice here, and even if the procedure could eliminate those mistakes, entire categories of other errors remain unaddressed. Some advocate actually leaving receivers unattended in static GPS. This idea seems unwise on the face of it.

High-wattage, private-line FM radios are quite useful when line of sight is available between them or when a repeater is available. The use of cell phones may eliminate the communication problem in some areas, but probably not in remote locations.

Despite the limitations of the systems available at the moment, achievement of the best possible communication between surveyors on a GPS project pays dividends in the long run.

GPS Equipment

Most GPS receivers capable of geodetic accuracy are designed to be mounted on a tripod, usually with a tribrach and adaptor. However, there is a trend toward bipod- or range-pole-mounted antennas. An advantage of these devices is that they ensure a constant height of the antenna above the station. The mismeasured height of the antenna above the mark is probably the most pervasive and frequent blunder in GPS control surveying.

The tape or rod used to measure the height of the antenna is sometimes built into the receiver and, sometimes a separate device. It is important that the height of instrument (H.I.) be measured accurately and consistently in both feet and meters, without merely converting from one to the other mathematically. It is also important that the value be recorded in the field notes and, where possible, also entered into the receiver itself.

Where tribrachs are used to mount the antenna, the tribrach's optical centering should be checked and calibrated. It is critical that the effort to perform GPS surveys to an accuracy of centimeters not be frustrated by inaccurate centering or H.I. measurement. Since many systems measure the height of the antenna to the edge of the ground plane or to the exterior of the receiver itself, the calibration of the tribrach affects both the centering and the H.I. measurement. The resetting of a receiver that occupies the same station in consecutive sessions is an important source of redundancy for many kinds of GPS networks. However, integrity can only be added if the tribrach has been accurately calibrated.

The checking of the carrier phase receivers themselves is also critical to the control of errors in a GPS survey, especially when different receivers or different models of antennas are to be used on the same work. The zero baseline test is a method that may be used to fulfill equipment calibration specifications where a three-dimensional test network of sufficient accuracy is not available. As a matter of fact, the simplicity of this test is an advantage. It is not dependent on special software or a test network. This test can also be used to separate receiver difficulties from antenna errors.

Two or more receivers are connected to one antenna with a *signal,* or *antenna splitter.* The antenna splitter can be purchased from specialty electronics shops and is also available online. An observation is done with the divided signal from the single antenna reaching both receivers simultaneously. Since the receivers are sharing the same antenna, satellite clock biases, ephemeris errors, atmospheric biases, and multipath are all canceled. In the absence of multipath, the only remaining errors are attributable to random noise and receiver biases. The success of this test depends on the signal from one antenna reaching both receivers, but the current from only one receiver can be allowed to power the antenna. This test checks not only the precision of the receiver measurements, but also the processing software. The results of the test should show a baseline of only a few millimeters.

Information is also available on National Geodetic Survey (NGS) calibration baselines throughout the United States. You can learn more by visiting the NGS site at http://www.ngs.noaa.gov/.

Auxiliary Equipment

Tools to repair the ends of connecting cables, a simple pencil eraser to clean the contacts of circuit boards, or any of a number of small implements have saved more than one GPS observation session from failure. Experience has shown that GPS surveying requires at least as much resourcefulness, if not more, than conventional surveying.

The health of the batteries is a constant concern in GPS. There is simply nothing to be done when a receiver's battery is drained but to resume power as soon as possible. A back-up power source is essential. Cables to connect a vehicle battery, an extra fully-charged battery unit, or both should be immediately available to every receiver operator.

Papers

The papers every GPS observer carries throughout a project ought to include emergency phone numbers; the names, addresses, and phone numbers of relevant property owners; and the combinations to necessary locks. Each member of the team should also have a copy of the virtual or hard copy project map, any other maps that are needed to clarify position or access, and, perhaps most important of all, the updated observation schedule.

The observation schedule for static GPS work will be revised daily based upon actual production (Table 7.1). It should specify the start–stop times and station for all the personnel during each session of the upcoming day. In this way, the schedule will not only serve to inform every receiver operator of his or her own expected occupations, but those of every other member of the project as well. This knowledge is most useful when a sudden revision requires observers to meet or replace one another.

STATION DATA SHEET

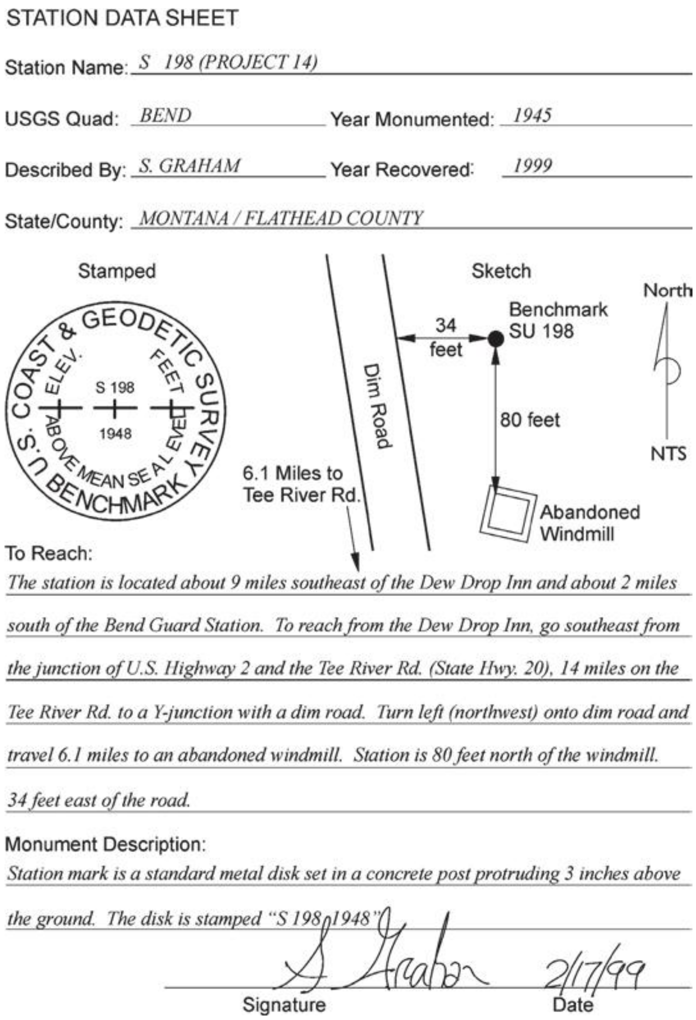
The principles of good field notes have a long tradition in land surveying, and they will continue to have validity for some time to come. In GPS, the ensuing paper trail will not only fill subsequent archives; it has immediate utility. For example, the station data sheet is often an important bridge between on-site reconnaissance and the actual occupation of a monument.

Though every organization develops its own unique system of handling its field records, most have some form of the station data sheet. The document illustrated in Figure 7.1 is merely one possible arrangement of the information needed to recover the station.

The station data sheet can be prepared at any period of the project, but perhaps the most usual times are during the reconnaissance of existing control or immediately after the monumentation of a new project point. Neatness and clarity, always paramount virtues of good field notes, are of particular interest when the station data sheet is to be later included in the final report to the client. The overriding principle

TABLE 7.1
Observation Schedule

	Session I	**8:10**	**Session 2**	**9:50**	**Session 3**	**11:15**	**Session 4**	**12:30**	**Session 5**
	Start 7:10	**to**	**Start 8:40**	**to**	**Start 10:15**	**to**	**Start 11:30**	**to**	**Start 14:00**
	Stop 8:10	**8:40**	**Stop 9:50**	**10:15**	**Stop 11:15**	**11:30**	**Stop 12:30**	**14:00**	**Stop 15:00**
Svs PRNs	**9,12,13,16,20,24**		**3,12,13,16,20,24**		**3,12,13,16,17,20,24**		**3,16,17,20,22,23,26**		**1,3,17,21,23,26,28**
Receiver A	Station 1		Station 1		Station 5		Station 5		Station 5
Dan H.	NGS Horiz.	Re-Set	NGS Horiz.	Move	NGS	Re-Set	NGS	Re-Set	NGS
	Control		Control		Benchmark		Benchmark		Benchmark
Receiver B	Station 3		Station 3		Station 6		Station 6		Station 1
Scott G.	NGS V&H	Re-Set	Move	Project	Re-Set	Project	Move	NGS Horiz.	
	Control		Control		Point		Point		Control
Receiver C	Station 8		Station 2		Station 2		Station 7		Station 10
Dewey A.	NGS Horiz.	Move	Project	Re-Set	Project	Movw	Project	Move	NGS
	Control		Point		Point		Point		Benchmark
Receiver D	Station 13		Station 4		Station 4		Station 9	Station 13	
Cindy E.	NGS Horiz.	Move	Project	Re-Set	Project	Move	Project	Move	NGS Horiz.
	Control		Point		Point		Point		Control



STATION DATA SHEET

Station Name: *S 198 (PROJECT 14)*

USGS Quad: *BEND* Year Monumented: *1945*

Described By: *S. GRAHAM* Year Recovered: *1999*

State/County: *MONTANA / FLATHEAD COUNTY*

To Reach:

The station is located about 9 miles southeast of the Dew Drop Inn and about 2 miles south of the Bend Guard Station. To reach from the Dew Drop Inn, go southeast from the junction of U.S. Highway 2 and the Tee River Rd. (State Hwy. 20), 14 miles on the Tee River Rd. to a Y-junction with a dim road. Turn left (northwest) onto dim road and travel 6.1 miles to an abandoned windmill. Station is 80 feet north of the windmill. 34 feet east of the road.

Monument Description:

Station mark is a standard metal disk set in a concrete post protruding 3 inches above the ground. The disk is stamped "S 198 1948"

S Graham 2/17/99

Signature Date

FIGURE 7.1 Station Data Sheet

in drafting a station data sheet is to guide succeeding visitors to the station without ambiguity. A GPS surveyor on the way to observe the position for the first time may be the initial user of a station data sheet. A poorly written document could void an entire session if the observer is unable to locate the monument. A client, later struggling to find a particular monument with an inadequate data sheet, may ultimately question the value of more than the field notes.

Station Name

The station name fills the first blank on the illustrated data sheet. Two names for a single monument are far from unusual. In this case the vertical control station, officially named S 198, is also serving as a project point, number 14. But two names purporting to represent the same position can present a difficulty. For example, when a horizontal control station is remonumented, a number 2 is sometimes added to the original name of the station and it can be confusing. For example, it can be easy to mistake station, “Thornton 2,” with an original station named, “Thornton,” that no longer exists. Both stations may still have a place in the published record, but with slightly different coordinates. Another unfortunate misunderstanding can occur when inexperienced field personnel mistake a reference mark, R.M., for the actual station itself. Taking rubbings and/or close-up photographs are widely recommended to avoid such blunders regarding stations names or authority.

Rubbings

The illustrated station data sheet provides an area to accommodate a rubbing. With the paper held on top of the monument’s disk, a pencil is run over it in a zigzag pattern producing a positive image of the stamping. This method is a bit more awkward than simply copying the information from the disk onto the data sheet, but it does have the advantage of ensuring the station was actually visited and that the stamping was faithfully recorded. Such rubbings or close-up photographs are required by the provisional Federal Geodetic Control Committee (FGCC) *Geometric Geodetic Accuracy Standards and Specifications for Using GPS Relative Positioning Techniques* for all orders of GPS surveys.

Photographs

The use of photographs is growing as a help for the perpetuation of monuments. It can be convenient to photograph the area around the mark as well as the monument itself. These exposures can be correlated with a sketch of the area. Such a sketch can show the spot where the photographer stood and the directions toward which the pictures were taken. The photographs can then provide valuable information in locating monuments, even if they are later obscured. Still, the traditional ties to prominent features in the area around the mark are the primary agent of their recovery.

Quad Sheet Name

Providing the name of the appropriate state, county, and U.S. Geological Survey (USGS) quad sheet helps to correlate the station data sheet with the project map. The year the mark was monumented, the monument description, the station name, and the “to-reach” description all help to associate the information with the correct official control data sheet and, most importantly, the correct station coordinates.

To-Reach Descriptions

The description of the route to the station is one of the most critical documents written during the reconnaissance. Even though it is difficult to prepare the information in unfamiliar territory and although every situation is somewhat different, there are some guidelines to be followed. It is best to begin with the general location of the station with respect to easily found local features.

The description in Figure 7.1 relies on a road junction, a guard station, and a local business. After defining the general location of the monument, the description should recount directions for reaching the station. Starting from a prominent location, the directions should adequately describe the roads and junctions. Where the route is difficult or confusing, the reconnaissance team should not only describe the junctions and turns needed to reach a station; it is wise to also mark them with lath and flagging, when possible. It is also a good idea to note gates. Even if they are open during reconnaissance, they may be locked later. When turns are called for, it is best to describe not only the direction of the turn, but the new course too. For example, in the description in Figure 7.1 the turn onto the dim road from the Tee River Road is described to the, "left (northwest)." Roads and highways should carry both local names and designations found on standard highway maps. For example in Figure 7.1, Tee River Road is also described as State Highway 20.

The "to-reach" description should certainly state the mileages as well as the travel times where they are appropriate, particularly where packing-in is required. Land ownership, especially if the owner's consent is required for access, should be mentioned. The reconnaissance party should obtain the permission to enter private property and should inform the GPS observer of any conditions of that entry. Alternate routes should be described where they may become necessary. It is also best to make special mention of any route that is likely to be difficult in inclement weather.

Where helicopter access is anticipated, information about the duration of flights from point to point, the distance of landing sites from the station, and flight time to fuel supplies should be included on the station data sheet.

Flagging and Describing the Monument

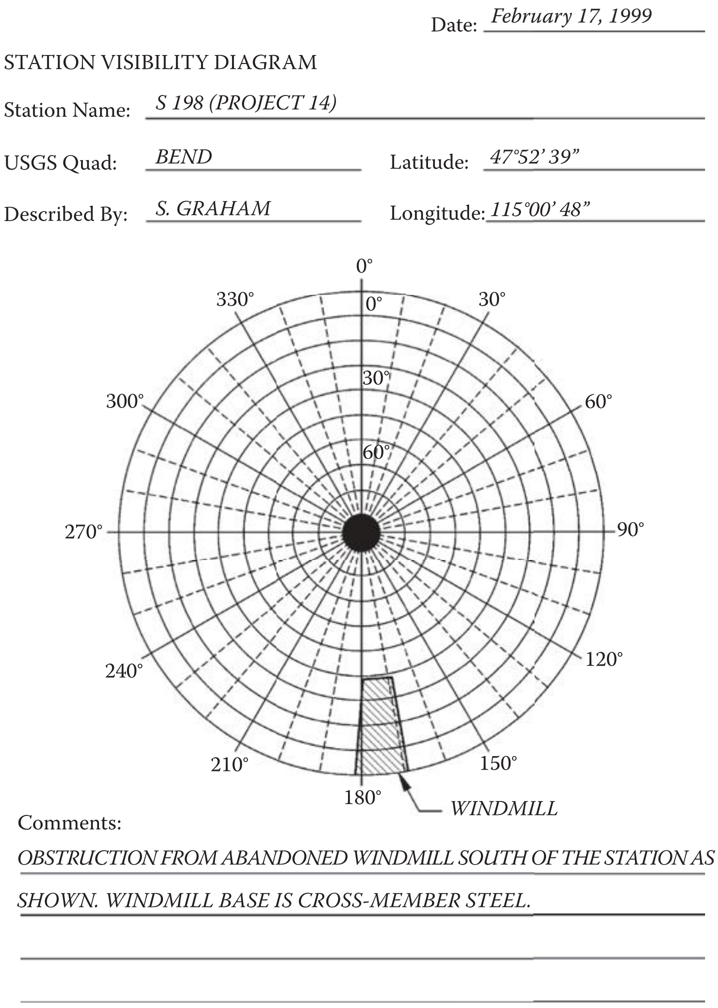
Flagging the station during reconnaissance may help the observer find the mark more quickly. On the station data sheet, the detailed description of the location of the station with respect to roads, fence lines, buildings, trees, and any other conspicuous features should include measured distances and directions. A clear description of the monument itself is important. It is wise to also show and describe any nearby marks, such as R.M.s, that may be mistaken for the station or aid in its recovery. The name of the preparer, a signature, and the date round out the initial documentation of a GPS station.

VISIBILITY DIAGRAMS

Obstructions above the mask angle of a GPS receiver must be taken into account in finalizing the observation schedule. A station that is blocked to some degree is not necessarily unusable, but its inclusion in any particular session is probably contingent on the position of the specific satellites involved.

An Example

The diagram in Figure 7.2 is widely used to record such obstructions during reconnaissance. It is known as a *station visibility diagram*, a *polar plot,* or a *skyplot.* The concentric circles are meant to indicate 10° increments along the upper half of the celestial sphere, from the observer's horizon at 0° on the perimeter, to the observer's zenith at 90° in the center. The hemisphere is cut by the observer's meridian, shown as a line from 0° in the north to 180° in the south. The prime vertical is signified as



Date: *February 17, 1999*

STATION VISIBILITY DIAGRAM

Station Name: *S 198 (PROJECT 14)*

USGS Quad: *BEND*

Latitude: *47°52' 39"*

Described By: *S. GRAHAM*

Longitude: *115°00' 48"*

Comments:

OBSTRUCTION FROM ABANDONED WINDMILL SOUTH OF THE STATION AS SHOWN. WINDMILL BASE IS CROSS-MEMBER STEEL.

FIGURE 7.2 Station Visibility Diagram

the line from 90° in the east to 270° in the west. The other numbers and solid lines radiating from the center, every 30° around the perimeter of the figure, are azimuths from north and are augmented by dashed lines every 10°.

Drawing Obstructions

Using a compass and a clinometer, a member of the reconnaissance team can fully describe possible obstructions of the satellite's signals on a visibility diagram. By standing at the station mark and measuring the azimuth and vertical angle of points outlining the obstruction, the observer can plot the object on the visibility diagram. For example, a windmill base is shown on in Figure 7.2 as a cross-hatched figure. It has been drawn from the observer's horizon up to 37° in vertical angle from 168°, to about 182° in azimuth at its widest point. This description by approximate angular values is entirely adequate for determining when particular satellites may be blocked at this station.

For example, suppose a 1-hour session from 9:10 to 10:10, illustrated in Table 7.2, was under consideration for the observation on station S 198. The station visibility chart might motivate a careful look at space vehicle (SV) pseudorandom noise (PRN) 16. Twenty minutes into the anticipated session, at 9:30 SV 16 has just risen above the 15° mask angle. Under normal circumstances, it would be available at station S 198, but it appears from the polar plot that the windmill will block its signals from reaching the receiver. In fact, the signals from SV 16 will apparently not reach station S 198 until sometime after the end of the session at 10:10.

Working around Obstructions

Under the circumstances, some consideration might be given to observing station S 198 during a session when none of the satellites would be blocked. However, the 9:10 to 10:10 session may be adequate after all. Even if SV 16 is completely blocked, the remaining 5 satellites will be unobstructed and the constellation still will have a relatively low position dilution of precision (PDOP). Still, the analysis must be carried to other stations that will be occupied during the same session. The success of the measurement of any baseline depends on common observations at both ends of the line. Therefore, if the signals from SV 16 are garbled or blocked from station S 198, any information collected during the same session from that satellite at the other end of a line that includes S 198 will be useless in processing the vector between those two stations.

But the material of the base of the abandoned windmill has been described on the visibility diagram as cross-membered steel, so it is possible that the signal from SV 16 will not be entirely obstructed during the whole session. There may actually be more concern of multipath interference from the structure than that of signal availability. One strategy for handling the situation might be to program the receiver at S 198 to ignore the signal from SV 16 completely if the particular receiver allows it.

The visibility diagram (Figure 7.2) and the azimuth-elevation table (Table 7.2) complement each other. They provide the field supervisor with the data needed to make informed judgments about the observation schedule. Even if the decision is taken to include station S 198 in the 9:10 to 10:10 session as originally planned, the supervisor will be forewarned that the blockage of SV 16 may introduce a bit of weakness at that particular station.

Approximate Station Coordinates

The latitude and longitude given on the station visibility diagram should be understood to be approximate. It is sometimes a scaled coordinate, or it may be taken from another source. In either case, its primary role is as input for the receiver at the beginning of its observation. The coordinate need only be close enough to the actual position of the receiver to minimize the time the receiver must take to lock onto the constellation of satellites it expects to find.

TABLE 7.2
Satellite Azimuth and Elevation Table

Time	El Az	El Az	El Az	El Az	El Az	El Az	El Az	El Az	PDOP
				Constellation of 5 SVs					
SV	3	12	13	20	24				
8:50	54 235	74 274	44 28	16 308	68 169				4.8
9:00	51 229	74 255	40 32	20 310	72 163				5.7
9:10	47 224	72 238	37 35	23 311	77 153				4.9
9:20	43 219	68 226	33 38	27 313	80 134				4.0
				Constellation of 6 SVs					
SV	3	12	13	16	20	24			
9:30	39 215	64 218	29 41	16 179	31 314	81 102			2.1
9:40	35 212	59 213	26 45	19 176	36 314	80 73			2.3
9:50	31 209	54 209	23 48	23 173	40 315	76 57			2.4
10:00	27 207	49 206	19 52	27 170	44 314	72 49			2.5
10:10	23 204	44 204	16 55	30 167	48 314	67 45			2.5

Multipath

The multipath condition is by no means unique to GPS. When a transmitted television signal reaches the receiving antenna by two or more paths, the resulting variations in amplitude and phase cause the picture to have ghosts. This kind of scattering of the signals can be caused by reflection from land, water, or man-made structures. In GPS, the problem can be particularly troublesome when signals are received from satellites at low elevation angles; hence the general use of a 15° to 20° mask angle. The use of choke ring antennas to mediate multipath may also be considered.

It is also wise, where it is possible, to avoid using stations that are near structures likely to be reflective or to scatter the signal. For example, chain-link fences that are found hard against a mark can cause multipath by forcing the satellite's signal to pass through the mesh to reach the antenna. The elevation of the antenna over the top of the fence with a survey mast is often the best way to work around this kind of obstruction. Metal structures with large flat surfaces are notorious for causing multipath problems. A long train moving near a project point could be a potential problem, but vehicles passing by on a highway or street usually are not, especially if they go by at high speed. It is important, of course, to avoid parked vehicles. It is best to remind new GPS observers that the survey vehicle should be parked far enough from the point to avert any multipath. A good way to handle these unfavorable conditions is to set an offset point.

Point Offsets

An offset must, of course, stand far enough away from the source of multipath or an attenuated signal to be unaffected. However, the longer the distance from the originally desired position the more important the accuracy of the bearing and distance between that position and the offset becomes. Recording the tie between the two correctly is crucial to avoid misunderstanding after the work is completed. Some receivers allow input of the information directly into the observations recorded in the receiver or data logger. However, during a control survey it is best to also record the information in a field book.

Offsets in GPS control surveying are an instance where conventional surveying equipment and expertise are necessary. Clearly the establishment of the tie requires a position for the occupation of the instrument, that is, a total station, and a position for the establishment of it orientation, that is, an azimuth (Figure 7.3). It is best to establish three intervisible rather than two points—one to occupy and two azimuth marks. This approach makes it possible to add a redundant check to the tie. The positions on these two, or three, points may be established by setting monuments and performing static observations on them all. Alternatively azimuthal control may be established by astronomic observations.

Look for Multipath

Both the GPS field supervisor and the reconnaissance team should be alert to any indications on the station visibility diagram that multipath may be a concern. Before the observations are done, there is nearly always a simple solution. Discovering multipath in the signals after the observations are done is not only frustrating, but often expensive.

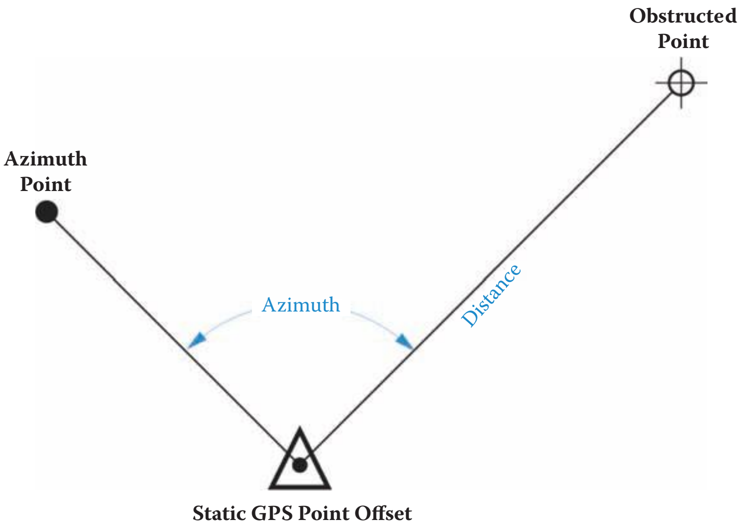


FIGURE 7.3 Static Point Offset

MONUMENTATION

The monumentation set for GPS projects varies widely and can range from brass tablets to aerial premarks, capped rebar or even pin flags. The objective of most station markers is to adequately serve the client's subsequent use. However, the time, trouble, and cost in most high-accuracy GPS work warrant the most permanent, stable monumentation.

Many experts predict that GPS will eventually make monumentation unnecessary. The idea foresees GPS receivers in constant operation at well-known master stations that will allow surveyors with receivers to determine highly accurate relative positions with such speed and ease that monumentation will be unnecessary. The idea may prove prophetic, but for now monumentation is an important part of most GPS projects. The suitability of a particular type of monument is an area still most often left to the professional judgment of the surveyors involved.

The FGCC recommends the use of traditional metal disks set in rock outcroppings, bridge abutments, or other large structural elements where possible. A three-dimensional (3-D) rod mark is approved as an alternative by the federal committee. It is described in detail in Appendix H of *Geometric Geodetic Accuracy Standards and Specifications for Using GPS Relative Positioning Techniques*.

LOGISTICS

Scheduling

Once all the station data sheets, visibility diagrams, and other field notes have been collected, the schedule can be finalized for the first observations. There will almost certainly be changes from the original plan. Some of the anticipated control stations may be unavailable or obstructed, some project points may be blocked, too difficult to reach, or simply not serve the purpose as well as a control station at an alternate location. When the final control has been chosen, the project points have been monumented, and the reconnaissance has been completed, the information can be brought together with some degree of certainty that it represents the actual conditions in the field.

Now that the access and travel time, the length of vectors, and the actual obstructions are more certainly known, the length and order of the sessions can be solidified. Despite all the care and planning that goes into preparing for a project, unexpected changes in the satellites' orbits or health can upset the best schedule at the last minute. It is always helpful to have a backup plan.

The receiver operators usually have been involved in the reconnaissance and are familiar with the area and many of the stations. Even though an observer may not have visited the particular stations scheduled for him, the copies of the project map, appropriate station data sheets, and visibility diagrams will usually prove adequate to their location.

OBSERVATION

When everything goes as planned a GPS observation is uneventful. However, even before the arrival of the receiver operator at the control or project point the session can get off-track. The simultaneity of the data collected at each end of a baseline is critical to the success of any measurement in GPS. When a receiver occupies a master station throughout a project there need be little concern on this subject. But most static applications depend on the sessions of many mobile receivers beginning and ending together.

Arrival

The number of possible delays that may befall an observer on the way to a station are too numerous to mention. With proper planning and reconnaissance, the observer will likely find that there is enough time for the trip from station to station and that sufficient information is on hand to guide him to the position, but this too cannot be guaranteed. When the observer is late to the station, the best course is usually to set up the receiver quickly and collect as much data as possible. The baselines into the late station may or may not be saved, but they will certainly be lost if the receiver operator collects no information at all. It is at times like these that good communication between the members of the GPS team is most useful. For example, some of the other observers in the session may be able to stay on their station a bit longer with the late arrival and make up some of the lost data. Along the same line, it is usually a good policy for those operators who are to remain on a station for two consecutive sessions to collect data as long as possible, while still leaving themselves enough time to reset between the two observation periods.

Setup

Centering an instrument over the station mark is always important. However, the centimeter-level accuracy of static GPS gives the centering of the antenna special significance. It is ironic that such a sophisticated system of surveying can be defeated from such a commonplace procedure. A tribrach with an optical plummet or any other device used for centering should be checked and, if necessary, adjusted before the project begins. With good centering and leveling procedures, an antenna should be within a few millimeters of the station mark. The FGCC's provisional specifications require that the antenna's centering be checked with a plumb bob at each station for surveys of the AA, A, and B orders.

Unfortunately, the centering of the antenna over the station does not ensure that its phase center is properly oriented. The contours of equal phase around the antenna's electronic center are not themselves perfectly spherical. Part of their eccentricity can be attributed to unavoidable inaccuracies in the manufacturing process. To compensate for some of this offset, it is a good practice to rotate all antennas in a session to the same direction. Many manufacturers provide reference marks on their antennas so that each one may be oriented to the same azimuth. That way they are expected to maintain the same relative position between their physical and electronic centers when observations are made.

The antenna's configuration also affects another measurement critical to successful GPS surveying: the height of the instrument. The frequency of mistakes in this important measurement is remarkable. Several methods have been devised to focus special attention on the height of the antenna. Not only should it be measured in both feet and meters, it should also be measured immediately after the instrument is setup and just before tearing it down to detect any settling of the tripod during the observation.

Height of Instrument

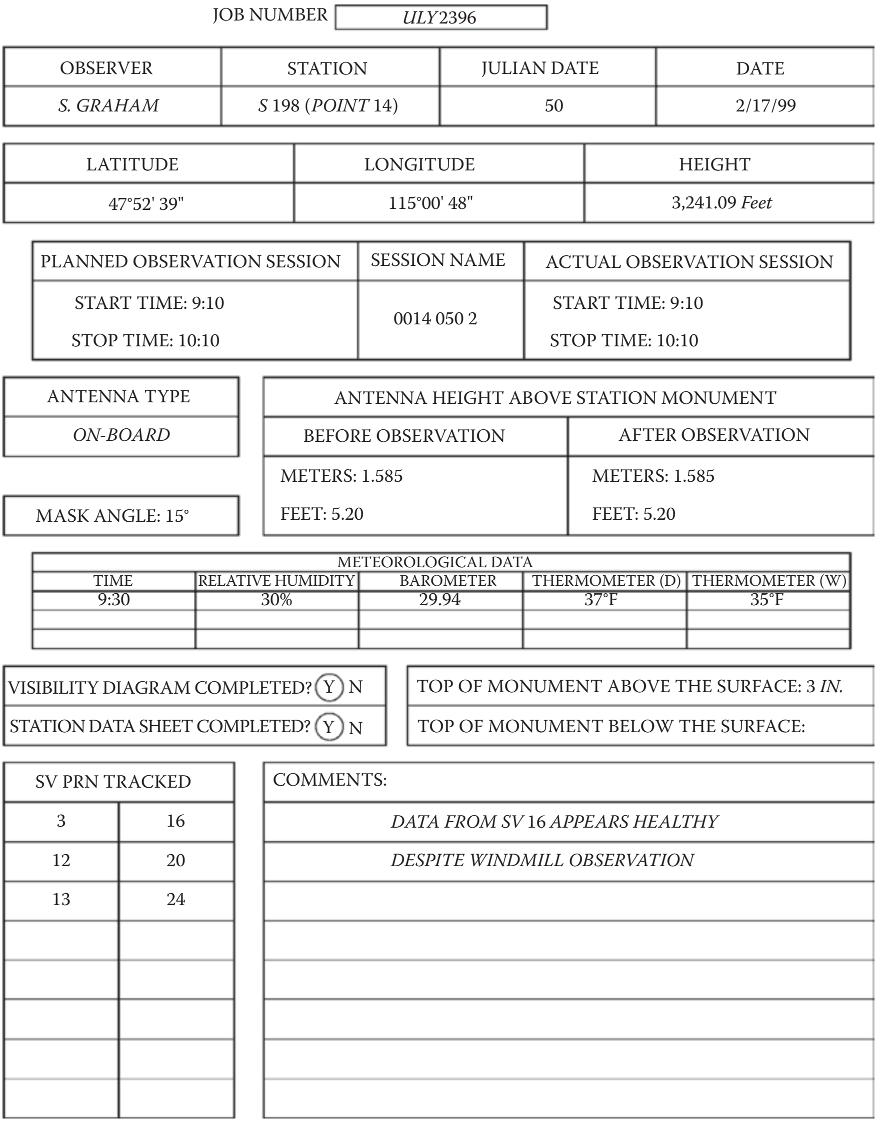
The measurement of the height of the antenna in a GPS survey is often not made on a plumb line. A tape is frequently stretched from the top of the station monument to some reference mark on the antenna or the receiver itself. Some GPS teams measure and record the height of the antenna to more than one reference mark on the ground plane. These measurements are usually mathematically corrected to plumb.

The care ascribed to the measurement of antenna heights is due to the same concern applied to centering. GPS has an extraordinary capability to achieve accurate heights, but those heights can be easily contaminated by incorrect H.I.s.

Observation Logs

Most GPS operations require its receiver operators to keep a careful log of each observation. Usually written on a standard form, these field notes provide a written record of the measurements, times, equipment, and other data that explains what actually occurred during the observation itself. It is difficult to overestimate the importance of this information. It is usually incorporated into the final report of the survey, the archives, and any subsequent effort to blue-book the project. However, the most immediate use of the observation log is in evaluation of the day's work by the on-site field supervisor.

An observation log may be organized in a number of ways. The log illustrated in Figure 7.4 is one method that includes some of the information that might be used to document one session at one station. Of course, the name of the observer and the station must be included, and while the date need not be expressed in both the Julian



JOB NUMBER *ULY* 2396

OBSERVER	STATION	JULIAN DATE	DATE
S. GRAHAM	*S* 198 (*POINT* 14)	50	2/17/99

LATITUDE	LONGITUDE	HEIGHT
47°52' 39"	115°00' 48"	3,241.09 *Feet*

PLANNED OBSERVATION SESSION	SESSION NAME	ACTUAL OBSERVATION SESSION
START TIME: 9:10 STOP TIME: 10:10	0014 050 2	START TIME: 9:10 STOP TIME: 10:10

ANTENNA TYPE
ON-BOARD

MASK ANGLE: 15°

ANTENNA HEIGHT ABOVE STATION MONUMENT	
BEFORE OBSERVATION	AFTER OBSERVATION
METERS: 1.585 FEET: 5.20	METERS: 1.585 FEET: 5.20

METEOROLOGICAL DATA				
TIME	RELATIVE HUMIDITY	BAROMETER	THERMOMETER (D)	THERMOMETER (W)
9:30	30%	29.94	37°F	35°F

VISIBILITY DIAGRAM COMPLETED? [✓ Y] N

STATION DATA SHEET COMPLETED? [✓ Y] N

TOP OF MONUMENT ABOVE THE SURFACE: 3 *IN.*

TOP OF MONUMENT BELOW THE SURFACE:

SV PRN TRACKED	
3	16
12	20
13	24

COMMENTS:
DATA FROM SV 16 *APPEARS HEALTHY*
DESPITE WINDMILL OBSERVATION

FIGURE 7.4 Observation Log

and Gregorian calendars that information may help in quick cataloging of the data. The approximate latitude, longitude, and height of the station are usually required by the receiver as a reference position for its search for satellites. The date of the planned session will not necessarily coincide with the actual session observed. The observer's arrival at the point may have been late, or the receiver may have been allowed to collect data beyond the scheduled end of the session.

There are various methods used to name observation sessions in terminology that is sensible to computers. A widely used system is noted here. The first four digits are the project point's number. In this case it is point 14 and is designated 0014. The next three digits are the Julian day of the session, in this case it is day 50, or 050. Finally, the session illustrated is the second of the day, or 2. Therefore, the full session name is 0014 050 2.

Whether onboard or separate, the type of the antenna used and the height of the antenna are critical pieces of information. The relation of the height of the station to the height of the antenna is vital to the station's later utility. The distance that the top of the station's monument is found above or below the surface of the surrounding soil is sometimes neglected. This information can not only be useful in later recovery of the monument, but can also be important in the proper evaluation of photo-control panel points.

Weather

The meteorological data is useful in modeling the atmospheric delay. This information is required at the beginning, middle, and end of each session of projects that are designed to satisfy the FGCC's provisional specifications for the AA and A orders of accuracy. Under those circumstances, measurement of the atmospheric pressure in millibars, the relative humidity, and the temperature in degrees Centigrade are expected to be included in the observation log. However, the general use is less stringent. The conditions of the day are observed, and unusual changes in the weather are noted.

A record of the satellites that is actually available during the observation and any comments about unique circumstances of the session round out the observation log.

DAILY PROGRESS EVALUATION

The planned observation schedules of a large GPS project usually change daily. The arrangement of upcoming sessions is often altered based on the success or failure of the previous day's plan. Such a regrouping follows evaluation of the day's data.

This evaluation involves examination of the observation logs as well as the data each receiver has collected. Unhealthy data, caused by cycle slips or any other source, are not always apparent to the receiver operator at the time of the observation. Therefore a daily quality control check is a necessary preliminary step before finalizing the next day's observation schedule.

Some field supervisors prefer to actually compute the independent baseline vectors of each day's work to ensure that the measurements are adequate. Neglecting the daily check could leave unsuccessful sessions undiscovered until the survey was thought to be completed. The consequences of such a situation could be expensive.

REAL-TIME KINEMATIC (RTK) AND DIFFERENTIAL GPS (DGPS) OBSERVATIONS

Some components of static GPS control methods are useful in RTK and DGPS work just as they are described above. However, there are some, such as station diagrams, observation logs, and to-reach descriptions that would rarely if ever be necessary in high-production dynamic work. And finally, there are aspects, though handled a bit differently in both categories of work, that have utility in each. One such technique is offsetting points to avoid multipath and signal attenuation.

Point Offsets

The need to offset points is much more prevalent in DGPS and RTK surveying (Figure 7.5). However, the methods of dealing with it are more varied. DGPS and RTK surveying generally have lower accuracy requirements than does static control work, and therefore the establishment of the tie between the offset and the originally desired position need not be so stringent. However, when circumstances arise that require a particularly accurate tie, the same conventional surveying procedures mentioned for static point offsets may be useful.

The offset point must be established far enough from the original position to avoid an obstructed signal, but close enough to prevent unacceptable positioning error. While the calculation of the allowable vertical and horizontal measurement errors can be done trigonometrically, the measurements themselves will be different than those for an offset point in a static survey. For example, rather than a total station point and an azimuth point, a magnetic fluxgate digital compass and laser may be used to measure the tie from the offset point to the original point. It is worth noting that magnetic declination must be accommodated and metal objects avoided when doing magnetic work. A declination calculator is available online at http://www.ngdc.noaa.gov/seg/geomag/jsp/Declination.jsp. And also such internal compasses should be carefully checked before they are relied upon.

The length of the tie may be measured by an external laser, a laser cabled directly into the GPS receiver, or even a tape and clinometer. Lasers are much more convenient since they can be used to measure longer distances more reliably and taping requires extra field crew members. And rather than recording the bearing and distance in a field book for postprocessing in DGPS or RTK the information

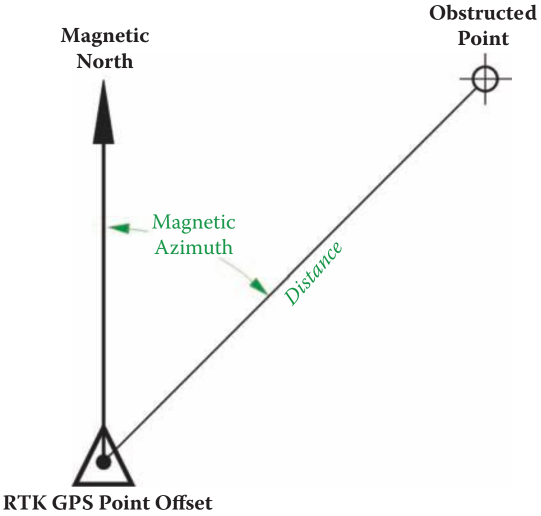


FIGURE 7.5 DGPS or RTK Point Offset

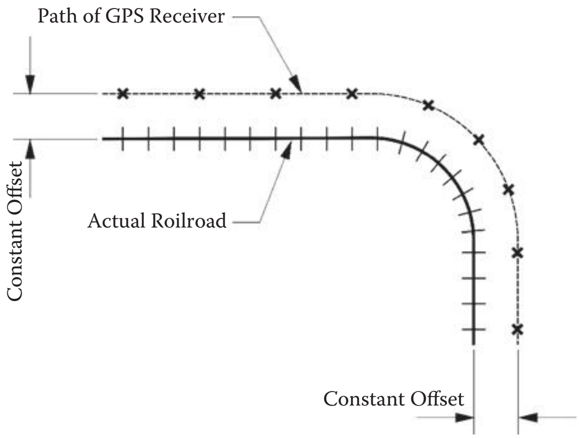


FIGURE 7.6 Line Offset

is usually stored directly in the data collector. In fact, often the receiver's real-time processor can combine the measured distance and the direction of the *sideshot* with the receiver's position and calculate the coordinate of the originally desired position on-the-fly (OTF).

Dynamic Lines

A technique unique to RTK and DGPS and used especially in mobile GPS application is the creation of dynamic lines. The GPS receiver typically moves along a route to be mapped logging positions at predetermined intervals of time or distance. These points can then be joined together to create a continuous line. Obstructions along the route present a clear difficulty for this procedure. Points may be in error, or lost completely due to multipath or signal attenuation. Also, in choosing the epoch interval the capacity of the receiver's memory must be considered, especially when long lines are collected. If the interval chosen is too short the receiver's storage capacity may be overwhelmed. If the interval is too long important deflections along the way may be missed.

Where it is impossible or unsafe to travel along the line to be collected in the field the dynamic line may be collected with a consistent offset (Figure 7.6). This technique is especially useful in the collection roads and railroads where it is possible to estimate the offset with some certainty due to the constant width of the feature. It is also possible, of course, to collect routes with individual discrete points with short occupations where that approach recommends itself.

Planning

Multipath and signal attenuation are particularly troublesome for the dynamic GPS of DGPS and RTK work. While the visibility diagrams mentioned earlier are not directly applicable it is nevertheless prudent to plan the work so that at least 5 GPS satellites are available above the mask angle in the area where data is to be collected. It also may be useful to lower the mask angle and/or reduce the signal-to-noise ratio

(SNR) when conditions warrant it. There are, of course, trade offs to such strategies that may involve a reduction in positional integrity. The balance between accuracy and productivity is always a consideration.

A Few RTK Procedures

As mentioned earlier redundancy in RTK work can be achieved by occupying each newly established position twice, and it is best if the second occupation is done using a different base station than was used to control the first. The control points occupied by base stations should not be too close to one another. A minimum of 300 meters is a good rule of thumb. Each time the base is set up and before it is taken down it is best to do a check shot on at least one known control point to verify the work. And in order to ensure that the GPS constellation during the second occupation differs substantially from that of the first, it is best if the second occupation takes place not less than 4 hours and not more than 8 hours later or earlier than the first.

To ensure that the centering is correct during the short occupations of RTK it is best if a bipod is used with a fixed height rod to eliminate the possibility of incorrect height of instrument measurement corrupting the results. Concerning heights, if orthometric heights in real-time are desired, a geoidal model is required and it is best if it is the most recent. However, please note that work retraced with a different geoidal model than was used initially will likely show vertical differences at the reoccupied points.

Some rover configurations facilitate *in-fill* surveys. In other words, when the correction signals from the base station fail to reach the rover the collected data is stored in the memory of the receiver for postprocessing after the work is completed.

Site Calibration

The area of interest that is the project area, covered by an RTK survey, is usually relatively small and defined. Typically a *site calibration* is performed to prepare such a GPS project to be done using plane coordinates. A site calibration establishes the relationship between geographical coordinates—latitude, longitude, and ellipsoidal height—with plane coordinates—northing, easting, and orthometric heights across the area. In the final analysis the relationship is expressed in three dimensions: translation, rotation, and scale. Because of the inevitable distortion that a site calibration must model one of the prerequisites for such *localization* is the enclosure of the area by the control stations that will be utilized during the work.

In the horizontal plane the method of using plane coordinates on an imaginary flat reference surface with northings and eastings, or *x*- and *y*-coordinates assumes a flat earth. That is incorrect of course but a viable simplification if the area is small enough and the distortion is negligible. Such local tangent planes fixed at discrete points, control points, by GPS site calibration have been long used by land surveyors. Such systems demand little if any manipulation of the field observations, and once the coordinates are derived they can be manipulated by straightforward plane trigonometry. In short, Cartesian systems are simple and convenient.

However, there are difficulties as the area grows as mentioned in Chapter 5. For example, typically each of these planes has a unique local coordinate system derived from its own unique site calibration. The axes, the scale, and the rotation of each one of these individual local systems will not be the same as those elements of its neighbor's coordinate system. Therefore, when a site calibration is done and a local flat plane coordinate system is created, it is important to keep all of the work in that system inside limits created by the control points used in its creation. In the simplest case, a single point calibration, a flat plane is brought tangent to the earth at one point, but a more typical approach is the utilization of three or four points enclosing the area of interest to be covered by the independent *local coordinate system.* Working outside of the limits created by those points should be avoided as it involves working where the distortion has not been modeled.

It might be said that a site calibration is a *best fit* of a plane onto a curved surface. It distributed the inevitable distortion in both the horizontal and vertical planes. The vertical aspect is particularly important. It is called upon to adjust the measured GPS ellipsoid heights to a desired local vertical datum. Therefore, it must account for undulations in the geoid because the separation between the ellipsoid and geoidal models is seldom if ever consistent over the project area. The separation is not consistent and usually can be modeled approximately as a trend across the area of interest so that the site calibration typically produces an inclined plane in the vertical aspect. Toward that end the set of control points used to establish the site calibration must have both geographical coordinates—latitude, longitude and ellipsoidal height- and plane coordinates—northing, easting, and orthometric heights in the desired local system. It is best if these control points are from the National Spatial Reference System (NSRS) when possible that enclose the project and are distributed evenly around its boundary.

PROCESSING

POSTPROCESSING GPS STATIC CONTROL SURVEYS

In many ways, processing is the heart of a GPS operation. Some processing should be performed on a daily basis during a GPS project. Blunders from operators, noisy data, and unhealthy satellites can corrupt entire sessions. And left undetected, such dissolution can jeopardize an entire survey. But with some daily processing, these weaknesses in the data can be discovered when they can still be eliminated with a timely amendment of the observation schedule.

But even after blunders and noisy data have been removed from the observation sets, GPS measurements are still composed of fundamentally biased ranges. Therefore, GPS data- processing procedures are really a series of interconnected computerized operations designed to remove these more difficult biases and extract the true ranges.

The biases originate from a number of sources: imperfect clocks, atmospheric delays, cycle ambiguities in carrier phase observations, and orbital errors. If a bias has a stable, well- understood structure, it can be estimated. In other cases, dual-frequency observations can be used to measure the bias directly, as in the ionospheric delay, or a model may be used to predict an effect, as in tropospheric delay. But one of the most effective strategies in eradicating biases is called *differencing.*

Correlation of Biases

When two or more receivers observe the same satellite constellation simultaneously, a set of correlated vectors are created between the co-observing stations. Most GPS practitioners use more than two receivers. Therefore, most GPS networks consist of many sets of correlated vectors for every separate session. The longest baselines between stations on the earth are usually relatively short when compared with the more than 20,000-km distances from the receivers to the GPS satellites. Therefore, even when several receivers are set up on widely spaced stations, as long as they collect their data simultaneously from the same constellation of satellites, they will record very similar errors. In other words, their vectors will be correlated. It is the simultaneity of observation and the resulting correlation of the carrier phase observables that make the extraordinary GPS accuracies possible. Biases that are correlated linearly can be virtually eliminated by differencing the data sets of a session.

QUANTITY OF DATA

Organization Is Essential

One of the difficulties of GPS processing is the huge amount of data that must be managed. For example, when even one single-frequency receiver with a 1-second sampling rate tracks 1 GPS satellite for an hour, it collects about 0.15 Mb of data. However, a more realistic scenario involves four receivers observing 6 satellites for 3600 epochs. There can be 4 X 6 X 3600 or 86,400 carrier phase observations in such a session. In other words, a real-life GPS survey with many sessions and many baselines creates a quantity of data in the gigabyte range. Some sort of structured approach must be implemented to process such a huge amount of information in a reasonable amount of time.

File Naming Conventions

One aspect of that structure is the naming conventions used to head GPS receivers' observation files. Many manufacturers recommend a file naming format that can be symbolized by *pppp-ddd-s.yyf*. The first letters, (*pppp*), of the file name indicate the point number of the station occupied. The day of the year, or Julian date, can be accommodated in the next three places (*ddd*), and the final place left of the period is the session number (*s*). The year (*yy*) and the file type (*f*) are sometimes added to the right of the period.

DOWNLOADING

The first step in GPS data processing is downloading the collected data from the internal memory of the receiver itself into a PC or laptop computer. When the observation sessions have been completed for the day, each receiver, in turn, is cabled to the computer and its data transferred. Nearly all GPS systems used in surveying are PC-compatible and can accommodate postprocessing in the field. But none can protect the user from a failure to back up this raw observational data onto some other form of semipermanent storage.

Making Room

Receiver memory capacity is usually somewhat limited, and older data must be cleared to make room for new sessions. Still, it is a good policy to create the necessary space with the minimum deletion and restrict it to only the oldest files in the receiver's memory. In this way the recent data can be retained as long as possible, and the data can provide an auxiliary back-up system. But when a receiver's memory is finally wiped of a particular session, if redundant raw data are not available reobservation may be the only remedy.

Most GPS receivers record data internally. The Navigation message, meteorological data, the observables, and all other raw data are usually in a manufacturer-specific, binary form. They are also available in RINEX (Receiver Independent Exchange Format). These raw data are usually saved in several distinct files. For example, the PC operator will likely find that the phase measurements downloaded from the receiver will reside in one file and the satellite's ephemeris data in another, and so forth. Likewise, the measured pseudorange information may be found in its own dedicated file, the ionospheric information in another, and so on. The particular division of the raw data files is designed to accommodate the suite of processing software and the data management system that the manufacturer has provided its customers, so each will be somewhat unique.

CONTROL

All postprocessing software suites require control. GPS static work often satisfies that requirement by inclusion of NGS monuments in the network design. In this approach the control stations are occupied by the surveyor building the network. There is an alternative. Continuously operating reference stations, CORS, already occupy many NGS control monuments and constantly collect observations; their data can be used to support carrier-phase static surveys as well as dynamic GPS work. The most direct method is to download the CORS data files posted on the Internet. The CORS data collected during the time of the survey can be combined with those collected in the field. They can be used to postprocess the baselines and derive positions for the new points.

A difficulty arises when the density of the CORS in and around the project area is not adequate to support the required accuracy of the survey. This problem occurs in both dynamic and static GPS surveying applications. Real-Time Networks are a useful strategy for dealing with this control spacing issue in RTK and DGPS work as discussed in Chapter 6. However, in static control surveys outside of urban or highly developed areas the sparseness of CORS may obviate their use. Still, the situation is changing around the world. While there may be distances of up to 100 km between CORS in the United States, GEONET in Japan has a spacing between stations of approximately 25 km and in Hong Kong the distance may be as little as 10 km.

THE FIRST POSITION

Baseline processing is usually begun with a point position solution at each end from pseudoranges. These differential code estimations of the approximate position of each receiver antenna can be thought of as establishing a search area, a three-dimensional volume of uncertainty at each receiver containing its correct position. The size of this search area is defined by the accuracy of the code solution, which also affects the computational time required to find the correct position among all the other potential solutions.

TRIPLE DIFFERENCE

The next step, usually the triple-difference, utilizes the carrier phase observable. Triple differences have several features to recommend them for this stage in the processing. They can achieve rather high accuracy even before cycle slips have been eliminated from the data sets, and they are insensitive to integer ambiguities in general.

Components of a Triple Difference

A triple difference is created by differencing two double differences at each end of the baseline. Each of the double differences involves two satellites and two receivers. A triple difference considers two double differences over two consecutive epochs. In other words, triple differences are formed by sequentially differencing double differences in time. For example, two triple differences can be created using double differences at epochs 1, 2, and 3. One is double difference 2 minus double difference 1. A second can be formed by double difference 3 minus double difference 2.

Since two receivers are recording the data from the same 2 satellites during two consecutive epochs across a baseline, a triple difference can temporarily eliminate any concern about the integer cycle ambiguity, because the cycle ambiguity is the same over the two observed epochs. However, the triple difference cannot have as much information content as a double difference. Therefore, while receiver coordinates estimated from triple differences are usually more accurate than pseudorange solutions, they are less accurate than those obtained from double differences, especially fixed-ambiguity solutions. Nevertheless, the estimates that come out of triple-difference solutions refine receiver coordinates and provide a starting point for the subsequent double-difference solutions. They are also very useful in spotting and correcting cycle slips. They also provide a first estimate of the receiver's positions.

DOUBLE DIFFERENCE

The next baseline processing steps usually involve two types of double differences, called the *float* and the *fixed* solutions.

The Integer Ambiguity

Double differences have both positive and negative features. On the positive side, they make the highest GPS accuracy possible, and they remove the satellite and receiver clock errors from the observations. On the other hand, the integer cycle ambiguity, sometimes known simply as the *ambiguity*, cannot be ignored in the double difference. In fact, the fixed double-difference solution, usually the most accurate technique of all, requires the resolution of this ambiguity.

The integer cycle ambiguity, usually symbolized by *N*, represents the number of full phase cycles between the receiver and the satellite at the first instant of the receiver's lock-on. N does not change from the moment of the lock is achieved, unless there is a cycle slip. Unfortunately, N is also an unknown quantity at the beginning of any carrier phase observation.

The Float Solution

Once again estimation plays a significant role in finding the appropriate integer value that will correctly resolve the ambiguity for component pairs in double differencing. In this first try, there is no effort to translate the biases into integers. It is sometimes said that the integers are allowed to float; hence, the initial process is called the *float solution*. Especially when phase measurements for only one frequency, L1 or L2, are available, a sort of calculated guess at the ambiguity is the most direct route to the correct solution. Not just N, but a number of unknowns, such as clock parameters and point coordinates, are estimated in this geometric approach. However, all of these estimated biases are affected by unmodeled errors and that causes the integer nature of N to be obscured. In other words, the initial estimation of N in a float solution is likely to appear as a real number rather than an integer.

However, where the data are sufficient, these floating real-number estimates are very close to integers—so close that they can next be rounded to their true integer values in a second adjustment of the data. Therefore, a second double difference solution follows. The estimation of N that is closest to an integer and has the minimum standard error is usually taken to be the most reliable and is rounded to the nearest integer. Now, with one less unknown, the process is repeated and another ambiguity can be fixed, and so on.

The Fixed Solution

This approach leads to the *fixed solution* in which N can be held to integer values. It is usually quite successful in double differences over short baselines. The resulting fixed solutions most often provide much more accurate results than were available from the initial floating estimates.

CYCLE SLIP DETECTION AND REPAIR

However, this process can be corrupted by the presence of cycle slips. A cycle slip is a discontinuity in a receiver's continuous phase lock on a satellite's signal. The coded pseudorange measurement is immune from this difficulty, but the carrier beat phase is not. In other words, even when a fixed double-difference solution can provide the correct integer ambiguity resolution, the moment the data set is interrupted by a cycle slip that solution is lost.

There are actually two components to the carrier phase observable that should not change from the moment of a receiver's lock onto a particular satellite. First is the fractional initial phase at the first moment of the lock. The receiver is highly unlikely to acquire the satellite's signal precisely at the beginning of a wavelength. It will grab on at some fractional part of a phase, and this fractional phase will remain unchanged for the duration of the observation. The other unchanged aspect of a normal carrier phase observable is an integer number of cycles. The integer cycle ambiguity is symbolized by N. It represents the number of full phase cycles between the receiver and the satellite at the first instant of the receiver's lock on. The integer ambiguity ought to remain constant throughout an observation as well. But when there is a cycle slip, lock is lost, and by the time the receiver reacquires the signal, the normally constant integer ambiguity has changed.

Cycle Slip Causes

A power loss, a very low signal-to-noise ratio, a failure of the receiver software, a malfunctioning satellite oscillator, or any event that breaks the receiver's continuous reception of the satellite's signal causes a cycle slip. Most common, however, is an obstruction that is so solid it prevents the satellite signal from being tracked by the receiver. Under such circumstances, when the satellite reappears, the tracking resumes. The fractional phase may be the same as if tracking had been maintained, but the integer number of cycles is not.

Repairing Cycle Slips

Cycle slips are repaired in postprocessing. Both their location and their size must be determined; then the data set can be repaired with the application of a fixed quantity to all the subsequent phase observations. One approach is to hold the initial positions of the stations occupied by the receivers as fixed and edit the data manually. This has proven to work, but would try the patience of Job. Another approach is to model the data on a satellite-dependent basis with continuous polynomials to find the breaks and then manually edit the data set a few cycles at a time. In fact, several methods are available to find the lost integer phase value, but they all involve testing quantities.

One of the most convenient of these methods is based on the triple difference. It can provide an automated cycle slip detection system that is not confused by clock drift and, once least-squares convergence has been achieved, it can provide initial station positions even using the unrepaired phase combinations. They may still contain cycle slips but can nevertheless be used to process approximate baseline vectors. Then the residuals of these solutions are tested, sometimes through several iterations. Proceeding from its own station solutions, the triple difference can predict how many cycles will occur over a particular time interval. Therefore, by evaluating triple-difference residuals over that particular interval, it is not only possible to determine which satellites have integer jumps, but also the number of cycles that have actually been lost. In a sound triple-difference solution without cycle slips, the residuals are usually limited to fractions of a cycle. Only those containing cycle slips have residuals close to one cycle or larger. Once cycle slips are discovered, their correction can be systematic.

For example, suppose the residuals of one component double difference of a triple-difference solution revealed that the residual of satellite PRN 16 minus the residual of satellite PRN 17 was 8.96 cycles. Further suppose that the residuals from the second component double difference showed that the residual of satellite PRN 17 minus the residual of satellite PRN 20 was 14.04 cycles. Then one might remove 9 cycles from PRN 16 and 14 cycles from PRN 20 for all the subsequent epochs of the observation. However, the process might result in a common integer error for PRNS 16, 17, and 20. Still, small jumps of a couple of cycles can be detected and fixed in the double-difference solutions.

In other words, before attempting double difference solutions, the observations should be corrected for cycle slips identified from the triple difference solution. And even though small jumps undiscovered in the triple difference solution might remain in the data sets, the double difference residuals will reveal them at the epoch where they occurred.

However, some conditions may prevent the resolution of cycle slips down to the one-cycle level. Inaccurate satellite ephemerides, noisy data, errors in the receiver's initial positions, or severe ionospheric effects all can limit the effectiveness of cycle-slip fixing. In difficult cases, a detailed inspection of the residuals might be the best way to locate the problem.

Fixing the Integer Ambiguity and Obtaining Vector Solutions

When cycle slips have been finally eliminated, the ambiguities can be fixed to integer values. First, the standard deviations of the adjusted integers are inspected. If it is found to be significantly less than one cycle, they can be safely constrained to the nearest integer value. This procedure is only pertinent to double differences, since phase ambiguities are moot in triple differencing. The number of parameters involved can be derived by multiplying one less the number of receivers by one less the number of satellites involved in the observation. And for dual-frequency observations the phase ambiguities for L1 and L2 are best fixed separately. But such a program of constraints is not typical in long baselines where the effects of the ionosphere and inaccuracies of the satellite ephemerides make the situation less determinable.

In small baselines, however, where these biases tend to be virtually identical at both ends, integer fixing is almost universal in GPS processing. Once the integer ambiguities of one baseline are fixed, the way is paved for constraint of additional ambiguities in subsequent iterations. Then, step by step, more and more integers are set, until all that can be fixed have been fixed. Baselines of several thousand kilometers can be constrained in this manner.

With the integer ambiguities fixed, the GPS observations produce a series of vectors, the raw material for the final adjustment of the survey. The observed baselines represent very accurately determined relative locations between the stations they connect. However, the absolute position of the whole network is usually much less accurately known, although more accurate absolute positioning may be on the horizon for GPS. For now, there remains a considerable difference in the accuracy of relative and absolute positioning. A GPS survey is usually related to the rest of the world by translation of its Cartesian coordinates into ellipsoidal latitude, longitude, and height. Most users are less comfortable with the original coordinate results, given in the WGS84 coordinate system of the satellites themselves, than latitude, longitude, and height.

LEAST-SQUARES ADJUSTMENT

There are numerous adjustment techniques, but least-squares adjustment is the most precise and most commonly used in GPS. The foundation of the idea of least-squares adjustment is the idea that the sum of the squares of all the residuals applied to the GPS vectors in their final adjustment should be held to the absolute minimum. But minimizing this sum requires first defining those residuals approximately. Therefore, in GPS it is based on equations where the observations are expressed as a function of unknown parameters, but parameters that are nonetheless given approximate initial values. This is the process that has been described above. Then by adding the squares of the terms thus formed and differentiating their sum, the derivatives can be set equal to zero. For complex work like GPS adjustments, the least-squares method has the advantage that it allows for the smallest possible changes to original estimated values.

The solution strategies of GPS adjustments themselves are best left to particular suites of software. Suffice it to say that the single baseline approach, that is, a baseline-by-baseline adjustment, has the disadvantage of ignoring the actual correlation of the observations of simultaneously occupied baselines. An alternative approach involves a network adjustment approach where the correlation between the baselines themselves can be more easily taken into account. And while the computations are simpler for the baseline-by-baseline approach, cycle slips are more conveniently repaired in network adjustment.

For the most meaningful network adjustment, the endpoints of every possible baseline should be connected to at least two other stations. Thereby, the quality of the work itself can be more realistically evaluated. For example, the most common observational mistake, the mismeasured antenna height, is very difficult to detect when adjusting baselines sequentially, one at a time, but a network solution spots such blunders more quickly.

For example, most GPS postprocessing adjustment begins with a minimally constrained least-squares adjustment. That means that all the observations in a network are adjusted together with only the constraints necessary to achieve a meaningful solution; for example, the adjustment of a GPS network with the coordinates of only one station fixed. The purpose of the minimally constrained approach is to detect large mistakes, like a misidentifying one of the stations. The residuals from a minimally constrained work should come pretty close to the precision of the observations themselves. If the residuals are particularly large, there are probably mistakes; if they are really small the network itself may not be a strong as it should be.

This minimally constrained solution is usually followed by an over-constrained solution. In other words, a least-squares adjustment where the coordinate values of selected control station are held fixed.

It is important to note that the downside of least-squares adjustment is its tendency to spread the effects of even one mistake throughout the work. In other words, it can cause large residuals to show up for several measurements that are actually correct. And when that happens, it can be hard to know exactly what is wrong. The adjustment may fail the *chi-square* test. That tells you there is a problem, but unfortunately, it cannot tell you where the problem is. The chi–square test is based in probability, and it can fail because there are still unmodeled biases in the measurements. Multipath, ionosphere and troposphere biases, and so forth, may cause it to fail.

Most programs look at the residuals in light of a limit at a specific probability, and when a particular measurement goes over the limit it gets highlighted. Trouble is, you cannot always be sure that the one that got tagged is the one that is the problem. Fortunately, least-squares does offer a high degree of comfort once all the hurdles have been cleared. If the residuals are within reason and the chi-square test is passed, it is very likely that the observations have been adjusted properly.

USING A PROCESSING SERVICE

There are several services available to GPS surveyors. While they differ somewhat in their requirements they are all based on the same idea. Static GPS data collected in the field may be uploaded to a web site on the Internet by the hosting organization, which will then return the final positions, often free of charge. Usually the user is directed to an ftp site where the results can be downloaded.

There can be advantages to the use of such a processing service. Aside from removing the necessity of having postprocessing software in-house, there is the strength of the network solution available from them. In other words, rather than the data sent in by the GPS surveyor being processed against a single CORS in the vicinity of the work, it is processed against a group of the nearest CORS. There are often three in the group. Such a network based solution improves the integrity of the final position markedly and may compensate for long baselines required by the sparseness of the CORS in the area of interest.

Among the online resources available for processing services is the National Geodetic Surveys *Online Positioning User Service* (*OPUS*). This service allows the user to submit RINEX files through NGS Web page. They are processed automatically with NGS computers and software utilizing data from three CORS that may be user selected, see http://www.ngs.noaa.gov/OPUS/.

There are others. Among them are: the Scripps Orbit and Permanent Array Center, SOPAC; *Scripps Coordinate Update Tool, SCOUT,* http://sopac.ucsd.edu/cgi-bin/SCOUT.cgi; the *Australian Online GPS Processing Service, AUSPOS* http://www.ga.gov.au/bin/gps.pl; and the Jet Propulsion Laboratories, *Auto Gipsy, AG,* http://milhouse.jpl.nasa.gov/ag/.

EXERCISES

1. Of the four listed below, what is the most frequent mistake made in GPS control surveying?

   (a) Failure to use a 15° mask angle.
   (b) Failure to have a fully charged battery.
   (c) Failure to properly measure the H.I. of the antenna.
   (d) Failure to properly center the antenna over the point.

2. Which of the following statements about the zero baseline test is not true?

   (a) The zero baseline text eliminates random noise and receiver biases from the results.
   (b) The zero baseline test requires that receivers share the same antenna using an antenna splitter.
   (c) The zero baseline test eliminates satellite clock biases, ephemeris errors, atmospheric biases, and multipath from the results.
   (d) The zero baseline test is not dependent on special software or a test network.

3. The place of a particular station in a static GPS control survey observation schedule may depend on a number of factors. Which of the following would not be one of them?

   (a) The obstructions around the station.
   (b) The difficulty involved in reaching the station in bad weather.
   (c) The previous day's successful and unsuccessful occupations.
   (d) The line-of-sight with other stations in the survey.

4. The *to-reach* description in a static GPS control survey would most likely appear on which of the following forms?

   (a) The Observation Log
   (b) The Station Data Sheet
   (c) The Station Visibility Diagram
   (d) The Observation Schedule

5. The H.I. of the antenna in a static GPS control survey would most likely appear on which of the following forms?

   (a) The Observation Log
   (b) The Station Data Sheet
   (c) The Station Visibility Diagram
   (d) The Observation Schedule

6. Which of the following forms used in a static GPS control survey would not include information on either the start and stop time of the observation or the approximate latitude and longitude of the station?

   (a) The Observation Log
   (b) The Station Data Sheet
   (c) The Station Visibility Diagram
   (d) The Observation Schedule

7. What is the overriding principle that ought to guide the preparation of a station data sheet?

   (a) To forewarn of obstructed satellites at the site.
   (b) To finalize the observation schedule.
   (c) To guide succeeding visitors to the station without ambiguity.
   (d) To explain what actually occurred during the observation itself.

8. Which statement below correctly identifies an advantage of using rubbings in recording the information on an existing monument?

   (a) It ensures that the station was actually visited and that the stamping was faithfully recorded.
   (b) It makes ties to prominent features in the area unnecessary.
   (c) It improves the neatness of the station data sheet.
   (d) It increases the efficiency of the actual occupation of the station.

9. What tools are necessary to prepare a visibility diagram at a station that is somewhat obstructed?

   (a) a theodolite and EDM
   (b) a compass and clinometer
   (c) a GPS receiver
   (d) a level and level rod

10. In planning a kinematic GPS survey, which statement below is the most likely use for control stations on placed on either side of a bridge?

    (a) The control stations may be used to reinitialize a receiver that lost lock while passing under the bridge.
    (b) The control stations might be occupied during the survey as a check of the accuracy of the survey.
    (c) The control stations might be used for an antenna swap.
    (d) The control stations might be set to make the kinematic survey unnecessary in the area.

11. Which of the following statements concerning the correlation of biases in GPS is not correct?

    (a) Even relatively long GPS baselines are short when compared with the more than 20,000-km distances from the receivers to the GPS satellites.
    (b) Every session included in a network is composed of many sets of vectors with correlated errors at each end.
    (c) The simultaneity of observation and the resulting correlation of errors helps make the extraordinary accuracy of GPS possible.
    (d) The effect of correlated errors in GPS baselines cannot be reduced by differencing.

12. Should the memory of GPS receivers be cleared routinely after their observations are downloaded to the memory of the processing computer. Why, or why not?

    (a) Yes, the receiver's memory is limited and should be made available for subsequent work.
    (b) No, it is always wise to have a copy of the receiver's observation in the receiver as long as possible as a backup.
    (c) Yes, once the file has been downloaded there is no reason to keep it in the receiver as well.
    (d) No, it is not possible to download critical information from the receiver's files to the processing computer.

13. Which statement concerning triple differences is most correct?

    (a) Triple differences are the difference between the carrier-phase observations of two receivers of the same satellite at the same epoch.
    (b) Triple differences are the differences of two single differences of the same epoch that refer to two different satellites.
    (c) Triple differences are the difference of two double differences of two different epochs.
    (d) Triple difference does depend on the integer cycle ambiguity, because this unknown is not constant in time.

14. What is the most useful aspect of triple differences in postprocessing GPS observations?

    (a) Triple differences can be an initial step to eliminate cycle slips.
    (b) Triple differences provide the best solution for the receiver positions.
    (c) Triple differences have more information than double differences.
    (d) Triple differences eliminate the need to model ionospheric biases.

15. What is the most useful aspect of a double difference in postprocessing GPS observations?

    (a) Double differences do not need resolution of the integer ambiguity.
    (b) Double differences eliminate clock errors.
    (c) Double differences eliminate multipath.
    (d) Double differences eliminate all atmospheric biases.

16. What would cause the integer cycle ambiguity, N, to change after a receiver has achieved lock-on?

    (a) An incorrect H.I.
    (b) Inclement weather.
    (c) Inaccurate centering over the station.
    (d) A cycle slip.

17. What is the difference between a float and a fixed solution?

    (a) The integer cycle ambiguity is resolved in a fixed solution, but not in a float solution.
    (b) The float solution is processed using a single difference, unlike a fixed solution.
    (c) The float solution is more accurate than a fixed solution.
    (d) The float solution may have cycle slips, but not a fixed solution.

18. What is the idea underlying the use of least-squares adjustment of GPS networks?

    (a) The sum of the squares of all the residuals applied to the GPS vectors in their final adjustment should be held to zero.
    (b) The multiplication of the GPS vectors by all the residuals in their final adjustment should be held to one.
    (c) The sum of the squares of all the residuals applied to the GPS vectors in their final adjustment should be held to the absolute minimum.
    (d) The least of the squares of all the residuals applied to the GPS vectors in their final adjustment should be held to zero.

19. What is usually the purpose of the initial minimally constrained least squares adjustment in GPS work?

    (a) To repair cycle slips.
    (b) To fix the integer cycle ambiguity.
    (c) To establish the correct coordinates of all the project points.
    (d) To find any large mistakes in the work.

20. What may cause an adjustment to fail the chi–square test?

    (a) Unmodeled biases in the measurements.
    (b) Smaller than expected residuals.
    (c) A measurement with too many epochs.
    (d) The lack of cycle slips.

ANSWERS AND EXPLANATIONS

1. Answer is (c)

Explanation: The mismeasured height of the antenna above the mark is probably the most pervasive and frequent blunder in GPS control surveying.

2. Answer is (a)

Explanation: The simplicity of the zero baseline test is an advantage. It is not dependent on special software or a test network and it can be used to separate receiver difficulties from antenna errors. Two or more receivers are connected to one antenna with a *signal* or *antenna splitter*.

An observation is done with the divided signal from the single antenna reaching both receivers simultaneously. Since the receivers are sharing the same antenna, satellite clock biases, ephemeris errors, atmospheric biases, and multipath are all canceled. The only remaining errors are attributable to random noise and receiver biases.

3. Answer is (d)

Explanation: In creating an observation, schedule consideration might be given to observing a particular station during a session when none of the satellites would be blocked by obstructions.

And while GPS is not restricted by inclement weather, particular access routes may not be so immune. Despite best efforts, a planned observation may have been unsuccessful at a required station on a previous day, and it may need to be revisited. However, the line-of-sight between a particular station and another in the survey is not likely to affect the GPS observation schedule, though such a consideration may be critical in a conventional survey.

4. Answer is (b)

Explanation: The *to-reach* description in a static GPS control survey would most likely be prepared during reconnaissance and would appear on the station data sheet.

5. Answer is (a)

Explanation: The H.I. of the antenna in a static GPS control survey would most likely be recorded before and after the actual observation on the station and would appear on the observation log.

6. Answer is (b)

Explanation: The station data sheet would not be likely to include either information. The other forms listed would probably include one or both categories.

7. Answer is (c)

Explanation: The station data sheet is often an important bridge between on-site reconnaissance and the actual occupation of a monument. Neatness and clarity, always paramount virtues of good field notes, are of particular interest when the station data sheet is to be later included in the final report to the client. The overriding principle in drafting a station data sheet is to guide succeeding visitors to the station without ambiguity. A GPS surveyor on the way to observe the position for the first time may be the initial user of a station data sheet. A poorly written document could void an entire session if the observer is unable to locate the monument. A client, later struggling to find a particular monument with an inadequate data sheet, may ultimately question the value of more than the field notes.

8. Answer is (a)

Explanation: Rubbings are performed with paper held on top of the monument's disk; a pencil is run over it in a zigzag pattern producing a positive image of the stamping. This method is a bit more awkward than simply copying the information from the disk onto the data sheet, but it does have the advantage of ensuring the station was actually visited and that the stamping was faithfully recorded. Such rubbings or close-up photographs are required by the provisional FGCC *Geometric Geodetic Accuracy Standards and Specifications for Using GPS Relative Positioning Techniques* for all orders of GPS surveys.

9. Answer is (b)

Explanation: Using a compass and a clinometer, a member of the reconnaissance team can fully describe possible obstructions of the satellite's signals on a visibility diagram. By standing at the station mark and measuring the azimuth and vertical angle of points outlining the obstruction, the observer can plot the object on the visibility diagram.

10. Answer is (a)

Explanation: Unavoidable obstructions like bridges and tunnels can be overcome by the placement of control stations on both sides of the barrier. If these control stations are coordinated by some type of static observation, they can later be used to reinitialize the kinematic receivers should they lose lock passing under the bridge.

11. Answer is (d)

Explanation: When two or more receivers observe the same satellite constellation simultaneously, a set of correlated vectors are created between the coobserving stations. Most GPS practitioners use more than two receivers. Therefore, most GPS networks consist of many sets of correlated vectors for every separate session. The longest baselines between stations on the earth are usually relatively short when compared with the 20,000-km distances from the receivers to the GPS satellites. Therefore, even when several receivers are set up on widely spaced stations, as long as they collect their data simultaneously from the same constellation of satellites, they will record very similar errors. In other words, their vectors will be correlated. It is the simultaneity of observation and the resulting correlation of the carrier phase observables that make the extraordinary GPS accuracies possible. Biases that are correlated linearly can be virtually eliminated by differencing the data sets of a session.

12. Answer is (c)

Explanation: Receiver memory capacity is usually somewhat limited, and older data must be cleared to make room for new sessions. Still, it is a good policy to create the necessary space with the minimum deletion and restrict it to only the oldest files in the receiver's memory. In this way the recent data can be retained as long as possible, the data can provide an auxiliary back-up system. But when a receiver's memory is finally wiped of a particular session, if redundant raw data are not available reobservation may be the only remedy.

13. Answer is (c)

Explanation: A triple difference is created by differencing two double differences at each end of the baseline. Each of the double differences involves two satellites and two receivers. A triple difference considers two double differences over two consecutive epochs. In other words, triple differences are formed by sequentially differencing double differences in time.

Since two receivers are recording the data from the same two satellites during two consecutive epochs across a baseline, a triple difference can temporarily eliminate any concern about the integer cycle ambiguity, because the cycle ambiguity is the same over the two observed epochs.

14. Answer is (a)

Explanation: The triple difference cannot have as much information content as a double difference. Therefore, while receiver coordinates estimated from triple differences are usually more accurate than pseudorange solutions, they are less accurate than those obtained from double differences, especially fixed-ambiguity solutions. Nevertheless, the estimates that come out of triple-difference solutions refine receiver coordinates and provide a starting point for the subsequent double-difference solutions. They are also very useful in spotting and correcting cycle slips. They also provide a first estimate of the receiver's positions.

15. Answer is (b)

Explanation: Double differences have both positive and negative features. On the positive side, they make the highest GPS accuracy possible, and they remove the satellite and receiver clock errors from the observations. On the other hand, the integer cycle ambiguity, sometimes known simply as the *ambiguity*, cannot be ignored in the double difference. In fact, the fixed double-difference solution, usually the most accurate technique of all, requires the resolution of this ambiguity.

16. Answer is (d)

Explanation: The integer cycle ambiguity, usually symbolized by *N*, represents the number of full phase cycles between the receiver and the satellite at the first instant of the receiver's lock-on. N does not change from the moment of the lock is achieved, unless there is a cycle slip.

17. Answer is (a)

Explanation: In other words, the initial estimation of N in a float solution is likely to appear as a real number rather than an integer. However, where the data are sufficient, these floating real-number estimates are very close to integers—so close that they can next be rounded to their true integer values in a second adjustment of the data. Therefore, a second double difference solution follows. The estimation of N that is closest to an integer and has the minimum standard error is usually taken to be the most reliable and is rounded to the nearest integer. Now, with one less unknown, the process is repeated and another ambiguity can be fixed, and so on. This approach leads to the fixed solution in which N can be held to integer values. It is usually quite successful in double differences over short baselines. The resulting fixed solutions most often provide much more accurate results than were available from the initial floating estimates.

18. Answer is (c)

Explanation: The foundation of the idea of least-squares adjustment is the idea that the sum of the squares of all the residuals applied to the GPS vectors in their final adjustment should be held to the absolute minimum.

19. Answer is (d)

Explanation: Most GPS postprocessing adjustment begins with a minimally constrained least-squares adjustment. That means that all the observations in a network are adjusted together with only the constraints necessary to achieve a meaningful solution; for example, the adjustment of a GPS network with the coordinates of only one station fixed. The purpose of the minimally constrained approach is to detect large mistakes, like a misidentifying one of the stations. The residuals from a minimally constrained work should come pretty close to the precision of the observations themselves. If the residuals are particularly large, there are probably mistakes; if they are really small, the network itself may not be a strong as it should be.

20. Answer is (a)

Explanation: The adjustment may fail the *chi-square* test. That tells you there is a problem; unfortunately, it cannot tell you where the problem is. The chi–square test is based in probability, and it can fail because there are still unmodeled biases in the measurements. Multipath, ionosphere and troposphere biases, and so forth, may cause it to fail.

Most programs look at the residuals in light of a limit at a specific probability, and when a particular measurement goes over the limit it gets highlighted. Trouble is you cannot always be sure that the one that got tagged is the one that is the problem. Fortunately, least-squares offer a high degree of comfort once all the hurdles have been cleared. If the residuals are within reason and the chi-square test is passed, it is very likely that the observations have been adjusted properly.

8 GPS Modernization and Global Navigation Satellite System (GNSS)

GPS MODERNIZATION

The configuration of the GPS Space Segment is well-known. A minimum of 24 GPS satellites ensure 24-hour worldwide coverage. But today there are more than that minimum on orbit. There are a few spares on hand in space. The redundancy is prudent. GPS, put in place with amazing speed considering the technological hurdles, is now critical to all sorts of positioning, navigation, and timing around the world. It is that very criticality that requires the GPS modernization. The oldest satellites in the current constellation were launched in 1989. Imagine using a personal computer of that vintage today. It is not surprising that there are plans in place to alter the system substantially. What might be unexpected is many of those plans will be implemented entirely outside of the GPS system itself. This chapter is about some of those changes.

BLOCK I, BLOCK II, AND BLOCK IIR SATELLITES

Currently, the GPS satellites in orbit around the earth include none of the first launched GPS satellites, which were known as Block I (1978–1985). None of these are functional today. The last Block I was retired in late 1995. These satellites needed help from the Control Segment to do the momentum dumping necessary to maintain their attitude control. They carried two cesium and two rubidium frequency standards and had a design life of 5 years, though some operated for double that.

The oldest of the satellites operating on orbit today are those in the original Block II (1989–1997). Remarkably, a majority of them are still healthy. The Block II satellites are radiation hardened against cosmic rays. They were built to provide Selective Availability (SA) and antispoofing (AS) capability and have onboard momentum dumping for the maintenance of velocity and attitude control. There are also upgraded Block II satellites known as Block IIA. The first of these was launched in 1990. Block IIA satellites can function continuously for 6 months without intervention from the Control Segment, but the broadcast ephemeris and clock correction would degrade if that were done. Like the Block I, the Block II satellites are also equipped with rubidium and cesium frequency standards. They are expected to have a Mean Mission Duration (MMD) of 10.6 years. That has obviously been exceeded in most cases. But the Block II satellites do wear out and as they do, the next generation goes up.

These are the Block IIR satellites, R is for replenishment. The first of these was launched in 1997. There are some significant differences between the Block II and the Block IIR satellites. Block IIR satellites can determine their own position using intersatellite crosslink ranging, called AutoNav. They have reprogrammable processors onboard to do their own fixes in flight. In other words, the Control Segment can change their flight software while the satellites orbit. Also, the Block IIR satellites can be moved into a new orbit with a 60-day advanced notice and they are more radiation hardened than their predecessors. In short they have more autonomy. On some of the Block IIR satellites there will be an improved antenna panel that will provide more signal power to GPS users on both L1 and L2. Eventually, there will be approximately 21 Block IIR satellites launched.

However, there is one significant way in which the Block II and the Block IIR satellites are very much the same. They both broadcast the same fundamental GPS signals that have been in place for a long time. Their frequencies are centered on L1 and L2. As mentioned before, the Coarse/Acquisition code or C/A-code is carried on L1 and has a chipping rate of 1.023 million chips per second. It has a code length of 1023 chips over the course of a millisecond before it repeats itself. There are actually 32 different code sequences that can be used in the C/A code, more than enough for each satellite. The Precise code or P-code on L1 and L2 has a chipping rate that is ten times faster at 10.23 million chips per second. It has a code length of about a week, approximately 6 trillion chips, before it repeats. If this code is encrypted it is known as the P(Y) code, or simply the Y-code.

POWER SPECTRAL DENSITY DIAGRAMS

A convenient way to visualize these signals is a diagram of the power spectral density function (PSD). In fact, a good deal of signal theory is expressed in PSDs. They are basically an expression of Watts per Hertz as a function of frequency. Another way of saying it is power per bandwidth. The actual definition of PSD is the Fourier transform of the autocorrelation function, but the idea behind them is to give you an idea of the power within a signal with regard to frequency. In GPS the diagram is usually represented with the frequency in MHz on the horizontal axis and the density, the power, represented on the perpendicular axes in decibels relative to one Hertz per Watt or dBW.

Perhaps a bit of explanation is in order for dBW. As mentioned in Chapter 4, the decibel unit was really begun in Bell Labs as a way to express power loss on telephone lines. It is logarithmic function, but can be thought of as a percentage or a ratio. In other words, a change of 1 decibel would be an increase or a decrease of 27%. A change of 3 decibels would be an increase or a decrease of 100%. Now a decibel per Watt (dBW) indicates actual power of a signal compared to a reference of 1 Watt, *W*. For example, the signal from a GPS satellite is about +13.4 dBW and the same signal as received is about –163 dBW. Actually the minimum level for power received from the C/A code is –158.5 dBW, for the P code it is –161.5 dBW, and for L2 it is –164.5 dBW.

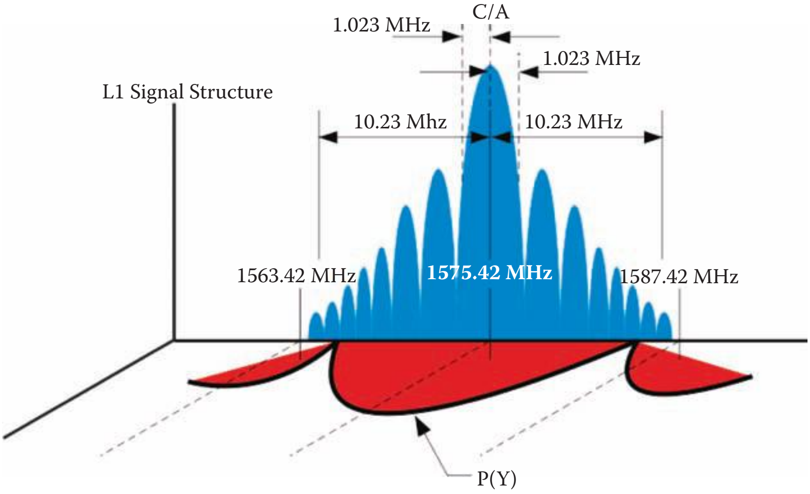


FIGURE 8.1 L1 Signal

Therefore, these graphics show the increase or decrease, in decibels, of power, in Watts with respect to frequency in Hertz. Let's start with a couple of PSD diagrams of the well-known codes on L1 and L2.

L1 SIGNAL

For example, in Figure 8.1 one can see the C/A code on the L1 signal, centered on the frequency 1575.42 MHz and spread over a bandwidth that is approximately 20.46 MHz, 10.23 MHz on each side of the center frequency. The P(Y) code is in quadrature that is 90 degrees from the C/A code. In both cases the majority of the power is close to the center frequency. The C/A code has many lobes but the P code with the same bandwidth but 10 times the clock rate has just the one main lobe.

L2 SIGNAL

In Figure 8.2 the L2 signal diagram is centered on 1227.60 MHz. As you can see it is similar to the L1 diagram except for the absence of the C/A code, which is, of course, not carried on the L2 frequency. As well-known as these are this state of affairs is changing.

NEW SIGNALS

An important aspect of GPS modernization is the advent of some new and different signals that are augmenting the old reliable codes. In GPS a dramatic step was taken in this direction on September 21, 2005 when the first Block IIR-M satellite was launched. One of the significant improvements coming with the Block IIR-M satellites is increased L-band power on both L1 and L2 by virtue of the new antenna panel. The Block IIR-M satellites will also broadcast new signals, such as the M-code.

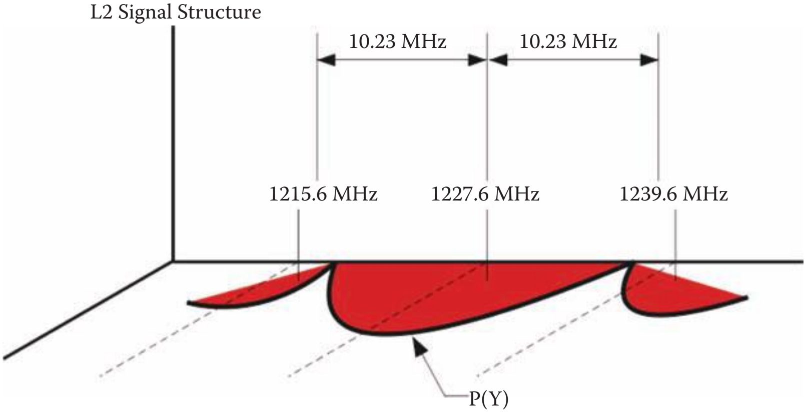


FIGURE 8.2 L2 Signal

The M-Code

Actually the M stands for modernized, but it is interesting to note that part of that modernization also includes a new M-code. Eight to twelve of these replenishment satellites are going to be modified to broadcast a new military code, the M-code. This code will be carried on both L1 and L2 and will probably replace the P(Y) code eventually and has the advantage of allowing the Department of Defense (DoD) to increase the power of the code to prevent jamming. There was consideration given to raising the power of the P(Y) code to accomplish the same end, but that strategy was discarded when it was shown to interfere with the C/A code.

The M-code was designed to share the same bands with existing signals, on both L1 and L2, and still be separate from them. See those two peaks in the M-code in Figure 8.3. They represent a split-spectrum signal about the carrier. Among other things this allows minimum overlap with the maximum power densities of the P(Y) code and the C/A code, which occur near the center frequency. That is because the actual modulation of the M-code is done differently. It is accomplished with

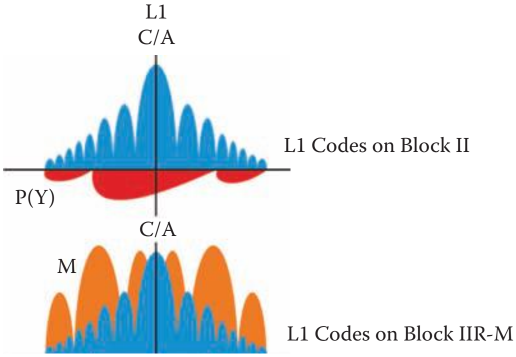


FIGURE 8.3 L1 with M-Code

binary offset carrier (BOC) modulation, which differs from the binary phase shift key (BPSK) used with the legacy C/A and P(Y) signals. An important characteristic of BOC modulation is the M-code has its greatest power density at the edges, that is at the nulls, of the L1. This architecture both simplifies implementation at the satellites and receivers and also mitigates interference with the existing codes. Suffice it to say that this aspect and others of the BOC modulation strategy offer even better spectral separation between the M-code and the older legacy signals.

Perhaps it would also be useful here to mention the notation used to describe the particular implementations of the Binary Offset Carrier. It is characteristic for it to be written BOC(α,β). Here the α indicates the frequency of the square wave modulation of the carrier, also know as the subcarrier frequency factor. The β describes the frequency of the pseudorandom noise modulation, also known as the spreading code factor. In the case of the M-code the notation BOC (10,5) describes the modulation of the signal. Both here are multiples of 1.023 MHz. In other words their actual values are

$$\alpha = 10 \times 1.023\ \text{MHz} = 10.23\ \text{MHz and } \beta = 5 \times 1.023\ \text{MHz} = 5.115\ \text{MHz (Betz).}$$

The M-code is tracked by direct acquisition. This means that as mentioned in Chapter 1, the receiver correlates the signal coming in from the satellite with a replica of the code that it has generated itself.

L2C

A new military code on L1 and L2 may not be terribly exciting to civilian users, but these IIR-M satellites have something else going for them. They are outfitted with new hardware that will allow them to broadcast a new civilian code. This is a code that was first announced back in March of 1998. It will be on L2 and will be known as L2C. The "C" is for civil.

We have been using the L2 carrier since the beginning of GPS of course, but now there will be two new codes broadcast on the carrier, L2, that previously only carried one military signal exclusively, the P(Y) code. Now L2 will carry a new military signal, the M-code, and a new civil signal as well, L2C as shown in Figure 8.4.

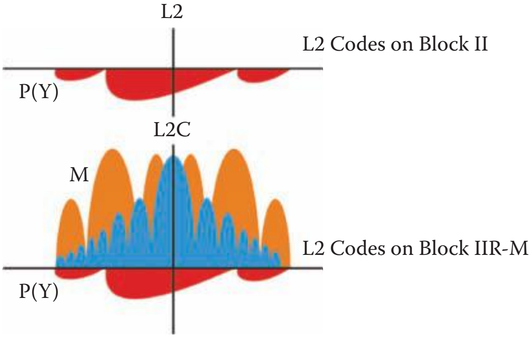


FIGURE 8.4 L2 with L2C and M-Code

Even though its 2.046 MHz from null-to-null gives it a very similar power spectrum to the C/A code, it is important to note that L2C is not merely a copy of the C/A-code. Still that was in fact the initial idea. The original plan was that it would be a replication of the venerable C/A code, but carried on L2 instead of L1. This concept changed when Colonel Douglas L. Loverro, Program Director for the GPS Joint Program Office (JPO), was asked if perhaps it was time for some improvement of C/A. The answer was yes. The C/A code is somewhat susceptible to both waveform distortion and narrow-band interference and its cross-correlation properties are marginal at best. So the new code on L2, known as L2 civil, or L2C was announced.

Civil-Moderate (CM) and Civil-Long (CL)

L2C will be a bit more sophisticated that C/A. L2C is actually composed of two codes, L2CM and L2CL. The CM designation stands for the civil-moderate length code. This signal carries data. The message that it carries is an improved Navigation code. You may recall that the legacy Navigation message was mentioned in Chapter 1. It is also known by the acronym NAV, and is broadcast at 50 bits per second (bps). The new Navigation message is known as CNAV. It is broadcast at 25 bps. Like the original NAV message, it has 300 bit subframes, but given the lower broadcast rate CNAV takes 12 seconds to transmit each of its subframes, whereas NAV requires only 6 seconds. Nevertheless because CNAV is a bit more flexible and compact than the original NAV it has the very desirable effect of allowing a receiver to get to its first fix on a satellite much faster than before.

CNAV

While the information in CNAV is fundamentally the same as that in the original Navigation message, including almanac, ephermides, time, and satellite health, the data is more accurate and provides higher precision. Also, instead of using the same frame-subframe format of the Navigation message illustrated in Chapter 1, CNAV uses 12-second 300-bit message packets. One of every four of these packets includes clock data, two of every four contains ephemeris data, and so on. CNAV can accommodate the transmission of the information in support of 32 satellites using 75% or less of its bandwidth and a fraction of the available packet types.

There could be a packet that would contain differential correction like that available from satellite based augmentation systems (SBAS). This could be used to improve the L1 NAV clock data. As it stands a packet is assigned to the time offset between GPS and GNSS, which is a boon for interoperability between GPS, GALILEO, and GLONASS. Also each packet contains a flag that can be toggled on within a few seconds of when a satellite is known to be unhealthy and should not be used. This is exactly the sort of quick access to information necessary to support safety-of-life applications. In other words, CNAV is designed to grow up to accommodate 63 satellites and change as the system requires.

There is also a very interesting aspect to the data broadcast on CM known as *Forward Error Correction* (*FEC*). An illustration of this technique is to imagine that every individual piece of data is sent to the receiver twice. If the receiver knows the details of the protocol to which the data ought to conform it can compare each of the two instances it has received to that protocol. If they both conform, there is no problem. If one does and one does not, the piece of data that conforms to the protocol is accepted and the other is rejected. If neither conforms then both are rejected. Using FEC allows the receiver to correct transmission errors itself, on the fly. The CL for civil-long, on the other hand, is a pilot signal and it carries no message. They utilize the same modulation scheme, binary phase shift key (BPSK), as the legacy signals.

Multiplexing

But wait a minute, how can you do that? How can you have two codes in one? L2C achieves this by multiplexing. Since the two codes have different lengths L2C alternates between chips of the CM code and chips of the CL code. It is called chip-by-chip time multiplexing. So even though the actual chipping rate is 511.5 KHz, half the chipping rate of the C/A code, with the time multiplexing it still works out that taken together L2C ends up having the same overall chip rate as L1 C/A code, 1.023 MHz.

The CM code, the moderate length code goes through 10,230 chips before it repeats. It does that every 20 milliseconds. But the CL code, the long code repeats after 767,250 chips every 1.5 seconds, and that length gives you very good cross-correlation protection. In fact, both are longer than the C/A code and present a subsequent improvement in autocorrelation, as well as cross-correlation. This is because the longer the code the easier it is to keep the desired signals separate from the background. In practice this means these signals can be acquired with more certainty by a receiver that can maintain lock on them more surely in marginal situations where the sky is obstructed.

Phase-Locked Loop

It is also important to note that L2C overall is approximately 2.3 dB weaker than is C/A on L1. Surprisingly, that is not a disadvantage due to the structure of the L2C signal. The receiver can track the long data-less CL with a phase-locked loop instead of a squaring Costas loop that is necessary to maintain lock on CM, C/A and P(Y). This allows for improved tracking from what is, in fact, a weaker signal and a subsequent improvement in protection against continuous wave interference. As a way to illustrate how this would work in practice, here is one normal sequence by which a receiver would lock onto L2C. First there would be acquisition of the CM code with a frequency locked or Costas loop, next there would be testing of the 75 possible phases of CL, and finally acquisition of CL. The CL as mentioned can be then tracked with a basic phase-locked loop. Using this strategy, even though L2C is weaker than C/A there is actually an improvement in the threshold of nearly 6 dB by tracking the CL with the phase-locked loop.

In other words the long data-less sequence of the CL provides for a correlation about 250 times stronger than the C/A code. So despite the fact that its transmission power is 2.3 dB weaker, compared to the C/A code L2C has 2.7 dB greater data recovery and 0.7 dB greater carrier-tracking.

PRACTICAL ADVANTAGES

Great, so what does all that mean in English? It means that L2C has better tolerance to interference. It also means increased stability. It means improved tracking in obstructed areas like woods, near buildings, and urban canyons. It means fewer cycle slips.

There are more solid practical advantages to the introduction of L2C. Before May 2, 2000 with Selective Availability on, a little handheld code-based receiver could get you within 30 to 100 meters of your true position. When SA was turned off that was whittled down to 15 to 20 meters or so under very good conditions. But with just one civilian code C/A on L1 there was no way to remove the second largest source of error in that position, the ionospheric delay. But now with two civilian signals, one on L1 (C/A) and one on L2 (L2C), it becomes possible to effectively model the ionosphere using code phase. In other words, it may become possible for an autonomous code-phase receiver to achieve positions with a 5-10m positional accuracy with some consistency. And there are more developments coming, developments that just might increase that accuracy to a submeter range.

So, even if it is the carrier-phase that ultimately delivers the wonderful positional accuracy we all depend on, the codes get us in the game and keep us out of trouble every time we turn on the receiver. The codes have helped us to lock on to the first satellite in a session and allowed us to get the advantage of cross-correlation techniques almost since the beginning of GPS. In other words, our receivers have been combining pseudorange and carrier-phase observables in innovative ways for some time now to measure the ionospheric delay, detect multipath, do wide laning, and so forth. But those techniques can be improved, because while the current methods work, the results can be noisy and not quite as stable as they might be, especially over long baselines. It will be cleaner to get the signal directly once there are two clear civilian codes, one on each carrier. It may also help reduce the complexity of the chipsets inside our receivers, and might just reduce their cost as well.

Along that line, it is worthwhile to recall that the L2C has an overall chip rate of 1.023 MHz, just like L1 C/A. Such a slow chip rate can seem to be a drawback until you consider that that rate affects the GPS chipset power consumption. In general, the slower the rate the longer the battery life and the improvement in receiver battery life could be very helpful. And not only that, the slower the chip rate, the smaller the chipset. That could mean more miniaturization of receiver components.

L2C is clearly going to be good for the GPS consumer market, but it also holds promise for surveyors. Nevertheless, there are a few obstacles to full utilization of the L2C signal. As mentioned earlier the first Block IIR-M, namely SVN53/PRN17, was launched on September 21, 2005. It will be some time before the constellation of Block IIR-M necessary to provide L2C at an operational level is up and functioning. Additionally aviation authorities do not support L2C. It is not in an *Aeronautical Radionavigation Service* (*ARNS*) protected band. It happens that L2 itself occupies a radiolocation band that includes ground-based radars.

L5

THE L5 CARRIER

Alright, L2C is fine, but what about the new carrier everybody has been talking about, L5? It will be centered on 1176.45 MHz, 115 times the fundamental clock

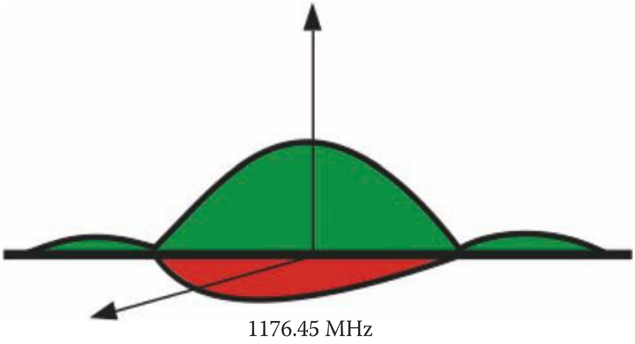


FIGURE 8.5 The L5 Carrier

rate. As you see from Figure 8.5 the basic structure of L5 looks similar to that of L1. There are two codes on this carrier in quadrature to each other. They have separate pseudorandom noise (PRN) codes, two codes per satellite, which are modulated using Quad Phase Shift Key (QPSK). However, borrowing a page from the newer developments the in-phase (I) signal carries a data message and the other, the quadraphase signal (Q), is data-less. Both have equal power. The data that is carried on L5 is a compact flexible message similar to that carried by L2C CM. L5 will also utilize time multiplexing in broadcasting its two codes as does L2C in broadcasting CM and CL.

Unlike L2C, L5 will have the benefits of its place in a band designated by the International Telecommunication Union for the Aeronautical Radionavigation Navigation Services (ARNS). Therefore, it will not be prone to interference with ground based navigation aids and will be available for aviation applications. It is also quite helpful that no other GPS signal occupies this band. However, L5 will share space with one of the signals, E5, from an entirely separate satellite system, GALILEO, a very good idea, more about that later in this chapter.

As mentioned, L5 will have one signal modulated with data and one without. And since L5 does not carry military signals, it achieves the power split by using two long equal-length codes in phase quadrature on each satellite. Both have a 10.23 MHz chipping rate, the same as the fundamental clock rate. It is worth noting that this is the same rate that has been available on the P(Y) code from the beginning of the system. However, as you know the P(Y) code is unavailable for civilian use. So this will be the fastest chipping rate available in a civilian code; it will have the same overall length as CM on L2C and L1C, 10,230 chips. But L5 will have the only civilian codes that are both ten times longer and ten times faster than the C/A code so the risk of interference is very low and, good news, the data-less signal will be much easier to acquire in unfavorable signal-to-noise ratio (SNR) conditions.

The fast chipping rate is also good news. As mentioned in Chapter 1 a rule of thumb is the maximum resolution available in a pseudorange is about 1% of the chipping rate of the code used. The faster the chipping rate the better the resolution, and it also improves multipath protection. L5 will also have higher power than L1, about four times more power. The increased power is also good news because the legacy signals from GPS satellites are weak. L5 will also have a wide bandwidth, about 20 MHz. L5 will also incorporate Forward Error Correction (FEC).

THE BLOCK IIF SATELLITES

L5 will be first introduced on the fourth generation GPS satellites, Block IIF; the F is for follow-on. These satellites have a somewhat longer expected lifetime of 12 to 15 years with faster processors and more memory onboard. They carry two rubidium atomic frequency standards and one cesium. Their onboard *navigation data units* (*NDU*) support the creation of new navigation messages with improved broadcast ephemeris and clock corrections. Like the Block IIR satellites, the Block IIF can be reprogrammed on orbit. But we are going to have to be a little patient. The IIR and IIR-M satellites are in line in front of the IIF satellites. The dozen or so IIR's will be launched, orbit, and reach the end of their functional lives before L5 will be on all the satellites. And the satellites already in orbit have performed well beyond their design life.

As a matter of fact, that development is somewhat responsible for the delay in implementing GPS III. GPS III is the effort to take a new look at the entire GPS architecture to ensure the best service for the next 30 years. We may have to wait a while before L5 and the L2C on L2 are fully operational. It is unfortunate that L5 cannot be built into the IIR's, the satellites that are going up now.

The first six Block IIF satellites are expected to have a Mean Mission Duration (MMD) of 9.9 years. Block IIF will have all the codes discussed earlier, including L2C on L2, and it will broadcast the new L5 carrier. L5 will not cause any interference to existing systems nor require their modification.

SUMMARY OF COARSE ACQUISITION (C/A), L2C, AND L5

GPS modernization is no longer a future development, it is underway. New spacecraft with better electronics, better navigation messages, newer and better clocks are just part of the story. Beginning with the launch of the first IIR-M satellite new civil signals began to appear, starting with L2C. It will be followed by others, including L5.

These signals tend to have longer codes, faster chipping rates, and more power than the C/A and P(Y) codes have. In practical terms these developments lead to faster first acquisition, better separation between codes, reduced multipath and better cross- correlation properties.

NEW SIGNAL AVAILABILITY

Figure 8.6 provides a general outline of all of the signals that are or will be available on the upcoming satellite blocks.

Availability

Given the current projected launch schedules it looks like the three new civil signals L1-C/A, L2-L2C, and the in-phase and quadraphase signals on L5 will attain initial operational capability (IOC) between 2010 and 2020 or so. Full operational

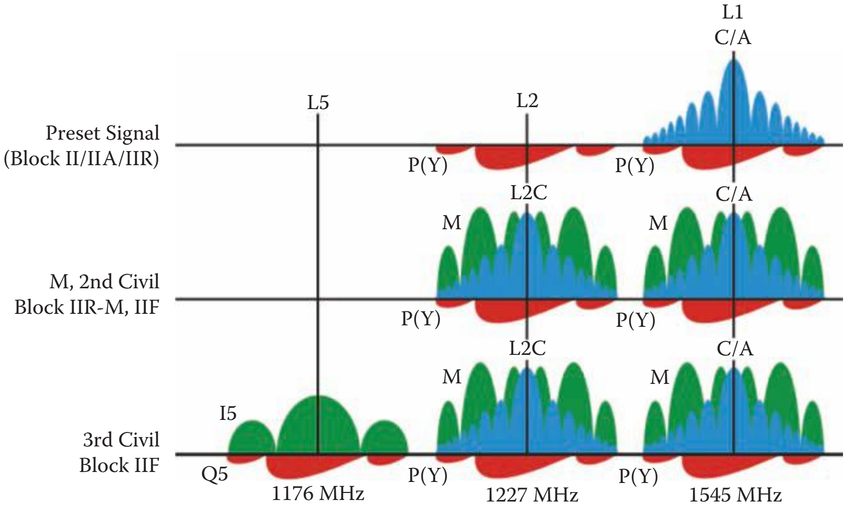


FIGURE 8.6 Modernized Signal Evolution (Adapted from Lazar, Steve, *Crosslink* [Summer 2002], Aerospace Corporation, http://www.aero.org/publications/crosslink/summer 2002/07.html.)

Carrier	Signal	Block IIR	Block IIR-M	Block IIF
L1	P/Y	•	•	•
L2	P/Y	•	•	•
L1	C/A	•	•	•
L2	L2C		•	•
L1	M		•	•
L2	M		•	•
L5	Civil			•

FIGURE 8.7 New Signal Availability

capability, FOC will follow IOC within 5 to 10 years. These years are, of course, impossible to predict with accuracy. Aviation, in particular, is looking forward to L5 for its promise of precision approach capability.

Ionospheric Bias

Concerning the effect of the ionosphere—as you know ionospheric delay is inversely proportional to frequency of the signal squared. So it is that L2's atmospheric bias is about 65% larger than L1, and it follows that the bias for L5 is the worst of the three at 79% larger than L1. L1 exhibits the least delay as it has the highest frequency of the three.

Correlation Protection

Where a receiver is in an environment where it collects some satellite signals that are quite strong and others that are weak, such as inside buildings or places where the sky is obstructed, correlation protection is vital. The slow chipping rate, short code length, and low power of L1 C/A means it has the lowest correlation protection of the three frequencies L1, L2, and L5. That means that a strong signal from one satellite can cross correlate with the codes a receiver uses to track other satellites. In other words the strong signal will actually block collection of the weak signals. To avoid this the receiver is forced to test every single signal so to avoid incorrectly tracking the strong signal it does not want instead of a weak signal that it does. This problem is much reduced with L2. It has a longer code length and higher power than L1. It is also reduced with L5 as compared to L1. L5 has a longer code length, much higher power, and a much faster chipping rate than L1. In short, both of the civilian codes on L2 and L5 have much better cross correlation protection and better narrowband interference protection than L1, but L5 is best of them all.

ANOTHER CIVIL SIGNAL—L1C

As the result of an agreement reached between the United States and the European Union (EU) in June of 2004, yet another civil signal is in the beginning stages of development. Part of that deal involved the creation of an interoperable GPS/GALILEO signal on L1. This signal, known as L1C will be only one of two common interoperable signals shared by both GPS and GALILEO's L1 Open Service signal. Work is also underway to allow L1C to be interoperable with QZSS, the Japanese Quasi-Zenith Satellite System, as well. These developments open extraordinary possibilities of improved accuracy and efficiency when one considers there may eventually be a combined constellation of 50 or more satellites all broadcasting this same civilian signal. All this is made possible by the fact that each of these different satellite systems utilizes carrier frequencies centered on the L1, 1575.42 MHz frequency.

While the details of L1C's design are somewhat provisional it is intended to provide a performance improvement over the C/A code and will have some similarities with L2C. L1C will have double the power of the C/A signal and a code of the same length as CM on L2C 10,230 chips. Like L2C, it will have a pilot signal that does not

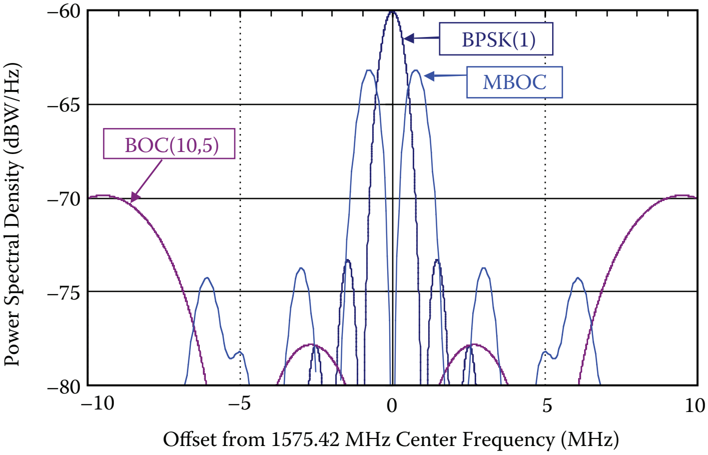


FIGURE 8.8 L1 with M-Code and L1C (Courtesy of Tom Stansell of Stansell Consulting.)

carry a message, and will also have one signal with a data message with exactly the same code length on both components. This approach means that all of the signal power can be used for acquisition of the signal. In other words, this strategy offers equal power splitting between the data and data-less portions.

The data portion of L1C will carry a Navigation message known as CNAV-2, just like L2C. Among other things this feature will allow the receiver to reach its first fix to the satellite faster. The details of this Navigation message are not yet as complete as other aspects of this signal. Also, as in L2C, L1C will likely incorporate FEC.

The L1C provisional design also shares some design similarities with the M-code, Binary Offset Carrier, BOC modulation. L1C will use MBOC, which is 90.9% BOC (1, 1) and 9.1% BOC (6, 1), and as with the M-code it will have good separation from the other signals on L1 and a good tracking threshold as well. Please see L1C in the blue in Figure 8.8. Please recall that this also allows tracking with the superior phase-locked loop as opposed to the Costas loop.

There are many signals on L1. Perhaps that has something to do with the fact that L1, having the highest frequency, experiences the least ionospheric delay of the carrier frequencies. There is the C/A code, the P(Y) code, the M-code and now perhaps the L1C code. It is a challenge to introduce yet another code on the crowded L1 frequency and still maintain separability.

GLOBAL NAVIGATION SATELLITE SYSTEM (GNSS)

As mentioned earlier in this chapter, it may be a bit unexpected but many of the plans that will change GPS as a practical utility will be implemented entirely outside of the GPS system itself.

The GPS system is one component of the worldwide effort now known as the *Global Navigation Satellite System* (*GNSS*). Another component of GNSS is the GLONASS system of the Russian Federation and a third is the GALILEO system administered by the EU. And it is likely that more constellations will eventually be included in GNSS.

The concept is nothing less than this: these networks of satellites and others will begin to work together. Further they will be augmented by both *ground-based augmentation systems* (*GBAS*) and *space-based augmentation systems* (*SBAS*) to provide positioning, navigation, and timing solutions to users around the world.

One goal of this cooperation is interoperability. Interoperability is the idea that properly equipped receivers will be able to obtain useful signals from all available satellites in all the constellations and have their solutions improved rather than impeded by the various configurations of the different satellite broadcasts.

One example of the scope of this increased horizon in global positioning is illustrated by the name change of the *International GPS Service* to the *International GNSS Service* (*IGS*). It is a federation of 200 worldwide agencies that generate information on the GPS & GLONASS systems, http://igscb.jpl.nasa.gov.

GNSS includes GPS, GLONASS, and GALILEO; it will also incorporate the Japanese *Quasi-Zenith Satellite System* (*QZSS*) and the *Chinese Beidou/Compass Satellite Navigation and Positioning System* as well as augmentation systems deployed by the United States, Europe, Japan, China, and Australia.

One immediate effect of GNSS is the substantial growth of the available constellation of satellites. The more signals that are available for positioning and navigation, the better. The two systems that are currently online and available are GPS and GLONASS.

GLONASS

Russia's *Globalnaya Navigationnaya Sputnikovaya Sistema* (*Global Orbiting Navigation Satellite System*), known as *GLONASS,* did not reach full operational status before the collapse of the Soviet Union. Its first satellites reached orbit in October 1982, a bit more than 4 years after the GPS constellation was begun. A nearly full constellation of 24 or so GLONASS satellites was achieved in 1996 but by 2001 only about 7 healthy satellites remained on orbit about 1000 km lower than the orbit of GPS satellites. And the remaining 7 were only expected to have a design life of 3 years. The situation was not helped by the independence of Kazakhstan, subsequent difficulties over the Baikonur Cosmodrome launch facility, and lack of funds. The system was in poor health.

Today there are signs of renewal. Since a decision in August 2001, which outlined a program to rebuild and modernize GLONASS, the constellation has increased dramatically. Full worldwide 24-hour coverage is expected between 2010 and 2020.

For example, there are improvements in the satellites themselves. The original GLONASS satellite was the Uragan. It was first launched in 1982 and had an intended life span of 4 years.

GLONASS URAGAN M AND K

As shown in Figure 8.9, the M version of the GLONASS Uragan has improved antennas over the earlier spacecraft. They are also expected to have extended lifetimes of 7 years and carry separate transmission frequencies that are dedicated to civilian users. The first of these were launched in 2003.

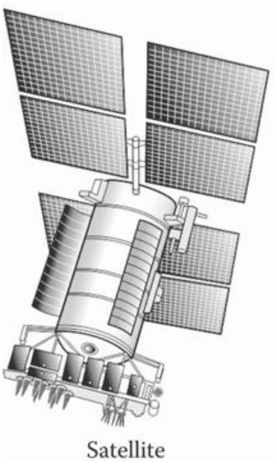


FIGURE 8.9 GLONASS Uragan M

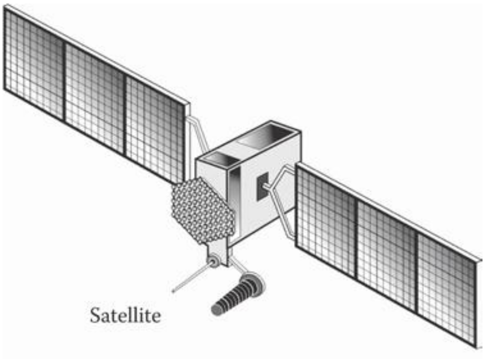


FIGURE 8.10 GLONASS Uragan K

As shown in Figure 8.10, the K version of the GLONASS Uragan has a third L-band transmitter for civilian users, an extended service life of 12 to 15 years, and costs less to produce. The first launch of these satellites is expected in 2009.

GLONASS Constellation

In other words, the GLONASS constellation has increased. A complete GLONASS system would contain 21 active and 3 spare satellites spread over 3 orbital planes at the altitude of 19,100 km inclined 64.8 degrees toward the Equator.

The day after Christmas in 2004, 3 GLONASS satellites were launched. The mission included 2 Uragan, *796 and 797* and a follow-on Uragan-M satellite, *712*. *Please note that the addition of a 7 to the satellite numbers converts them to their GLONASS numbers, that is, No. 98—GLONASS Number 798*. The next year, on Christmas Day, 3 more GLONASS satellites a regular Uragan spacecraft *798* and a pair of upgraded Uragan-M satellites, *713* and *714*, achieved orbit. The launch pattern was repeated with 3 GLONASS satellites sent up on Christmas Day 2006 and 2007 each. There was also a launch of 3 GLONASS satellites in October 2007.

While the GPS system has six orbital planes, as mentioned GLONASS has three. In September of 2006, 3 out of the 8 satellites in the third plane were deactivated, apparently in anticipation of the reconfiguration of the constellation for the launch on Christmas Day in 2006. On that day again 3 more GLONASS satellites, *715*, *716*, *and 717,* all Uragan M satellites, were launched. The goal is to achieve full worldwide 24-hour coverage by the first part of the second decade of the 21st century with 21 Uragan-M satellites in 3 orbital planes, with 3 on-orbit spares in place.

There has also been a recent approval of plans to substantially increase the funding and accelerate the restoration of a complete constellation including upgrading its ground sector. India is assisting Russia with GLONASS, it appears it will be possible to remove the legal barriers to civil use of the GNSS receivers and develop a civilian mass market in GNSS in Russia including digital mapping and equipment.

GLONASS Signals

Regarding the signals broadcast by these satellites, the original objective was similar to the plan embraced by GPS a system that would provide 100 meters accuracy with a deliberately degraded standard C/A signal and 10- to 20-meter accuracy with its P signals available exclusively to the military. However, that changed at the end of 2004, when the *Federal Space Agency* (*FKA*) announced a plan to provide access to the high-precision navigation data to all users. This is, of course, based on the code solution.

Code Division Multiple Access (CDMA) and Frequency Division Multiple Access (FDMA)

Since the revitalization of GLONASS is underway, interoperability between it and other systems is enticing. However, as you know a receiver collecting signals from GPS, or GALILEO for that matter, gets a different segment of the P code and the C/A code from each satellite. In other words, as mentioned in Chapter 1, a particular segment of the 37-week long P code is assigned to each satellite. For example, SV14 is so named because it broadcasts the fourteenth week of the P code. Also, each GPS satellite broadcasts its own completely unique segment of the C/A code that it repeats. However, even though the segments of the P code and the C/A code coming into a receiver on L1 are unique to their satellite or origin, they all arrive at the same frequency, 1575.42 MHz. The same is true of the P code arriving from satellites on L2; even though the segments of the P code coming into a receiver on L2 are unique to their satellite or origin, they all arrive at the same frequency, 1227.60 MHz.

This approach is known as *CDMA* (*Code Division Multiple Access*). CDMA technology was originally developed by the military during World War II. Researchers were looking for ways of communicating that would be secure in the presence of jamming. CDMA does not use frequency channels or time slots. As in GPS, CDMA usually involves a narrow band message multiplied by a large

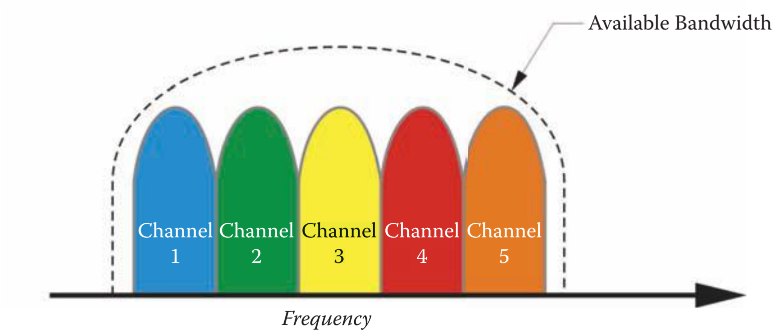


FIGURE 8.11 Frequency Division Multiple Access (FDMA)

bandwidth PRN (pseudorandom noise) signal. As you have read these PRN codes are attached to the GPS carrier by changes in phase. Then all the users can receive the same frequency bands. And again, just as in GPS with CDMA, the transmitted message is recovered by correlating the received signal with the PRN code available at the receiver.

GLONASS uses a different strategy. As shown in Figure 8.11, the satellites transmit L-band signals, and unlike GPS each code a GLONASS receiver collects from any one of the GLONASS satellites is exactly the same. Also unlike GPS each GLONASS satellite broadcasts its codes at its own unique assigned frequency. This is known as *FDMA* (*Frequency Division Multiple Access*).

The two GLONASS L bands have a range of frequencies to assign to satellites. In the future there may be a GLONASS L3. L1 is centered on 1602 MHz with a range between ~1598.0625 to ~1607.0625 MHz, and L2 is centered on 1246 MHz with a range between ~1242.9375 to ~1249.9375 MHz. However, within those ranges there can be up to 25 channels of L-band signals; currently there are 16 channels on each to accommodate the available satellites. Each channel is separated from the others by a ΔF which is 0.5625 MHz on L1 and 0.4375 MHz on L2. In other words each GLONASS satellite broadcasts the same code, but each satellite gets its own frequencies. The standard code chip lengths on the GLONASS L1 are 0.511MHz, which is 3135.03 L1 cycles/chip standard and 5.11 MHz precise, which is 313.503 L1 cycles/chip. On L2 they are 0.511MHz, which is 2438.36 L2 cycles/chip standard and 5.11 MHz precise, which is 243.836 L2 cycles/chip.

A civil reference signal on the L2 frequency satellites will substantially increase the accuracy of navigation relying on civil signals just as L2C will add capability on the GPS L2 signal. GLONASS signals do not overlap GPS frequencies, but the third civil reference signal on L3 that will be available on the K satellites will be within a new frequency band that includes 1201.743 to 1208.511 MHz and will overlap GALILEO's E5B signal. This could be good news.

While there are some differences in the signals available from GPS, the EU's GALILEO system, and Russia's GLONASS, they are surmountable. And Russia has discussed development and use of GLONASS in parallel with the American GPS and European GALILEO systems.

Changes to FDMA

There may be some changes to the FDMA approach in the future. Recently Russia agreed to alter the architecture a bit. In order to use only half as many bands GLONASS will now assign the same frequency to satellites that are in the same orbital plane but are always on opposite sides of the earth.

This will not only reduce the amount of the radio spectrum used by GLONASS; it may actually improve its broadcast ephemeris information. Utilizing so many frequencies makes it difficult to accommodate the wide variety of propagation rates and keep the ephemeris information sent to the receivers within good limits. There are a number of receiver manufacturers that have GPS/GLONASS receivers available but the differences between FDMA and CDMA signals increases the technical difficulty and costs of such equipment. In the last few months of 2006, it was mentioned that GLONASS probably will be able to implement CDMA signals on the third frequency and at L1. This could make it easier for GPS and GALILEO to be interoperable with GLONASS.

In fact, there are many efforts underway to improve the GLONASS accuracy. The stability of the satellites' onboard clocks has improved from 5 x 10^{-13} to 1 x 10^{-13} over 24 hours with precision thermal stabilization. The GLONASS Navigation Message will include the difference between GPS time and GLONASS time, which is significant.

GLONASS TIME

As you know there are no leap seconds introduced to GPS Time in synch with UTC. However things are different in GLONASS. Leap seconds are incorporated into the time standard of the system. Therefore, there is no integer-second difference between GLONASS Time and UTC as there is with GPS. Still that is not the whole story. The epoch and rate of Russian time, relative to UTC (BIH), is monitored and corrected periodically by the Main Metrological Center of Russian Time and Frequency Service (VNIIFTRI) at Mendeleevo, Moscow. They establish the regional version of UTC, which is known as UTC (CIS). There is a constant offset of 3 hours between GLONASS Time and UTC (CIS). However, with these differences available in the Navigation Message from GLONASS they can be accommodated. There are also efforts to increase the number of available tracking facilities in the GLONASS Ground Segment from 9 to 12, tie the GLONASS coordinate system to the International Terrestrial Reference System (ITRS), and launch the improved Uragan K spacecraft, which may add a third L3 frequency band including differential ephemeris and time corrections that would allow submeter real-time positioning accuracy.

Considering interoperability, given the fact that the first GALILEO satellite reached orbit atop a Soyuz-Fregat rocket, there is every reason to believe that the GALILEO/GLONASS agreement concerning the signal compatibility and interoperability at the GLONASS L3 and Galileo E5 or E6 bands will be successful.

GALILEO

Just over a dozen years after the idea was first proposed, the work on GALILEO culminated in the launch of GIOVE-A (Galileo In Orbit Validation Experiment-A) on

December 28, 2005. The name GIOVE, Italian for Jupiter, is also a tribute to Galileo Galilei, discoverer of Jupiter's moons. In any case, it is intended to be the first of 30 satellites of the constellation. These satellites will orbit in 3 planes, 10 in each plane at approximately 3600 km higher than the GPS constellation. Like the GPS system, GALILEO will utilize *CDMA*. The full constellation of GALILEO satellites and full operational capability is expected to be in place around 2010–2020.

The Galileo Joint Undertaking (*GJU*) is the body set up by the European Commission and the European Space Agency (*ESA*) to oversee Galileo's development phase; this phase may be financed in public–private partnerships (*PPP*). And it is the responsibility of the GJU to help mobilize the public and private sector funds required to complete the various phases of the program. In other words, GALILEO will be controlled by a civilian agency with a more of a business-operating model than is the case of GPS. The European Commission owns the physical system, the ground stations, and satellites and so on. They are a public asset. Nevertheless, the day-to-day operations will be the responsibility of a concessionaire. GALILEO is a civil system and is clearly designed to reduce Europe's dependence on GPS.

GIOVE-A AND GIOVE-B

In fact one of the motivations for launching GIOVE-A was to allow European government authorities to register its Galileo frequencies with international regulators. Registration is necessary to prevent the frequency registration from expiring. It has done its job and continues in orbit and bought time for Europe to build additional satellites without facing a confiscation of its frequency reservations.

The first follow-on satellite GIOVE-B is more like the satellites that will eventually comprise the GALILEO constellation than is GIOVE-A. As an illustration of the partnerships necessary for the success of GALILEO, it is instructive to note that Galileo Industries, the consortium that is building GIOVE-B and the first 4 operational Galileo satellites, has scheduled their launches aboard Russian Soyuz rockets.

GALILEO Signals and Services

The GALILEO signals are known as L1, E5a, and E5b (Figure 8.12). These signals will be compatible with the existing L1 GPS signal and the coming L5 signal. The system will also broadcast a third frequency band—E6.

GALILEO has defined five levels of service that will be provided by the system. They include the *Open Service,* which uses the basic signals and is quite similar to GPS and GLONASS. The *Safety of Life Service* is along the same line but provides increased guarantees including integrity monitoring, meaning that users are warned if there are signal problems. The *Public Regulated Service* (*PRS*) is encrypted and is meant to assist public security and civil authorities. It provides users with protection against jamming. The *Search and Rescue Service* is intended to enhance space-based services and improve response time to distress beacons and alert messages. Encrypted custom solutions for unique applications are provided in the *Commercial Service*. The business model is still under development, but the GALILEO concessionaire, while delivering agreed service for the other four, will probably find that the Commercial Service will generate the most profits.

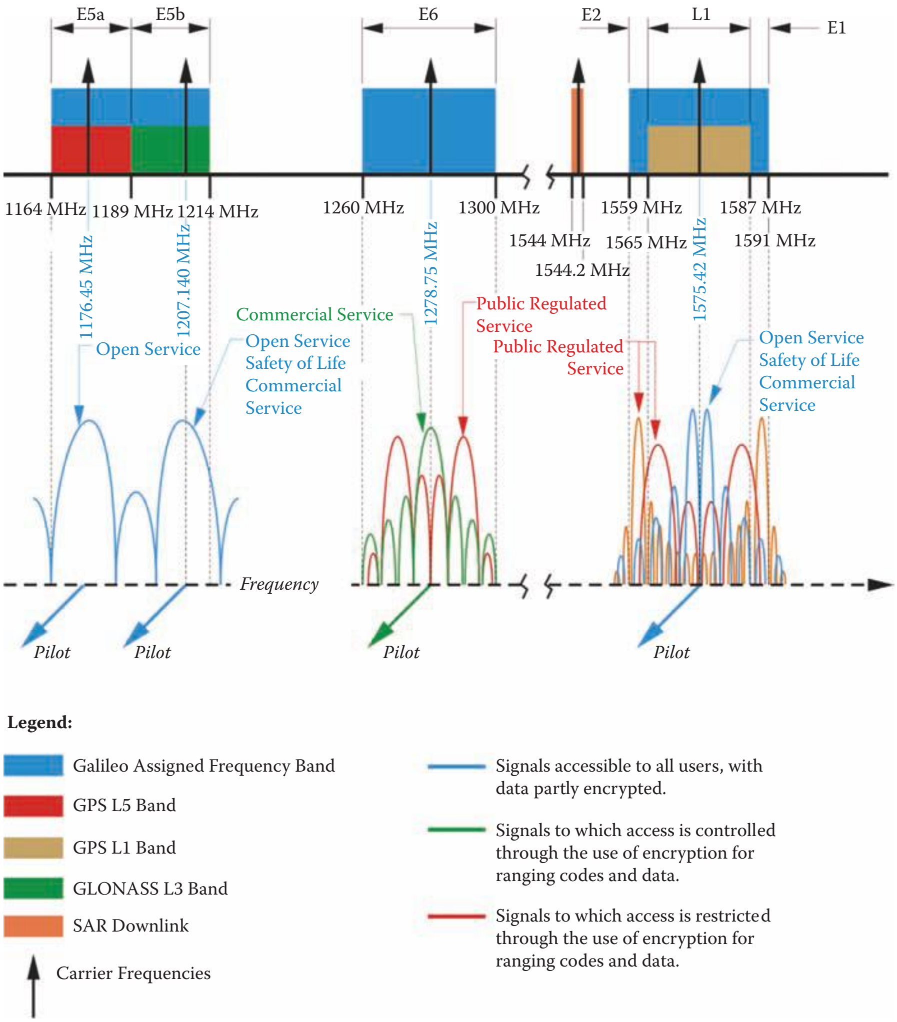


FIGURE 8.12 GALILEO–GPS–GLONASS Signals (Courtesy of Hein, Guenter W., Director of the Institute of Geodesy and Navigation, University FAF Munich, http://www.ifen.unibw-muenchen.de/research/signal.htm.)

INTEROPERABILITY

Any discussion of interoperability between GPS and GALILEO must consider the overlapping signals. It is helpful that the signals center on the same frequency if they are to be used in a combined fashion. For example, please recall that the third GLONASS civil reference signal on L3 that will be available on the K satellites will be within a new frequency band that includes 1201.743-1208.511 MHz and will overlap GALILEO's E5b signal.

In Figure 8.12, the GALILEO signals are shown on the top and the GPS signals on the bottom. The GALILEO satellites broadcast signals in several frequency ranges including 1176-1207 MHz, near GPS L5. Please note in Figure 8.12 that GALILEO's E5a signal is centered exactly at 1176 as is L5. The other overlapping signals can be seen at 1575.42 MHz where both GALILEO's L1 and the GPS Ll frequency are both centered. Also, notice that in both cases the GPS signal is based on the binary phase shift key (BPSK) and the GALILEO signal is accomplished with the binary offset carrier (BOC) method. The compatibility of these methods can be seen graphically in Figure 8.12. An important characteristic of BOC modulation is that the code's greatest power density is at the edges that is at the *nulls* which, as it did with the M code on GPS, mitigates interference with the existing codes. In this case, not only will there not be interference between the codes on GALILEO and GPS where they overlap, they can actually be used together. GALILEO also has a signal E6 at 1278.75 MHz. As you can see this band does not overlap any GPS frequency; however, it does happen to coincide with the band that Russia is considering for L3 on GLONASS.

GALILEO Signals

There are GALILEO signals available to all users; they are known as *Open Service* or *OS*. They include three data-less channels or *pilot tones*. Pilot tones are ranging codes not modulated by data.

The signal E5 will be spread from 1164 to1215 MHz. If they are separately modulated, E5a will be centered on 1178.45 MHz; this corresponds with the coming GPS L5. And E5b at 1207.14 MHz. will be in the range of GLONASS L3.

From 1260 to 1300 MHz the signal designated E6 is part of the Radio Navigation Satellite Service (RNSS) allocation for GALILEO. The GALILEO signal E2-L1-E1 from 1559 to 1592 MHz is also part of the Radio Navigation Satellite Service. This signal is often known as simply L1. That is a convenient name since the GPS L1 is right there too. Spectral separation of GPS and GALILEO L1 signals is accomplished by use of different modulation schemes. This strategy allows jamming of civil signals, if that should prove necessary, without affecting GPS M-code or the Galileo PRS service. You can see the modulation method—*BOC or BPSK*, chipping rates, data rates in Figure 8.12. Also please note the places where the carrier frequencies and frequency bands are common between GPS, GLONASS, and GALILEO.

There are also two signals on E6 with encrypted ranging codes, including one data-less channel that is only accessible to users who gain access through a given Commercial Service, CS, provider. And last, there are two signals, one in E6 band and one in E2–L1–E1, with encrypted ranging codes and data that are accessible to authorized users of the Public Regulated Service (*PRS*).

Frequency Coincidence

The fortuitous coincidences of frequencies between GPS and GALILEO did not happen without discussion. As negotiations proceeded between the United States and the European Union, one of the most contentious issues arose just as the European Union was moving to get GALILEO off the ground. They announced their intention to overlay GALILEO's Public Regulated Service (PRS) code on the U.S. Military's M-code. The possibility that this would make it difficult for the DoD to jam the GALILEO signal in wartime without also jamming the U.S. signal was considered.

It became known as the M-code overlay issue. In June of 2004, the United States and the European Union reached an agreement that ensured the GALILEO's signals would not harm the navigation warfare capabilities of the United States and the North Atlantic Treaty Organization (NATO).

So it looks as if some of the hurdles to interoperability between GALILEO and GPS are falling away. If the two systems can be compatible, when GALILEO is fully operational they will provide more than twice the signal-in-space resource available to Global Navigation Satellite System (GNSS) users today.

It looks as if interoperability between its L1 and E5A frequencies and the GPS L1 and L5 frequencies, respectively, can be accomplished. Therefore, for 1- to 10-meter accuracy the frequencies are already matched at L1 and L5, and the code rates are synchronized around the multiples of 1.023 MHz. However, submeter users may have some issues with time drift between the systems. The problem may require a broadcast correction.

BEIDOU/COMPASS

Perhaps it is appropriate to say a word or two about the Chinese System. The fourth GNSS system, joining those undertaken by the United States—*GPS*, Russia—*GLONASS* and Europe—*GALILEO*, will be the Chinese Beidou or *North Dipper,* also known as *Compass*. The system is already operational but is expected to expand substantially. The Chinese government had launched 3 Beidou geostationary (GEO) Compass Navigation Satellite System (*CNSS*) satellites, by May 2003. The Beidou-1 and 1B, launched in 2000 at 140° and 80° E longitude, respectively, were followed by 1C in 2003 over 110.5° E longitude. Since the system only requires two satellites to function 1C is actually an in-orbit spare.

In an announcement made through the government news agency, Xinhua, the People's Republic of China National Space Administration said that two more GEO will be launched in 2007 on the way to an enhanced system of up to 35 satellites that will cover all of China and neighboring nations. It is expected to include 5 GEO and 30 MEO or medium earth orbit satellites in six orbital planes, the latter operating near the GPS and GALILEO altitude of 20,000 kilometers.

There will be two levels of service. One will be a 10-m Open Service and the other will be an Authorized Service. The Open Service will offer an accuracy of 10m, 0.2 mps velocity accuracy, and timing accuracy within 50 nanoseconds. The Authorized Service will be available to subscribers and will provide more reliable positioning and system integrity information. This bifurcation will probably resemble GALILEO's publicly regulated service (*PRS*), or the encrypted *P code* and *M* GPS military signals. However, details are difficult to come by.

This is all actually a bit of a surprise since there had been some expectation that China would preempt the expansion of its system in favor of participation in the GALILEO project.

FREQUENCIES

Frequency requests have been made by China for bands for the Beidou signals that may interfere with the Galileo Public Regulated Service (PRS) and the GPS M-code on Ll and L2. The requests are under consideration by the International Telecommunications Union (ITU), a United Nations affiliate responsible for achieving handling the use of radio spectrum worldwide.

China has said that it is willing to cooperate with other countries in arranging Beidou so that it will operate with other global satellite positioning systems. However, the proposal to overlay the M-code would prevent the U.S. Military from jamming Beidou transmissions without jamming its own signals. This would raise an alarm in the Department of Defense as did the original GALILEO Public Regulated Service design. The issue is also a concern to the European Union, which has wanted to engage China's support in GALILEO.

THE QUASI-ZENITH SATELLITE SYSTEM (QZSS)

The Japanese Quasi-Zenith Satellite System (*QZSS*) was originally proposed by a private sector consortium, but now the Japanese government plans to launch 3 geosynchronous satellites broadcasting GPS-like signals. The configuration is intended to provide satellites at high elevation angles over Japan. This is the origin of the term *quasi-zenith*. It is actually a multisatellite augmentation system designed to benefit modified GPS receivers operating in areas with significant signal obstructions such as urban canyons. The first demonstration QZSS satellite will be launched in 2008 and the satellites will also pass over parts of the Asia–Pacific region and will effectively increase the number of satellites available to suitably equipped GPS users in that region.

THE FUTURE

So what is coming? Some day there may be as many as 80 satellites from GPS, GLONASS, GALILEO, QZSS, and others. If so the systems will provide users with a variety of signals and codes. The availability of many more satellites will enable new applications in areas where the current lack of satellites has been a hindrance. For civil users, new signals will provide more protection from interference, ability to compensate for ionospheric delays with pseudoranges, and wide-laning or even tri-laning capabilities. For military users, there will be greater antijam capability and security. For everybody, there will be improvement in accuracy, availability, integrity, and reliability.

It looks as if the GNSS is on its way. So it is no surprise that there is great anticipation from a business perspective, but from a user's point of view the situation is not unlike the advent of GPS more than 30 years ago. Much is promised but little assured. New capabilities will be available, but exactly what and exactly when is by no means certain. Nevertheless, it is prudent to consider the ramifications of a constellation including QZSS, GLONASS, Beidou, GALILEO, and GPS satellites.

MORE SATELLITES

How many more? GPS and GLONASS together provide the user with ~2 times the satellites than does GPS alone. In other words, if one considers that 6 satellites are normally above a user's horizon with GPS alone, there will usually be about 12 available with GPS and GLONASS combined. If GPS and GALILEO are considered together there are ~2½ times or about 15 satellites typically available to a user. The number increases to 21 or ~3½ times more satellites with all three GPS, GLONASS, and GALILEO together and particularly if Beidou and QZSS are included.

Accessibility

In a sense, the more satellites the better the performance, particularly among trees and in urban canyons, those places where signals bounce and scatter, and multipath abounds.

Flexibility

When more satellites are overhead, the user has more flexibility. For example, since there are 6 satellites in a window available to a GPS receiver in Figure 8.13, the user may be able to increase the mask angle to decrease the multipath and still have 4 satellites to observe. Imagine if there were 12, 15, or even 21 satellites in the picture and you can see how more satellites can mean better accessibility in restricted environments.

Reliability

Also, the more diverse the maintenance of the components of GNSS, the less chance of overall system failure. The United States, Russia, Japan, the European Union, and China all have infrastructure in place to support their contribution to GNSS. Under such circumstances simultaneous outages across the entire GNSS constellation are extremely unlikely.

Faster Positioning

More measurements in shorter time means observation periods can be shortened without degrading accuracy and interference can be ameliorated more easily. In short, better accuracy can be achieved in less time.

Faster Initialization

Also, with more satellites available, the time to first fix for carrier-phase receivers, the period when the receiver is solving for the integers, downloading the almanac and so forth, also know as *initialization*, will be shortened significantly. And fixed solution accuracy will be achieved more quickly. Today dual-frequency carrier-phase solutions are accurate but noisy, but with the new signals available on L2C, L5, and other GNSS signals, dual-frequency solutions will be directly enhanced.

GPS Accuracy

As discussed, from the beginning it has been, and continues to be, possible to achieve centimeter, even millimeter, positional accuracy with GPS. Using the signals of two carrier frequencies, *L1 and L2,* and two PRN codes, *P and C/A,* on all available satellites the users of GPS can have accuracy commensurate with virtually

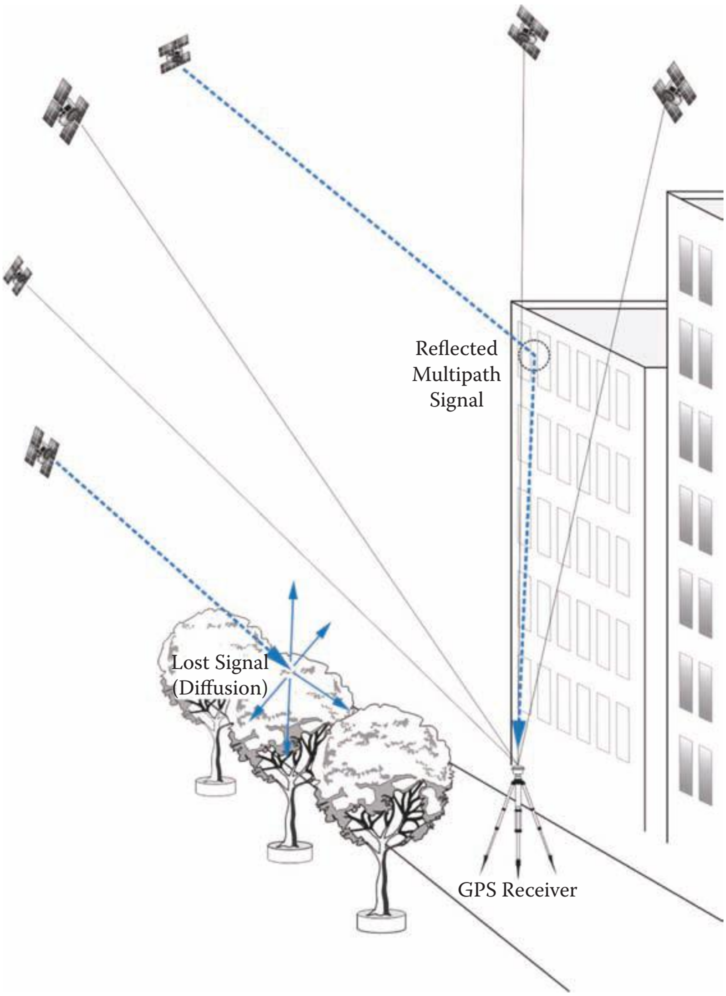


FIGURE 8.13 The Urban Canyon

any requirement. The cost and time required to accomplish this high accuracy has declined steadily, but are still more substantial than many would like.

While a GNSS capable receiver may offer a user improved availability and reliability, it may not necessarily offer higher accuracy than is available from GPS. However, the achievement of high accuracy more conveniently and in more places—that seems to be within reach with GNSS.

GNSS Accuracy—Faster

When more satellites are available to a real-time GPS solution, high accuracy is also available. The same is true with GNSS, and along with more signals that means better ionospheric correction too. Remember the ionospheric delay is frequency dependent. More signals also means the number of observations available for ambiguity resolution increases and the integers can be fixed more rapidly. Also consider the utility of dual-frequency measurements in GPS. Three frequencies, that is, L5, will increase performance even more.

Simplification

The algorithms currently necessary for the achievement of high accuracy with carrier-phase ranging may be simplified since many of the new GNSS signals will be carrying a civilian code. Generally speaking, code correlation is a more straightforward problem than is carrier differencing. This may lead to less complicated receivers. This presents the possibility that they will be less expensive.

INTEROPERABILITY

GPS–GALILEO–GLONASS CONSTELLATIONS

Inconsistency

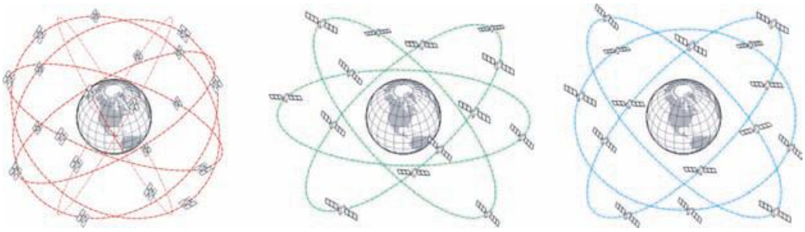
Despite similarities, there are some issues in the consistency GNSS issues as you know. For example, you may recall that the roll out of the new systems is not coordinated. In other words GPS Modernization, GLONASS replenishment, and QZSS and Galileo deployment have not been synchronized. Despite much cooperation there is no clear agreement among nations that launches and operational capabilities of GLONASS, GALILEO, GPS, QZSS, and Beidou will happen in the same time frame. Also there are differences between GLONASS and GPS regarding CDMA and FDMA. There are also differences in the time standards between the two systems. Not to mention the overlap between the GPS M-code and Beidou. It is important that these inconsistencies are worked out technologically because apart from the less sophisticated applications for GNSS interoperability will be required. For the full potential of the system to realize multiple GNSS, frequencies will need to work together. In other words they require as many satellites as possible delivering signals that can be used in conjunction with one another any time, any place.

Consistency

However, there is signal compatibility among subsets of the 80 satellites that will be broadcasting at the same frequencies. Interoperability is achieved by partial frequency overlap using different signal structures and/or different code sequences for spectral separation. In Figure 8.12, you can see the overlap of GPS L1 and GALILEO L1. We can also look forward to GPS L5 and GALILEO E5a. It may also be possible for GLONASS L3 to be interoperable with GALILEO E5b. Also please note that the Galileo satellites will make use of code division multiple access CDMA techniques which, as you know, are compatible with the GPS approach.

Robust Solutions

In fact, high accuracy and interoperability are not only a matter of convenience—robust, reliable solutions are becoming a business necessity. Consider *safety-of-life*



GPS
– 6 Orbital Planes
– 24 Satellites + Spare
– 55° Inclination Angle
– Altitude 20,200 km

GALILEO
– 3 Orbital Planes
– 27 Satellites + 3 Spares
– 56° Inclination Angle
– Altitude 23,616 km

GLONASS
– 3 Orbital Planes
– 21 Satellites + 3 Spares
– 64.8° Inclination Angle
– Altitude 19,100 km

FIGURE 8.14 Some GNSS Constellations

uses for things such as routing of emergency vehicles, or the GPS based automated machine control system now in used in construction. Mining, agriculture, aircraft control, and so forth, are depending more and more on satellite navigation systems. These industries have high costs and high risks and not only require high accuracy but reliability as well. If GNSS can deliver inexpensive receivers tracking the maximum number of satellites broadcasting the maximum number of signals, it will live up to the fondest hopes of not only many individuals but also many industries as well.

Summary

Though certainly not assured, it is possible that receivers that track GPS alone will soon be able to utilize carrier phase on L1, be codeless on L2 with L2C, and perhaps L5 if the Block IIF satellites are operational. Further, when it is available, receivers that track only GALILEO will be able to utilize L1, E5, and E6 from a full, or nearly full, constellation. It is probable that there will be a charge for observation of GALILEO's E6 signals. However, receivers with both GPS and GALILEO capability may have the carrier phase on L1, codeless on L2 with L2C, and L5 as well as Galileo L1 and E5 signals. It is also possible that some receivers may also be available to track GLONASS and QZSS.

Some years later, the modernized GPS constellation may be in place, including L5 and GALILEO. Just considering GPS and GALILEO there could be 60 or so satellites in orbit and available. Including GLONASS and QZSS it is feasible that ~80 satellites could be available. If these constellations (such as those shown in Figure 8.14) become reality a typical user could find 10-20 satellites above the horizon anywhere, any time.

In fact, the goal of a single receiver that can track all the old and new satellite signals with a significant performance improvement looks possible. But after all, the main attraction of interoperability between these systems is the greatly increased number of satellites and signals, better satellite availability, better dilution of precision, immediate ambiguity resolution on long baselines with three-frequency data, better accuracy in urban settings, and fewer multipath worries. Those are some of the things we look forward to. It is beginning to look like at least some of those things are achievable.

EXERCISES

1. How many GPS satellites are needed for worldwide 24-hour coverage?

   (a) 18 satellites
   (b) 12 satellites
   (c) 20 satellites
   (d) 24 satellites

2. Which of the following statements is true of the current GPS constellation?

   (a) There are 4 orbital planes.
   (b) There are 8 satellites in each orbital plane.
   (c) The satellites orbit approximately 20,000 nautical miles above the earth.
   (d) Each satellite completes an orbit in 12 sidereal hours.

3. Which of the following statements is not true of the Block I and II satellites?

   (a) All of the Block II satellites built by Rockwell have been launched.
   (b) None of the Block I satellites remain in service.
   (c) The Block I satellites will broadcast the new M-code.
   (d) The Block I satellites were launched from 1978 to 1985.

4. Eight to twelve of the new Block IIR satellites are going to be modified to broadcast a new military code called the M-code. What code will the M-code eventually replace?

   (a) The navigation code.
   (b) The carrier.
   (c) The P(Y) code.
   (d) The C/A code.

5. Which will carry the M-code?

   (a) L1
   (b) L1 and L2
   (c) L2
   (d) L5

6. Which of the following statements about L2C is not correct?

   (a) It will be broadcast by the IIR-M satellites.
   (b) It was originally announced more than 5 years ago.
   (c) It will be carried on L1 and L2.
   (d) It will have somewhat higher power than the C/A code.

7. What characteristic of L5 is an improvement over both L1 and L2?

   (a) It will have a higher chipping rate.
   (b) Its frequency will be a multiple of the GPS fundamental clock rate.
   (c) It will have codes broadcast in quadrature.
   (d) It will be available on the Block IIF satellites.

8. The L5 will be a bit stronger than the current carriers. How much stronger?

   (a) 6 db
   (b) 2 db
   (c) 10 db
   (d) 12 db

9. Which country will be assisting Russia in launching new GLONASS satellites?

   (a) China
   (b) Japan
   (c) India
   (d) Taiwan

10. A sticking point between GALILEO and GPS was the possible conflict between the EU's PRS code and which GPS code?

    (a) C/A code
    (b) P(Y) code
    (c) L2C code
    (d) M code

11. Current plans call for the GALILEO constellation to be in place in what year?

    (a) 2005
    (b) 2008
    (c) 2010
    (d) 2012

12. Which satellites will be the first to broadcast L5?

    (a) Block I
    (b) Block II
    (c) Block IIR
    (d) Block IIF

13. On the Block IIF satellites what is a code that will be carried on L1 and not carried on L2?

    (a) C/A code
    (b) P(Y) code
    (c) L2C code
    (d) M code

14. On the Block IIF satellites what is a code that will be carried on L2 and not carried on L1?

    (a) C/A code
    (b) P(Y) code
    (c) L2C code
    (d) M code

15. Which of the following statements is not true of the Block IIR satellites?

    (a) Block IIR satellites can determine their own position.
    (b) Block IIR satellites have programmable processors onboard to do their own fixes in flight.
    (c) Block IIR satellites can be moved into a particular orbit with a 60-day advanced notice.
    (d) Block IIR satellites cost about one-third more than the earlier Block II satellites did.

ANSWERS AND EXPLANATIONS

1. Answer is (d)

Explanation: The basics of its configuration are well known. Twenty-four GPS satellites are needed for 24-hour worldwide coverage but there are actually more than 24 birds up at any one time to assure that the minimum will always be available.

2. Answer is (d)

Explanation: The satellites are in 6 orbital planes, 4 or more satellites per plane in orbit about 11,000 nautical miles above the earth. Each completes an orbit every 12 hours sidereal time.

3. Answer is (c)

Explanation: All the Block II's built by Rockwell have been launched. The Block orbit satellites were launched first, 1978–1985. None are functional today. Now 8 to 12 of these, IIR, replenishment satellites are going to be modified to broadcast a new military code. It is known as the M-code. This code will be carried on both L1 and L2 probably will replace the P(Y) code. It has the advantage of allowing the DoD to increase the power of the code to prevent jamming. The upgraded IIR satellites that have the M-code are known as IIR-M. See Figure 8.7.

4. Answer is (c)

Explanation: This code will replace the P(Y) code. See Figure 8.6.

5. Answer is (b)

Explanation: This code will be carried on both L1 and L2.

6. Answer is (c)

Explanation: A new military code may not be terribly exciting to civilian users like us, but these IIR-M satellites have something else going for them. They will broadcast a new civilian code. This is a code that was announced way back in March 1998. It will be on L2 and will be known as L2C. And this new code is not a merely a copy of the C/A-code. L2C will be a bit more sophisticated and have somewhat higher power.

7. Answer is (a)

Explanation: Alright, L2C is fine, but what about the new carrier everybody has been talking about, L5? It will be centered on 1176.45 MHz, 115 times the fundamental clock rate, in the Aeronautical Radio Navigation Services (ARNS) band. The L5 codes will have a higher chipping rate. L5 will be first introduced on the fourth generation GPS satellites, Block IIF. Even though both codes are broadcast on L1, they are distinguishable from one another by their transmission in *quadrature*. That means that the C/A code modulation on the L1 carrier is phase shifted 90 degrees from the P code modulation on the same carrier. L5 will use the same strategy.

8. Answer is (a)

Explanation: It will have higher power than L1, about 6 dB higher.

9. Answer is (c)

Explanation: According to the *Moscow Times* India will be helping Moscow get the system to that point.

10. Answer is (d)

Explanation: One of the most contentious issues arose just as the European Union was moving to get GALILEO off the ground. They announced their intention to overlay GALILEO's Public Regulated Service (PRS) code on the U.S. Military M-code.

11. Answer is (b)

Explanation: The current plan is for the GALILEO 30-satellite constellation to be fully operational in 2008.

12. Answer is (d)

Explanation: L5 will be first introduced on the fourth generation GPS satellites, Block IIF.

13. Answer is (a)

Explanation: See Figure 8.7.

14. Answer is (c)

Explanation: See Figure 8.7.

15. Answer is (d)

Explanation: There are some significant changes with the Block IIR satellites. Block IIR satellites can determine their own position. They have programmable processors onboard to do their own fixes in flight. The Block IIR satellites can be moved into a particular orbit with a 60-day advanced notice. In short they have more autonomy. They are also more radiation hardened then their predecessors and they cost about a third less than the Block II satellites did.

References

Brunner, F.K., and W.M. Welsch. 1993. Effect of the Troposphere on GPS Measurements. In *GPS World* 4(1): 42–51, Advanstar Communications.

Dvorkin, V., and S. Karurtin**.** 22 June 2006. GLONASS: Current Status and Perspectives, 3rd AllSAT Open Conference. Slide 13 of 24, Hannover.

Federal Geodetic Control Committee. 1 Aug. 1989. Geometric Geodetic Accuracy Standards and Specifications for Using GPS Relative Positioning Techniques. Version 5.

Issler, J-L., G.W. Hein, J. Godet, J.C. Martin, P. Erhard, R. Lucas-Rodriguez, and T. Pratt. June 2003. Galileo Frequency & Signal Design. *GPS World* 14(6):30–37.

Kalman, R. E. 1960. A New Approach to Linear Filtering and Prediction Problems. *Trans. of the ASME-Journal of Basic Engineering* 82:35–45.

Kaplan, E.D. 1996. *Understanding GPS, Principles and Applications.* Artech House Publishers: Boston, MA.

Langley, R.B. 1991. The GPS Receiver: An Introduction. *GPS World* 2(1):50–53, Advanstar Communications.

Langley, R.B. 1993. The GPS Observables. *GPS World* 4(4):52-59, Advanstar Communications.

Langley, R.B. Sept. 1998. RTK GPS Innovation. *GPS World* 9(9):70–76.

Lazar, S., Summer 2002. Crosslink. Aerospace Corporation, http://www.aero.org/publications/crosslink/summer2002/07.html.

Leick, A. The Least-Squares Toolbox. Dec. 1993. *ACSM Bulletin* (146).

National Geodetic Survey. 1986. Geodetic Glossary. *NOAA Technical Publications*, U.S. Department of Commerce:209.

Reilly, J.P. April 1996. GPS Calibration. *The GPS Observer in Point of Beginning* 21(5):20–23.

Reilly, J.P. May 1997. Elevations from GPS. *The GPS Observer in Point of Beginning* 22:24–26.

Reilly, J.P. Jan. 2000. GEOID99. *The GPS Observer in Point of Beginning* 25(4):12–13.

Sky DSP. 1.3.1 Frequency Division Multiple Access. http://www.skydsp.com/publications/4thyrthesis/chapter1.htm.

Trimble Navigation Limited. WAVE Baseline Processor and SV Azimuth & Elevation Table. GPSurvey Software Suite, Version 1.10D.

Trimble Navigation Limited. Model 4000ST GPS Surveyor, Operation Manual. July 1989 3.2X, page D-1.

U.S. Department of Commerce. NOAA, National Ocean Survey. National Geodetic Survey. Control Data Sheets.

U.S. Department of Commerce. NOAA, National Ocean Survey. National Geodetic Survey. DSDATA format specifications. www.ngs.noaa.gov.

van Dierendock, A. J., S. S. Russell, E. R. Kopitzke, and M. Birnbaum. 1978. The GPS Navigation Message. *Global Positioning System*. Papers published in *Navigation*.

Wells, D. (ed.) 1986. *Guide to GPS Positioning*. Fredericton, New Brunswick: Canadian GPS Associates, University of New Brunswick, pp. 8.2–8.11.

Glossary

Absolute Positioning (*Also known as* **Point Positioning, Point Solution, or Single Receiver Positioning**) A single receiver position, defined by a coordinate system. Most often the coordinate system used in absolute positioning is geocentric. In other words, the origin of the coordinate system is intended to coincide with the center of mass of the earth.

Accuracy The agreement of a value, whether measured or computed, with the standard or accepted true value. In the absolute sense, the true value is unknown and therefore, accuracy can only be estimated. Nevertheless, in measurement, accuracy is considered to be directly proportional to the attention given to the removal of systematic errors and mistakes. In GPS specifically, the values derived are usually the position, time, or velocity at GPS receivers.

Almanac A data file containing a summary of the orbital parameters of all GPS satellites. The almanac is found in subframe 5 of the Navigation Message. This information can be acquired by a GPS receiver from a single GPS satellite. It helps the receiver find the other GPS satellites.

Analog Representation of data by a continuous physical variable. For example, the modulated carrier wave used to convey information from a GPS satellite to a GPS receiver is an analog mechanism.

Antenna An antenna is a resonant device that collects and often amplifies a satellite's signals. It converts the faint GPS signal's electromagnetic waves into electric currents sensible to a GPS receiver. Microstrip, also known as patch antennas, are the most often used with GPS receivers. Choke-ring antennas are intended to minimize multipath error.

Antenna Splitter (*See also* **Zero Baseline Test**) An antenna splitter is an attachment that divides a single GPS antenna's signal in two. The divided signal goes to two GPS receivers. This is the foundation of Zero Baseline test. The two receivers are using the same antenna so the base line length should be measured as zero; since perfection remains elusive, it is usually a bit more.

Antispoofing (AS) The encryption of the P code to render spoofing ineffective. Spoofing is generation of false transmissions masquerading as the Precise Code (P Code). This countermeasure, Antispoofing, is actually accomplished by the modulation of a W-Code to generate the more secure Y-Code that replaces the P code. Commercial GPS receiver manufacturers are not authorized to use the P code directly. Therefore, most have developed proprietary techniques both for carrier wave and pseudorange measurements on L2 indirectly. Dual-frequency GPS receivers must also overcome AS. Antispoofing was first activated on all Block II satellites on January 31, 1994.

Anywhere Fix A receiver's ability to achieve lock-on without being given a somewhat correct beginning position and time.

Atomic Clock (*See also* **Cesium Clock**) A clock regulated by the resonance frequency of atoms or molecules. In GPS satellites the substances used to regulate atomic clocks are cesium, hydrogen, or rubidium.

Attosecond One-quintillionth (10^{18}) of a second.

Attribute Information about features of interest. Attributes such as date, size, material, color, and so forth are frequently recorded during data collection for Geographic Information Systems (GIS).

Availability The period, expressed as a percentage, when positioning from a particular system is likely to be successful.

Bandwidth The bandwidth of a signal is its range of frequencies. It is a measurement of the difference between the highest frequency and the lowest frequency expressed in Hertz. For example, the bandwidth of a voice signal is about 3 kHz, but for television signals it is around 6 MHz.

Baseline A line described by two stations from which GPS observations have been made simultaneously. A vector of coordinate differences between this pair of stations is one way to express a baseline.

Base Station (*Also known as* **Reference Station**) A known location where a static GPS receiver is set. The base station is intended to provide data with which to differentially correct the GPS measurements collected by one or more roving receivers. A base station used in Differential GPS (DGPS) is used to correct pseudorange measurements collected by roving receivers to improve their accuracy. In carrier-phase surveys, the base station data are combined with the other receivers' measurements to calculate double-differenced observations or used in real-time kinematic (RTK) configurations.

Beat Frequency When two signals of different frequencies are combined, two additional frequencies are created. One is the sum of the two original frequencies. One is the difference of the two original frequencies. Either of these new frequencies can be called a beat frequency.

Between-Epoch Difference The difference in the phase of the signal on one frequency from one satellite as measured between two epochs observed by one receiver. Because a GPS satellite and a GPS receiver are always in motion relative to one another the frequency of the signal broadcast by the satellite is not the same as the frequency received. Therefore, the fundamental Doppler observable in GPS is the measurement of the change of phase between two epochs.

Between-Receiver Single Difference The difference in the phase measurement between two receivers simultaneously observing the signal from one satellite on one frequency. For a pair of receivers simultaneously observing the same satellite, a between-receiver single difference pseudorange or carrier phase observable can virtually eliminate errors attributable to the satellite's clock. When baselines are short the between-receiver single difference can also greatly reduce errors attributable to orbit and atmospheric discontinuities.

Between-Satellite Single Difference The difference in the phase measurement between signals from two satellites on one frequency simultaneously observed by one receiver. For one receiver simultaneously observing two satellites, a between-satellite single difference pseudorange or carrier phase observable can virtually eliminate errors attributable to the receiver's clock.

Bias A systematic error. Biases affect all GPS measurements and hence the coordinates and baselines derived from them. Biases may have a physical origin, satellite orbits, atmospheric conditions, clock errors, and so forth. They may also originate from less than perfect control coordinates, incorrect ephemeris information, and so on. Modeling is one method used to eliminate or at least limit the effect of biases.

Binary Biphase Modulation The method used to impress the pseudorandom noise codes onto GPS carrier waves using two states of phase modulations, since the code is binary. In binary biphase modulation, the phase changes that occur on the carrier wave are either 0° or 180°.

Bit A unit of information. In a binary system either a 1 or a 0.

Block I, II, IIR, and IIF Satellites Classification of the GPS satellite's generations. Block I satellites were officially experimental. There were 11 Block I satellites. PRN4, the first GPS satellite, was launched February 22, 1978. The last Block I satellite was launched October 9, 1985, and none remain in operation. Block II satellites are operational satellites. The GPS constellation was declared operational in 1995. Twenty-seven Block II satellites have been launched. Block IIRs are replenishment satellites and have the ability to receive signals from other IIR satellites. These satellites have the capability to measure ranges between themselves. Block IIF will be the follow-on series of satellites in the new century.

Broadcast A modulated electromagnetic wave transmitted across a large geographical area.

Broadcast Ephemeris (*See* Ephemeris)

Byte A sequence of eight binary digits that represents a single character, that is, a letter, a number.

C/A Code (*Also known as* S-Code) A binary code known by the names, Civilian/Access code, Clear/Access code, Clear/Acquisition code, Coarse/Acquisition code, and various other combinations of similar words. It is a standard spread-spectrum GPS pseudorandom noise code modulated on the L1 carrier using binary biphase modulations. The C/A code is not carried on L2. Each C/A code is unique to the particular GPS satellite broadcasting it. It has a chipping rate 1.023 MHZ, more than a million bits per second. The C/A code is a direct sequence code and a source of information for pseudorange measurements for commercial GPS receivers.

Carrier An electromagnetic wave, usually sinusoidal, that can be modulated to carry information. Common methods of modulation are frequency modulation and amplitude modulation. However, in GPS the phase of the carrier is modulated and there are currently two carrier waves. The two carriers in GPS are L1 and L2, which are broadcast at 1575.42 MHz and 1227.60 MHz, respectively.

Carrier Beat Phase (*See also* **Beat Frequency; Doppler Shift**) The phase of a beat frequency created when the carrier frequency generated by a GPS receiver combines with an incoming carrier from a GPS satellite. The two carriers have slightly different frequencies as a consequence of the Doppler shift of the satellite carrier compared with the nominally constant receiver generated carrier. Because the two signals have different frequencies, two beat frequencies are created. One beat is the sum of the two frequencies. One is the difference of the two frequencies.

Carrier Frequency In GPS the frequency of the unmodulated signals broadcast by the satellites. Currently, the carrier frequencies are, L1 and L2, which are broadcast at 1575.42 MHz and 1227.60 MHz, respectively.

Carrier Phase (*Also known as* **Reconstructed Carrier Phase**) (1) The term is usually used to mean a GPS measurement based on the carrier signal itself, either L1 or L2, rather than measurements based on the codes modulated onto the carrier wave. (2) Carrier phase may also be used to mean a part of a full carrier wavelength. An L1 wavelength is about 19 cm, and an L2 wavelength is about 24 cm. A part of the full wavelength may be expressed in a phase angle from 0° to 360°, a fraction of a wavelength, cycle, or some other way.

Carrier Tracking Loop (*See also* **Code Tracking Loop**) A feedback loop that a GPS receiver uses to generate and match the incoming carrier wave from a GPS satellite.

CEP (*See* **Circular Error Probable**)

Cesium Clock (*Also known as* **Cesium Frequency Standard**) An atomic clock that is regulated by the element cesium. Cesium atoms, specifically atoms of the isotope Cs-133, are exposed to microwaves and the atoms vibrate at one of their resonant frequencies. By counting the corresponding cycles, time is measured.

Channel A channel of a GPS receiver consists of the circuitry necessary to receive the signal from a single GPS satellite on one of the two carrier frequencies. A channel includes the digital hardware and software.

Check Point A reference point from an independent source of higher accuracy used in the estimation of the positional accuracy of a data set.

Chipping Rate In GPS the rate at which chips, binary 1s and 0s, are produced. The P code chipping rate is about 10 million bits per second. The C/A code chipping rate is about 1 million bits per second.

Circular Error Probable (CEP) A description of two-dimensional precision. The radius of a circle, with its center at the actual position, that is expected to be large enough to include half (50%) of the normal distribution of the scatter of points observed for the position.

Civilian Code (*See* **C/A Code**)

Civilian/Access Code (*See* **C/A Code**)

Class of Survey Class of Survey is a means to generalize and prioritize the precision of surveys. The foundation of the categories is often taken from geodetic surveying classifications supplemented with standards of higher precision applicable to GPS work. Different classifications frequently apply to horizontal and vertical surveys. The categories themselves, their notation, and accuracy tolerances are unique to each nation. A Class of Survey should reflect the quality of network design, the instruments and methods used, and processing techniques as well as the precision of the measurements.

Clear/Access Code (*See* C/A Code)

Clear/Acquisition Code (*See* C/A Code)

Clock Bias The discrepancy between a moment of time per a GPS receiver's clock or a GPS satellite's clock and the same moment of time per GPS Time, or another reference, such as Coordinated Universal Time or International Atomic Time.

Clock Offset A difference between the same moments of time as indicated by two clocks.

Coarse/Acquisition (C/A) (*See* C/A Code)

Code Phase Measurements based on the C/A code rather than measurements based on the carrier waves. Sometimes used to mean pseudorange measurements expressed in units of cycles.

Code Tracking Loop A feedback loop used by a GPS receiver to generate and match the incoming codes, C/A or P codes, from a GPS satellite.

Complete Instantaneous Phase Measurement (See Fractional Instantaneous Phase Measurement) A measurement including the integer number of cycles of carrier beat phase since the initial phase measurement.

Confidence Level The probability that the true value is within a particular range of values, expressed as a percentage.

Confidence Region A region within which the true value is expected to fall, attended by a confidence level.

Constellation (1) The Space Segment, all GPS satellites in orbit. (2) The particular group of satellites used to derive a position. (3) The satellites available to a GPS receiver at a particular moment.

Continuously Operating Reference Stations (CORS) A system of base stations that originated from the stations built to support air and marine navigation with real-time differential GPS correction signals.

Continuous-Tracking Receiver (*See* Multichannel Receiver)

Control Segment The U.S. Department of Defense's network of GPS monitoring and upload stations around the world. The tracking data derived from this network is used to prepare, among other things, the broadcast ephemerides and clock corrections by the Master Control Station at Schriever Air Force Base (formerly Falcon AFB), Colorado Springs, Colorado. Then included in the Navigation Message, the calculated information is uploaded to the satellites.

Coordinated Universal Time (*Also known as* UTC, Zulu Time) Universal Time systems are international, based on atomic clocks around the world. Coordinated atomic clock time called Tempes Atomique International (TAI) was established in the 1970s and it is very stable. It is more stable that the actual rotation of the earth. TAI would drift out of alignment with the planet if leap seconds were not introduced periodically as they are in UTC. UTC is in fact one of several of the Universal Time standards. There are standards more refined than UTC's tenth of a second. UTC is maintained by the U.S. Naval Observatory (USNO).

Correlation-Type Channel A channel in a GPS receiver used to shift or compare the incoming signal with an internally generated signal. The code generated by the receiver is cross-correlated with the incoming signal to find the correct delay. Once they are aligned the delay lock loop keeps them so. Correlator designs are sometimes optimized for acquisition of signal under foliage, accuracy, multipath mitigation, and so forth.

Cutoff Angle (*See* **Mask Angle)**

Cycle Ambiguity (*Also known as* **Integer Ambiguity)** The number of full wavelengths, the integer number of wavelengths, between a particular receiver and satellite is initially unknown in a carrier phase measurement. It is called the cycle or integer ambiguity. If a single frequency receiver is tracking several satellites there is a different ambiguity for each. If a dual-frequency receiver is tracking several satellites, there is a different ambiguity for each satellite's L1 and L2. However, in every case the ambiguity is a constant number as long as the tracking, lock, is not interrupted. However, should the signal be blocked, if a cycle slip occurs, there is a new ambiguity to be resolved. Once the initial integer ambiguity value is resolved in a fixed solution for each satellite-receiver pair, the integrated carrier phase measurements can yield very precise positions. But the cycle ambiguity resolution processes actually utilize double-differenced carrier phase observables, not single satellite-receiver measurements. The great progress in carrier phase GPS systems has reduced the length of observation data needed, as in rapid static techniques. The integer ambiguity resolution can now occur even with the receiver in motion, the on-the-fly approach.

Cycle Slip A discontinuity of an integer number of cycles in the carrier phase observable. A jump of a whole number of cycles in the carrier tracking loop of a GPS receiver. Usually the result of a temporarily blocked GPS signal. A cycle slip causes the cycle ambiguity to change suddenly. The repair of a cycle slip includes the discovery of the number of missing cycles during an outage. Cycle slips must be repaired before carrier phase data can be successfully processed in double-differenced observables.

Data Message (*See* **Navigation Message)**

Data Logger (*Also known as* **Data Recorder and Data Collector)** A data entry computer, usually small, lightweight, and often handheld. A data logger stores information supplemental to the measurements of a GPS receiver.

Data Transfer Transporting data from one computer or software to another, often accompanied by a change in the format of the data.

Dataset An organized collection of related data compiled specifically for computer processing.

Datum (1) Any reference point line or surface used as a basis for calculation or measurement of other quantities. (2) A means of relating coordinates determined by any means to a well-defined reference frame. (3) The singular of data.

Decibel (dB) Decibel, a tenth of a bel. The bel was named for Alexander Graham Bell. Decibel does not actually indicate the power of the antenna, it refers to a comparison. Most GPS antennas have a gain of about 3 dB (decibels). This indicates that the GPS antenna has about 50% of the capability of an *isotropic* antenna. An isotropic antenna is a hypothetical, lossless antenna that has equal capabilities in all directions.

DGPS (*See* Differential GPS)

Differential GPS (1) A method that improves GPS pseudorange accuracy. A GPS receiver at a base station, a known position, measures pseudoranges to the same satellites at the same time as other roving GPS receivers. The roving receivers occupy unknown positions in the same geographic area. Occupying a known position, the base station receiver finds corrective factors that can either be communicated in real-time to the roving receivers, or may be applied in postprocessing. (2) The term *Differential GPS* is sometimes used to describe Relative Positioning. In this context it refers to more precise carrier phase measurement to determine the relative positions of two receivers tracking the same GPS signals in contrast to Absolute, or Point, Positioning.

Digital Involving or using numerical digits. Information in a binary state, either a one or a zero, a plus or a minus, is digital. Computers utilize the digital form almost universally.

Dilution of Precision (DOP) In GPS positioning, an indication of geometric strength of the configuration of satellites in a particular constellation at a particular moment and hence the quality of the results that can be expected. A high DOP anticipates poorer results than a low DOP. A low DOP indicates that the satellites are widely separated. Since it is based solely on the geometry of the satellites, DOP can be computed without actual measurements. There are various categories of DOP, depending on the components of the position fix that are of most interest.

Standard varieties of Dilution of Precision:

PDOP: Position (three coordinates)
GDOP: Geometric (three coordinates and clock offset)
RDOP: Relative (normalized to 60 seconds)
HDOP: Horizontal (two horizontal coordinates)
VDOP: Vertical (height)
TDOP: Time (clock offset)

Distance Root Mean Square (drms) A statistical measurement that can characterize the scatter in a set of randomly varying measurements on a plane.The drms is calculated from a set of data by finding the root-mean-square value of the radial errors from the mean position. In other words, the root-mean-square value of the linear distances between each measured position and the ostensibly true location. In GPS positioning, 2 drms is more commonly used. Two drms does *not* mean two-dimensional (2D) rms. Two drms *does* mean twice the distance root mean square. In practical terms, a particular 2 drms value is the radius of a circle that is expected to contain from 95% to 98% of the positions a receiver collects in one occupation, depending on the nature of the particular error ellipse involved. Two drms is convenient to calculate. It can be predicted using covariance analysis by multiplying the HDOP by the standard deviation of the observed pseudoranges.

Dithering Intentional introduction of digital noise. Dithering the satellite's clocks is the method the Department of Defense (DoD) uses to degrade the accuracy of the Standard Positioning Service. This degradation was known as Selective availability. Selective availability was switched off on May 2, 2000 by presidential order.

DoD Department of Defense

DOT Department of Transportation

DOP (*See* Dilution of Precision)

Doppler-Aiding A method of receiver signal processing that relies on the Doppler shift to smooth tracking.

Doppler Shift In GPS, a systematic change in the apparent frequency of the received signal caused by the motion of the satellite and receiver relative to one other.

Double-Difference A method of GPS data processing. In this method simultaneous measurements made by two GPS receivers, either pseudorange or carrier phase, are combined. Both satellite and receiver clock errors are virtually eliminated from the solution. Most high-precision GPS positioning methods use double-difference processing in some form.

Dual-Capability GPS/GLONASS Receiver (*Also known as* Interoperability) A receiver that has the capability to track both GPS satellites and GLONASS satellites.

Dual-Frequency Receivers or GPS measurements that utilize both L1 and L2. Dual-frequency implies that advantage is taken of pseudorange and/or carrier phase on both L-band frequencies. Dual-frequency allows modeling of ionospheric bias and attendant improvement in long baseline measurements particularly.

Dynamic Positioning (*See* Kinematic Positioning)

Earth-Centered Earth-Fixed (ECEF) A Cartesian system of coordinates with three axes in which the origin of the 3D-system is the earth's center of mass, the geocenter. The z-axis passes through the North Pole, that is, the International Reference Pole (IRP) as defined by the International Earth Rotation Service (IERS). The x-axis passes through the intersection of zero longitude, near the Greenwich meridian, and the equator. The y-axis extends through the geocenter along a line perpendicular from the x-axis. It completes the right-handed system with the x- and z-axes and all three rotate with the earth. Three coordinate reference systems-NAD83, WGS 84, and ITRS-are all similarly defined.

ECEF (*See* Earth-Centered Earth-Fixed)

Elevation The distance measured along the direction of gravity above a surface of constant potential. Usually the reference surface is the geoid. Mean Sea Level (MSL) once utilized as a reference surface approximating the geoid is known to differ from the geoid up to a meter or more. The term height, sometimes considered synonymous with elevation, refers to the distance above an ellipsoid in geodesy.

Elevation Mask Angle (*See* **Mask Angle)**

Ellipsoid of Revolution (*Also known as* **Spheroid)** A biaxial closed surface, whose planar sections are either ellipses or circles that are formed by revolving an ellipse about its minor axis. Two quantities fully define an ellipsoid of revolution, the semimajor axis, a, and the flattening, $f = (a - b)/a$, where b is the length of the semiminor axis. The computations for the North American Datum of 1983 were done with respect to the GRS80 ellipsoid, the Geodetic Reference System of 1980. The GRS80 ellipsoid is defined by:

$$a = 6738137 \text{ meters}$$
$$1/f = 298.257222101$$

Ellipsoidal Height (*Also known as* **Geodetic Height)** The distance from an ellipsoid of reference to a point on the earth's surface, as measured along the perpendicular from the ellipsoid. Ellipsoidal height is not the same as elevation above mean sea level nor orthometric height. Ellipsoidal height also differs from geoidal height.

Ephemeris A table of values including locations and related data from which it is possible to derive a satellite's position and velocity. In GPS broadcast ephemerides are compiled by the Master Control Station, uploaded to the satellites by the Control Segment, and transmitted to receivers in the Navigation Message. It is designed to provide orbital elements quickly and is not as accurate as the precise ephemeris. Broadcast ephemeris errors are mitigated by differential correction or in double-differenced observables over short baselines. Precise ephemerides are postprocessed tables available to users via the Internet.

Epoch A time interval. In GPS the period of each observation in seconds.

Error Budget A summary of the magnitudes and sources of statistical errors that can help in approximating the actual errors that will accrue when observations are made.

FAA Federal Aviation Administration

Fast-Switching Channel (*See* **Multiplexing Channel)**

Federal Radionavigation Plan A document mandated by Congress that is published every other year by the Department of Defense and the Department of Transportation. In an effort to reduce the functional overlap of federal initiatives in radionavigation, it summarizes plans for promotion, maintenance, and discontinuation of domestic and international systems.

Femtosecond One-millionth of a nanosecond (10^{15}) of a second.

FM Frequency Modulation

Fractional Instantaneous Phase A carrier beat phase measurement not including the integer cycle count. It is always between zero and one cycle.

Frequency (*See also* **Wavelength**) The number of cycles per unit of time. In GPS the frequency of the unmodulated carrier waves are L1 at 1575.42 MHz, and L2 at 1227.60 MHz.

Frequency Band Within the electromagnetic spectrum, a particular range of frequencies.

FRP Federal Radionavigation Plan

Gain The *gain, or gain pattern,* of a GPS antenna refers to its ability to collect more energy from above the mask angle, and less from below the mask angle.

General Theory of Relativity (*See also* **Special Theory of Relativity**) A physical theory published by A. Einstein in 1916. In this theory space and time are no longer viewed as separate but rather form a four-dimensional continuum called space-time that is curved in the neighborhood of massive objects. The theory of general relativity replaces the concept of absolute motion with that of relative motion between two systems or frames of reference and defines the changes that occur in length, mass, and time when a moving object or light passes through a gravitational field. General Relativity predicts that as gravity weakens the rate of clocks increase; they tick faster. On the other hand, Special Relativity predicts that moving clocks appear to tick more slowly than stationary clocks, since the rate of a moving clock seems to decrease as its velocity increases. Therefore, for GPS satellites, General Relativity predicts that the atomic clocks in orbit on GPS satellites tick faster than the atomic clocks on earth by about 45,900 ns/day. Special Relativity predicts that the velocity of atomic clocks moving at GPS orbital speeds tick slower by about 7200 ns/day than clocks on earth. The rates of the clocks in GPS satellites are reset before launch to compensate for these predicted effects.

Geodesy The science concerned with the size and shape of the earth. Geodesy also involves the determination of positions on the earth's surface and the description of variations of the planet's gravity field.

Geodetic Datum A model defined by an ellipsoid and the relationship between the ellipsoid and the surface of the earth, including a Cartesian coordinate system. In modern usage, eight constants are used to form the coordinate system used for geodetic control. Two constants are required to define the dimensions of the reference ellipsoid. Three constants are needed to specify the location of the origin of the coordinate system, and three more constants are needed specify the orientation of the coordinate system. In the past, a geodetic datum was defined by five quantities: the latitude and longitude of an initial point, the azimuth of a line from this point, and the two constants to define the reference ellipsoid.

Geodetic Survey A survey that takes into account the size and shape of the earth. A geodetic survey may be performed using terrestrial or satellite positioning techniques.

Geographic Information System (GIS) A computer system used to acquire, store, manipulate, analyze, and display spatial data. A GIS can also be used to conduct analysis and display the results of queries.

Geoid The equipotential surface of the earth's gravity field which approximates mean sea level more closely. The geoid surface is everywhere perpendicular to gravity. Several sources have contributed to models of the Geoid, including ocean gravity anomalies derived from satellite altimetry, satellite-derived potential models, and land surface gravity observations.

Geoidal Height The distance from the ellipsoid of reference to the geoid measured along a perpendicular to the ellipsoid of reference.

Geometrical Dilution of Precision (GDOP) (*See* **Dilution of Precision**)

Gigahertz (GHZ) One billion cycles per second.

Global Navigation Satellite System (GNSS) GNSS is a reference to a combination of proposed GPS, GLONASS, and other systems both space and ground-based.

Global Orbiting Navigation Satellite System (GLONASS) A satellite radionavigation system financed by the Soviet Commonwealth. It is comprised of 21 satellites and 3 active spares arranged in three orbital rings approximately 11,232 nautical miles above the earth, though the number of functioning satellites may vary because of funding. Frequencies are broadcast in the ranges of 1597 to 1617 MHz and 1240 to 1260 MHz. GLONASS positions reference the datum PZ90 rather than WGS84.

Global Positioning System (GPS) A radionavigation system for providing the location of GPS receivers with great accuracy. Receivers may be stationary on the surface of the earth or in vehicles: aircraft, ships, or in earth orbiting satellites. The Global Positioning System is comprised of three segments; the User Segment, the Space Segment, and the Control Segment.

GPSIC GPS Information Center, U.S. Coast Guard

GPS Receiver An apparatus that captures modulated GPS satellite signals to derive measurements of time, position, and velocity.

GPS Time GPS time is the time given by all GPS Monitoring Stations and satellite clocks. GPS time is regulated by Coordinated Universal Time (UTC). GPS time and Coordinated Universal Time were the same at midnight UT on January 6, 1980. Since then GPS time has not been adjusted as leap seconds were inserted into UTC approximately every 18 months. These leap seconds keep UTC approximately synchronized with the earth's rotation. GPS time has no leap seconds and is offset from UTC by an integer number of seconds; the number of seconds is in the Navigation Message and most GPS receivers apply the correction automatically. The exact difference is contained in two constants, within the Navigation Message, A0 and A1, providing both the time difference and rate of system time relative to UTC. Disregarding the leap second offset, GPS time is mandated to stay within one microsecond of UTC, but over several years it has actually remained within a few hundred nanoseconds.

GRS80 Geocentric Reference System 1980. An ellipsoid adopted by the International Association of Geodesy.

Major semiaxis (a) = 6378137 meters
Flattening (1/f) = 298.257222101

Ground Antennas The S-band antennas used to upload information to the GPS satellites, including broadcast ephemeris data and broadcast clock corrections.

Hand-Over Word Information used to transfer tracking from the C/A-code to the P code. This time synchronization information is in the Navigation Message.

Height, Ellipsoidal (*See* Ellipsoidal Height)

Height, Orthometric (*See* Orthometric Height)

Hertz One cycle per second.

Heuristic Approach Fancy words for trial and error.

Hydrogen Maser An atomic clock. A device that uses microwave amplification by stimulated emission of radiation is called a maser. Actually, the microwave designation is not entirely accurate since masers have been developed to operate at many wavelengths. In any case, a maser is an oscillator whose frequency is derived from atomic resonance. One of the most useful types of maser is based on transitions in hydrogen, which occur 1421 MHz. The hydrogen maser provides a very sharp, constant oscillating signal, and thus serves as a time standard for an atomic clock. The active hydrogen maser provides the best known frequency stability for a frequency generator commercially available. At a 1 hour averaging time the active maser exceeds the stability of the best known cesium oscillators by a factor of at least 100 and the hydrogen maser is also extremely environmentally rugged.

Independent Baselines (*Also known as* Nontrivial Baselines) These are baselines observed using GPS Relative Positioning techniques. When more than two receivers are observing at the same time both independent (nontrivial) and trivial baselines are generated. For example, where r is the number of receivers, every complete static session yields r-1 independent (nontrivial) baselines. If four receivers are used simultaneously, six baselines are created. However, three of those lines will fully define the position of each occupied station in relation to the others. Therefore, the user can consider three (r-1) of the six lines independent (nontrivial), but once the decision is made only those three baselines are included in the network.

Inmarsat International Maritime Satellite

Integrity A quality measure of GPS performance including a system to provide a warning when the system should not be used for navigation because of some inadequacy.

Interferometry (*See* Relative Positioning)

International Earth Rotation Service (IERS) IERS was created in 1988 by the International Union of Geodesy and Geophysics (IUGG) and the International Astronomical Union (IAU). It is an interdisciplinary service organization that includes astronomy, geodesy, and geophysics. IERS maintains the International Celestial Reference System and the International Terrestrial Reference System.

International GPS Service (IGS) The IGS is comprised of many civilian agencies that operate a worldwide GPS tracking network. IGS produces postmission ephemerides, tracking station coordinates, earth orientation parameters, satellite clock corrections, tropospheric and ionospheric models. It is an initiative of the International Association of Geodesy and other scientific organizations, established in 1994 and originally intended to serve precise surveys for monitoring crustal motion. The range of users has since expanded.

International Terrestrial Reference System (ITRS) A very precise, geocentric coordinate system. The ITRS is geocentric, including the oceans and the atmosphere. Its length unit is the meter. Its axes are consistent with the Bureau International de l'Heuer (BIH) System at 1984.0 within +/–3 milliarcseconds. The IERS Reference Pole (IRP) and Reference Meridian (IRM) are consistent with the corresponding directions in the BIH directions within +/- 0.005". The BIH reference pole was adjusted to the Conventional International Origin (CIO) in 1967 and kept stable independently until 1987. The uncertainty between IRP and CIO is +/–0.03". The ITRS is realized by estimates of the coordinates and velocities of a set of stations, some of which are Satellite Laser Ranging (SLR) stations, or Very Long Baseline Interferometry (VLBI) stations, but the vast majority are GPS tracking stations of the IGS network.

Ionosphere A layer of atmosphere extending from about 50 to 1000 kilometers above the earth's surface in which gas molecules are ionized by ultraviolet radiation from the sun. The apparent speed, polarization, and direction of GPS signals are affected by the density of free electrons in this non-homogeneous and dispersive band of atmosphere.

Ionospheric Delay (*Also known as* **Ionospheric Refraction**) The difference in the propagation time for a signal passing through the ionosphere compared with the propagation time for the same signal passing through a vacuum. The magnitude of the ionospheric delay changes with the time of day, the latitude, the season, solar activity, and the observing direction. For example, it is usually least at the zenith and increases as a satellite gets closer to the horizon. The ionospheric delay is frequency dependent and, unlike the troposphere, affects the L1 and L2 carriers differently. There are two categories of the ionospheric delay, phase and group. Group delay affects the codes, the modulations on the carriers; phase advance affects the carriers themselves. Group delay and phase advance have the same magnitude but opposite signs. Since the ionospheric delay is frequency dependent it can be nearly eliminated by combination of pseudorange or carrier phase observations on both the L1 and L2 carriers. Still, even with dual-frequency observation and relative positioning methods in place, over long baselines the residual ionospheric delay can remain a substantial bias for high precision GPS. Single frequency receivers cannot significantly mitigate the error at all and must depend on the ionospheric correction available in the Navigation Message to remove even 50% of the effect.

Isotropic Antenna An isotropic antenna is a hypothetical, lossless antenna that has equal capabilities in all directions.

Iteration Converging on a solution by repetitive operations.

IVHS Intelligent Vehicle Highway System

Joint Program Office (JPO) The office responsible for the management of the GPS system in the U.S. Department of Defense.

Kalman Filter In GPS a numerical data combiner used in determining an instantaneous position estimate from multiple statistical measurements on a time-varying signal in the presence of noise. The Kalman filter is a set of mathematical equations that provides an estimation technique based on least squares. This recursive solution to the discrete-data linear filtering problem was proposed by R.E. Kalman in 1960. Since then the Kalman filter has been applied to radionavigation in general and GPS in particular, among other methods of measurement.

Kinematic Positioning (*Also known as* **Stop & Go Positioning**) (*See also* **Real-Time Kinematic Positioning**) (1) A version of relative positioning in which one receiver is a stationary reference and at least one other roving receiver coordinates unknown positions with short occupation times while both track the same satellites and maintain constant lock. If lock is lost reinitialization is necessary to fix the integer ambiguity. (2) GPS applications in which receivers on vehicles are in continuous motion.

L-Band Radio frequencies from 390 MHz to 1550 MHz.

L1 A GPS signal with the C/A code, the P code, and the Navigation Message modulated onto a carrier with the frequency 1575.42 MHz.

L2 A GPS signal with the P code and the Navigation Message modulated onto a carrier with the frequency 1227.60 MHz.

Latency The time taken for a system to compute corrections and transmit them to users in real-time GPS.

Latitude An angular coordinate, the angle measured from an equatorial plane to a line. On Earth the angle measured northward from the equator is positive, southward is negative. The geodetic latitude of a point is the angle between the equatorial plane of the ellipsoid and a normal to the ellipsoid through the point. At that point the astronomic latitude differs from the geodetic latitude by the meridional component of the deflection of the vertical.

Longitude An angular coordinate. On Earth, the dihedral angle from the plane of reference, 0° meridian, to a plane through a point of concern, and both planes perpendicular to the plane of the equator. In 1884, the Greenwich Meridian was designated the initial meridian for longitudes. The geodetic longitude of a point is the angle between the plane of the geodetic meridian through the point and the plane of the 0° meridian. At that point the astronomic longitude differs from the geodetic longitude by the amount of the component in the prime vertical of the local deflection of the vertical divided by the cosine of the latitude.

Loop Closure A procedure by which the internal consistency of a GPS network is discovered. A series of baseline vector components from more than one GPS session, forming a loop or closed figure, is added together. The closure error is the ratio of the length of the line representing the combined errors of all the vector's components to the length of the perimeter of the figure.

Mask Angle An elevation angle below which satellites are not tracked. The technique is used to mitigate atmospheric, multipath, and attenuation errors. A usual mask angle is 15°.

Master Control Station A facility manned by the 2nd Space Operations Squadron at Schriever Air Force Base (formerly Falcon AFB) in Colorado Springs, Colorado. The Master Control Station is the central facility in a network of worldwide tracking and upload stations that comprise the GPS Control Segment.

MCS GPS Master Control Station

Megahertz (MHz) One million cycles per second.

Microsecond (µsec) One-millionth (10^6) of a second.

Millisecond (ms or msec) One-thousandth of a second.

Minimally Constrained (*See* **Network Adjustment)** A least squares adjustment of all observations in a network with only the constraints necessary to achieve a meaningful solution. For example, the adjustment of a GPS network with the coordinates of only one station fixed.

Monitor Stations Stations used in the GPS control segment to track satellite clock and orbital parameters. Data collected at monitor stations are linked to a master control station at which corrections are calculated and from which correction data is uploaded to the satellites as needed.

Multichannel Receiver (*Also known as* **Parallel Receiver)** A receiver with many independent channels. Each channel can be dedicated to tracking one satellite continuously.

Multipath (*Also known as* **Multipath Error)** The error that results when a portion of the GPS signal is reflected. When the signal reaches the receiver by two or more different paths, the reflected paths are longer and cause incorrect pseudoranges or carrier phase measurements and subsequent positioning errors. Multipath is mitigated with various preventative antenna designs and filtering algorithms.

Multiplexing Channel (*Also known as* **Fast-Switching, Fast-Sequencing, and Fast-Multiplexing)** A channel of a GPS receiver that tracks through a series of a satellite's signals, from one signal to the next in a rapid sequence.

Multiplexing Receiver A GPS receiver that tracks a satellite's signals sequentially and differs from a Multichannel Receiver in which individual channels are dedicated to each satellite signal.

Modem (Modulator/Demodulator) A device that converts digital signals to analog signals and analog signals to digital signals. Computers sharing data usually require a modem at each end of a phone line to perform the conversion.

NAD83 (North American Datum, 1983) The horizontal control datum for positioning in Canada, the United States, Mexico, and Central America based on a geocentric origin and the Geodetic Reference System 1980 (GRS80) ellipsoid. The values for GRS 80 adopted by the International Union of Geodesy and Geophysics in 1979 are a = 6378137 meters and reciprocal of flattening = 1 /f = 298.257222101. It was designed to be compatible with the Bureau International de l'Heuer (BIH) Terrestrial System BTS84. The ellipsoid is geocentric. The origin was defined by satellite laser ranging (SLR), orientation by astronomic observations, and scale by both SLR and very long baseline interferometry (VLBI). NAD83 was actually realized through Doppler observations using internationally accepted transformations from the Doppler reference frame to BTS84 and adjustment of some 250,000 points. VLBI stations were also included to provide an accurate connection to other reference frames. While NAD83 is similar to the World Geodetic System of 1984 (WGS 84), it is not the same as WGS 84. Defined and maintained by the U.S. Department of Defense, WGS84 is a global geodetic datum used by the GPS Control Segment. Access to NAD83 is through a national network of horizontal control monuments established mainly by conventional horizontal control methods, triangulation, trilateration, and astronomic azimuths. Some GPS baselines were used. The adjustment of these horizontal observations, together with several hundred observed Doppler positions, provides a practical realization of NAD83.

Nanosecond (ns or nsec) One-billionth, (10^9) of a second.

NANU Notice Advisory to NAVSTAR Users

Navigation Message (*Also known as* **Data Message)** A message modulated on L1 and L2 of the GPS signal that includes an ionospheric model, the satellite's broadcast ephemeris, broadcast clock correction, constellation almanac, and health, among other things.

NAVSAT A European radionavigation system under development.

NAVSTAR (NAVigation Satellite Timing and Ranging) The GPS satellite system.

Network Adjustment A least squares solution in which baselines vectors are treated as observations. It may be minimally constrained. It may be constrained by more than one known coordinates, as is usual in a GPS survey to densify previously established control or a geodetic framework.

North American Vertical Datum of 1988 (NAVD88) A minimally constrained adjustment of U.S., Canadian, and Mexican leveling observations holding fixed the height of the primary tidal benchmark of the new International Great Lakes Datum of 1985 (IGLD85) at Father Point/Rimouski, Quebec, Canada. NAVD88 and IGLD85 are now the same. Between NAVD88 orthometric heights and those referred to the National Geodetic Vertical Datum of 1929 (NGVD29) there are differences ranging from –40 cm to +150 cm within the lower 48 states. The differences range from +94 cm to +240 cm in Alaska. GPS derived orthometric heights estimated using the precise geoid models now available are compatible with NAVD88. NAVD88 includes 81,500 km of new leveling data never before adjusted to NGVD29. The principal impetus for NAVD88 was minimizing the recompilation of national mapping products. The NAVD88 datum does not correspond exactly to the theoretical level surface defined by the GRS80 definitions.

Observing Session (*Also known as* **Observation**) Continuous and simultaneous collection of GPS data by two or more receivers.

Omega A radionavigation system that can provide global coverage with only eight ground-based transmitting stations.

On-the-Fly (OTF) A method of resolving the carrier phase ambiguity very quickly. The method requires a dual-frequency GPS receivers capable of making both carrier phase and precise pseudorange measurements. The receiver is not required to remain stationary, making the technique useful for initializing in carrier phase kinematic GPS.

Orthometric Height The vertical distance from the geoid to the surface of the earth. GPS heights are ellipsoidal. Ellipsoidal heights are the vertical distance from an ellipsoid of reference to the earth's surface. Ellipsoidal heights differ from leveled, orthometric heights. And conversion from ellipsoidal to orthometric heights requires the vertical distance from the ellipsoid of reference to the geoid, the geoidal height. The vertical distance from the ellipsoid of reference to the geoid around the world varies from +75 to –100 meters; in the coterminous United States it varies from –8 meters to –53 meters. The geoidal heights are negative because the geoid is beneath the ellipsoid. In other words, the ellipsoid is overhead. The relationship between these three heights is:

$$h = H + N$$

where h is the ellipsoid height, H is the orthometric height and N is the geoidal height.

Outage GPS positioning service is unavailable. Possible reasons for an outage include sufficient number of satellites are not visible, the dilution of precision value is too large, or the signal to noise ratio value is too small.

P Code (*Also known as* **Protected Code**) A binary code known by the names P Code, Precise Code, and Protected Code. It is a standard spread spectrum GPS pseudorandom noise code. It is modulated on the L1 and the L2 carrier using binary biphase modulations. Each week long segment of the P Code is unique to a single GPS satellite and is repeated each week. It has a chipping rate 10.23 MHZ, more than ten million bits per second. The P Code is sometimes replaced with the more secure Y-Code in a process known as antispoofing.

Parallel Receiver (*See* **Multichannel Receiver**)

PDOP-Position Dilution of Precision (*See* **Dilution of Precision**)

Perturbation Deviation in the path of an object in orbit. Deviation being departure of the actual orbit from the predicted Keplerian orbit. Perturbing forces of earth orbiting satellites are caused by atmospheric drag, radiation pressure from the sun, the gravity of the moon and the sun, the geomagnetic field, and the noncentral aspect of earth's gravity.

Phase Lock The adjustment of the phase of an oscillator signal to match the phase of a reference signal. First, the receiver compares the phases of the two signals. Next, using the resulting phase difference signal the reference oscillator frequency is adjusted. When the two signals are next compared the phase difference between them is eliminated.

Phase Observable (*See* **Carrier Phase)**

Phase Smoothed Pseudorange A pseudorange measurement with its random errors reduced by combination with carrier phase information.

Picosecond One-trillionth (10^{12}) of a second, one-millionth of a microsecond.

Point Positioning (*See* **Absolute Positioning)**

Position A description, frequently by coordinates, of the location and orientation of a point or object.

Postprocessed GPS A method of deriving positions from GPS observations in which base and roving receivers do not communicate as they do in Real-Time Kinematic (RTK) GPS. Each receiver records the satellite observations independently. Their collections are combined later. The method can be applied to pseudoranges to be differential corrected or carrier phase measurements to be processed by double-differencing.

Precise Code (*See* **P Code)**

Precise Ephemeris (*See* **Ephemeris)**

Precise Positioning Service (PPS) GPS positioning for the military at a higher level of absolute positioning accuracy than is available to C/A code receivers, which relies on SPS (Standard Positioning Service). PPS is based on the dual-frequency P code.

Precision Agreement among measurements of the same quantity; widely scattered results are less precise than those that are closely grouped. The higher the precision, the smaller the random errors in a series of measurements. The precision of a GPS survey depends on the network design, surveying methods, processing procedures, and equipment.

Protected Code (*See* **P Code)**

Pseudolite (*Also known as* **Pseudo Satellite)** A ground-based differential station which simulates the signal of a GPS satellite with a typical maximum range of 50 km. Pseudolites can enhance the accuracy and extend the coverage of the GPS constellation. Pseudolite signals are designed to minimize their interference with the GPS signal.

Pseudorandom Noise (*Also known as* **PRN)** A sequence of digital 1s and 0s that appear to be randomly distributed, but can be reproduced exactly. Binary signals with noise-like properties are modulated on the GPS carrier waves as the C/A codes and the P codes. Each GPS satellite has unique C/A and P codes. A satellite may be identified according to its PRN number. Thirty-two GPS satellite pseudorandom noise codes are currently defined.

Pseudorange In GPS a time biased distance measurement. It is based on code transmitted by a GPS satellite, collected by a GPS receiver and then correlated with a replica of the same code generated in the receiver. However, there is no account for errors in synchronization between the satellite's clock and the receiver's clock in a pseudorange. The precision of the measurement is a function of the resolution of the code, therefore a C/A code pseudorange is less precise than a P code pseudorange.

Quartz Crystal Controlled Oscillator GPS receivers rely on a quartz crystal oscillator to provide a stable reference so that other frequencies of the system can be compared with or generated from this reference. The fundamental component is the quartz crystal resonator. It utilizes the piezoelectric effect. When an electrical signal is applied the quartz resonates at a frequency unique to its shape, size, and cut. The first study of the use of quartz crystal resonators to control the frequency of vacuum tube oscillators was made by Walter G. Cady in 1921. Important contributions were made by G. W. Pierce, who showed that plates of quartz cut in a certain way could be made to vibrate so as to control frequencies proportional to their thickness.

R95 A representation of positional accuracy. The radius of a theoretical circle centered at the true position that would enclose 95% of the other positions.

Radionavigation The determination of position, direction, and distance using the properties of transmitted radio waves.

Range (*Also known as* **Geometric Range**) A distance between two points, particularly the distance between a GPS receiver and satellite.

Range Rate The rate at which the range between a GPS receiver and satellite changes. Usually measured by tracking the variation in the Doppler shift.

Real-Time DGPS A method of determining relative positions between known control and unknown positions using pseudorange measurements. A base station or satellite at the known position transmits corrections to the roving receiver or receivers. The procedure offers less accuracy than RTK. However, the results are immediate available, in real-time and need not be postprocessed.

Real-Time Kinematic (RTK) Positioning A method of determining relative positions between known control and unknown positions using carrier phase measurements. A base station at the known position transmits corrections to the roving receiver or receivers. The procedure offers high accuracy immediately, in real-time. The results need not be postprocessed. In the earliest use of GPS kinematic and rapid static positioning were not frequently used because ambiguity resolution methods were still inefficient. Later when ambiguity resolution such as on-the-fly (OTF) became available, real-time kinematic and similar surveying methods became more widely used.

Receiver Channel (*See* **Channel**)

Reconstructed Carrier Phase (*See* **Carrier Phase**)

Reference Station (*See* **Base Station**)

Relative Positioning A GPS surveying method that improves the precision of carrier phase measurements. One or more GPS receivers occupy a base station, a known position. They collect the signals from the same satellites at the same time as other GPS receivers that may be stationary or moving. The other receivers occupy unknown positions in the same geographic area. Occupying known positions, the base station receivers find corrective factors that can either be communicated in real-time to the other receivers, as in RTK, or may be applied in postprocessing, as in static positioning. Relative positioning is in contrast to Absolute, or Point Positioning. In relative positioning errors that are common to both receivers, such as satellite clock biases, ephemeris errors, propagation delays, and so forth are mitigated.

RINEX (Receiver Independent Exchange Format) A package of GPS data formats and definitions that allow interchangeable use of data from dissimilar receiver models and postprocessing software developed by the Astronomical Institute of the University of Berne in 1989. More than 60 receivers from 4 different manufacturers were used in the GPS survey EUREF 89. RINEX was developed for the exchange of the GPS data collected in that project.

Root Mean Square (RMS) The square root of the mean of squared errors for a sample.

Rover (*Also known as* **Mobile Receiver)** A GPS receiver that is in motion relative to a stationary base station during a session.

RTCM (Radio Technical Commission on Maritime Services) In DGPS the abbreviation RTCM has come to mean the correction messages transmitted by some reference stations using a protocol developed by the Radio Technical Commission on Maritime services Special Committee 104. These corrections can be collected and decoded by DGPS receivers designed to accept the signal. The corrections allow the receiver to generate corrected coordinates in real-time. There are several sources of RTCM broadcasts. One source is the U.S. Coast Guard system of beacons in coastal areas. Other sources are commercial services, some of which broadcast RTCM corrections by satellite, some use FM subcarriers.

Rubidium Clock An environmentally tolerant and very accurate atomic clock whose working element is gaseous rubidium. The resonant transition frequency of the Rb-87 atom (6,834,682,614 Hz) is used as a reference. Rubidium frequency standards are small, light, and have low power consumption.

SA (*See* **Selective Availability)**

S-Code (*See* **C/A-Code)**

Satellite Clocks Two rubidium (Rb) and two cesium (Cs) atomic clocks are aboard Block II/IIA satellites. Three rubidium clocks are on Block IIR satellites. Hydrogen Maser time standards may be used in future satellites.

Satellite Constellation In GPS, 4 satellites in each of six orbital planes. In GLONASS 8 satellites in each of three orbital planes.

Second Base unit of time in the International System of Units. The duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom undisturbed by external fields.

Selective Availability (SA) By dithering the timing and ephermides data in the satellites, standard positioning service (SPS) users are denied access to the full GPS accuracy. The intentional degradation of the signals transmitted from the satellites may be removed by precise positioning service (PPS) users. It is also virtually eliminated in carrier phase relative positioning techniques and double-differencing processing. Selective availability was begun in 1990 and was removed from August 10, 1990 to July 1, 1991. It was enabled on November 15, 1991 and finally switched off by presidential order on May 2, 2000.

Single Difference (*See* **Between-Receiver Difference**) (*See also* **Between-Satellite Single Difference**)

Space Segment The portion of the GPS system in space, including the satellites and their signals.

Spatial Data (*Also known as* **Geospatial Data**) Information that identifies the geographic location and characteristics of natural or constructed features and boundaries of earth. This information may be derived from, among other things, remote sensing, mapping, and surveying technologies.

Special Theory of Relativity (*See also* **General Theory of Relativity**) A theory developed by Albert Einstein, predicting, among other things, the changes that occur in length, mass, and time at speeds approaching the speed of light. General relativity predicts that as gravity weakens the rate of clocks increase and they tick faster. On the other hand, special relativity predicts that moving clocks appear to tick more slowly than stationary clocks, since the rate of a moving clock seems to decrease as its velocity increases. Therefore, for GPS satellites, general relativity predicts that the atomic clocks in orbit on GPS satellites tick faster than the atomic clocks on earth by about 45,900 ns/day. Special relativity predicts that the velocity of atomic clocks moving at GPS orbital speeds tick slower by about 7200 ns/day than clocks on earth. The rates of the clocks in GPS satellites are reset before launch to compensate for these predicted effects.

Spherical Error Probable (SEP) A description of three-dimensional precision. The radius of a sphere with its center at the actual position that is expected to be large enough to include half (50%) of the normal distribution of the scatter of points observed for the position.

Spread Spectrum Signal A signal spread over a frequency band wider than needed to carry its information. In GPS a spread spectrum signal is used to prevent jamming, mitigate multipath and allow unambiguous satellite tracking.

SPS (*See* **Standard Positioning Service**)

Standard Deviation (*Also known as* **1 sigma, 1σ**) An indication of the dispersion of random errors in a series of measurements of the same quantity. The more tightly grouped the measurements around their average (mean), the smaller the standard deviation. Approximately 68% of the individual measurements will be within the range expressed by the standard deviation.

Standard Positioning Service (SPS) Civilian absolute positioning accuracy using pseudorange measurements from a single frequency C/A code receiver, ±20 to ±40 meters 95% of the time (2 drms) with Selective availability switched off.

Static Positioning A relative, differential, surveying method in which at least two stationary GPS receivers collect signals simultaneously from the same constellation of satellites during long observation sessions. Generally, static GPS measurements are postprocessed and the relative position of the two units can be determined to a very high accuracy. Static positioning is in contrast to kinematic and real-time kinematic positioning where one or more receivers track satellites while in motion. Static positioning is in contrast to absolute, or point, positioning, which has no relative positioning component. Static positioning is in contrast to rapid-static positioning where the observation sessions are short.

Stop-and-Go Positioning (*See* **Kinematic Positioning**)

SV Space Vehicle

Time Dilation (*Also known as* **Relativistic Time Dilation**) (*See* **Special Theory of Relativity**) (*See* **General Theory of Relativity**) Systematic variation in time's rate on an orbiting GPS satellite relative to time's rate on earth. The variation is predicted by the special theory of relativity and the general theory of relativity as presented by A. Einstein.

Transit Navigation System A satellite based Doppler positioning system funded by the U.S. Navy.

Triple-Difference (*Also known as* **Receiver-Satellite-Time Triple Difference**) The combination of two double differences. Each of the double differences involves 2 satellites and 2 receivers. The triple difference is derived between 2 epochs. In other words, a triple difference involves 2 satellites, 2 receivers, and time. Triple differences ease the detection of cycle slips.

Trivial Baseline Baselines observed using GPS Relative Positioning techniques. When more than two receivers are observing at the same time both independent (nontrivial) and trivial baselines are generated. For example, where r is the number of receivers, every complete static session yields r-1 independent (nontrivial) baselines and the remaining baselines are trivial. For example, if four receivers are used simultaneously, six baselines are created. Three are trivial and three are independent. The three independent baselines will fully define the position of each occupied station in relation to the others. The three trivial baselines may be processed, but the observational data used for them has already produced the independent baselines. Therefore, only the independent baseline results should be used for the network adjustment or quality control.

Tropospheric Effect (*Also known as* **Tropospheric Delay**) The troposphere comprises approximately 9 km of the atmosphere over the poles and 16 km over the equator. The tropospheric effect is non-dispersive for frequencies under 30 GHz. Therefore, it affects both L1 and L2 equally. Refraction in the troposphere has a dry component and a wet component. The dry component is related to the atmospheric pressure and contributes about 90% of the effect. It is more easily modeled than the wet component. The GPS signal that travels the shortest path through the atmosphere will be the least affected by it. Therefore, the tropospheric delay is least at the zenith and most near the horizon. GPS receivers at the ends of short baselines collect signals that pass through substantially the same atmosphere and the tropospheric delay may not be troublesome. However, the atmosphere may be very different at the ends of long baselines.

UKOOA Member companies of the United Kingdom Offshore Operators Association (UKOOA) are licensed by the government to search for and produce oil and gas in U.K. waters. It publishes, among other things "The Use of Differential GPS in Offshore Surveying." These guidelines cover installation and operation, quality measures, minimum training standards, receiver outputs, and data exchange format.

Universal Transverse Mercator (UTM) The transverse Mercator projection represents ellipsoidal positions, latitude and longitude, as grid coordinates, northing and easting, on a cylindrical surface that can be developed into a flat surface. Universal Transverse Mercator is a particular type of transverse Mercator projection. It was adopted by the U.S. Army for large-scale military maps and is shown on all 7.5 minute quadrangle maps and 15 minute quadrangle maps prepared by the U.S. Geological Survey. The earth is divided into 60 zones between 84° N latitude and 80° S latitude, most of which are 6° of longitude wide. Each of these UTM zones have a unique central meridian and the scale varies by 1 part in 1000 from true scale at equator.

USCG U.S. Coast Guard

User-Equivalent Range Error (UERE) The contribution in range units of individual uncorrelated biases to the range measurement error.

User Interface The software and hardware that activate displays and controls that are the means of communication between a GPS receiver and the receiver's operator.

User Segment That component of the GPS system that includes the user equipment, applications, and operational procedures. The part of the whole GPS system that includes the receivers of GPS signals.

UTC (*See* Coordinated Universal Time)

UTM (*See* Universal Transverse Mercator)

Very Long Baseline Interferometry (VLBI) By measuring the arrival time of the wave front emitted by a distant quasar at two widely separated earth-based antennas, the relative position of the antennas is determined. Because the time difference measurements are precise to a few picoseconds, the relative positions of the antennas are accurate within a few millimeters and the quasar positions to fractions of a milliarcsecond.

Voltage-Controlled Quartz Crystal Oscillator (*See also* Quartz Crystal Controlled Oscillator) A quartz crystal oscillator with a voltage controlled frequency.

Wavelength (*See also* Frequency) Along a sine wave the distance between adjacent points of equal phase. The distance required for one complete cycle.

Waypoint A two-dimensional coordinate to be reached by GPS navigation.

Wide Area Augmentation System (WAAS) A U.S. Federal Aviation Authority (FAA) system that augments GPS accuracy, availability, and integrity. It provides a satellite signal for users to support en route and precision approach aircraft navigation. Similar systems are Europe's European Geostationary Navigation Overlay System and Japan's MT-SAT.

World Geodetic System 1984 (WGS84) A world geodetic earth-centered, earth-fixed terrestrial reference system. The origin of the WGS 84 reference frame is the center of mass of the earth. The GPS Control Segment has worked in WGS84 since January 1987; and therefore, GPS positions are said to be in this datum.

Y-Code When Antispoofing is on the P code is encrypted into the Y code and transmitted on L1 and L2.

Z-Count Word The GPS satellite clock time at the leading edge of the next data subframe of the transmitted GPS message (usually expressed as an integer number of 1.5 second periods) [van Dierendock et al., 1978].

Zero Baseline Test A setup using two GPS receivers connected to one antenna. Nearly all biases are identical for both receivers and only random observation errors attributable to both receivers remain. The baseline measured should be zero if the receivers' calibration was ideal.

Zulu Time (*See* **Coordinated Universal Time)**

Index

Page references in *italics* refer to figures.

Page references in **bold** refer to tables.

A

absolute positioning, 47
access, site, 187–188, 236
accuracy
  clock corrections, 41
  ephemerides, 41
  FGCC standards, 177–178, 181, 194, 199, 235, 241, 245
  GLONASS, 284
  GNSS, 291–292
  GPS, **70**, 290–291
  kinematic positioning, 110–111
  navigation message, 3–4
  receivers, 72–73
  SPS, 72
Active Stations; *See* CORS (continuously operating reference stations)
Adams, Oscar, 146–148
adaptive gain, 82
Adjusted, 179
AEP (Architecture Evolution Plan), 41, 81
AFN (Australian Fiducial Network), 83
almanac, 7, 18
  carrier phase measurement, 97
  downloading, 290
AM frequency, 15
ambiguity; *See* cycle ambiguity
amplitude modulation, 2
Antenna Reference Point (ARP), 183–184
  bandwidth, 93
  centering, 94, 96, 231, 242–243
  coverage of, 93–94
  dish, 91
  efficiency, 92
  electronic center, 94
  gain pattern, 93
  ground, 81
  height of, 94, 213, 231, 245
  minimizing multipath, 45, *46*, 93, 240
  mounting, 231
  orientation, 94
  phase center, 94, 96, 243
  physical center, 94
  radio receiving, 211
  shared, 231
  tracking facilities, 41
  types of, 93
antenna splitter, 231
antispoofing (AS), 7, 80, 105
  Block II satellites, 267
apogee, 165
Army POS/NAV Master Plan, 71
ARNS (Aeronautical Radionavigation Service), 274, 275
ARP; *See* Antenna Reference Point
astronomic latitude, 125
astronomical coordinates, 65
atmosphere
  and antenna coverage, 93–94
  effect on signals, 2
  effects on SLR, 66
  ionospheric effects, 37–40
  modeling, 43
  tropospheric effects, 42–46
AUSLIG (Australian Surveying and Land Information Group), 83
AUSPOS (Australian Online GPS Processing Service), 257
Australian Geodetic Datum of 1966, 130, *131*
Authorized Service, 288
Auto Gipsy, 257
autocorrelation, 18–19
AutoNav, 268
autonomous positioning, 47
auxiliary equipment, 232
azimuth, 93, 94, 181
  deflection of the vertical, 125
  measuring, 237
  planning marks, 188
azimuth and elevation, 190, **191**, **193**, **239**

B

back-dating, 206
Bailey, Ann, 131
ballistic cameras, 65
bandpass filter, 95, 97
bandwidth
  antennas, 93
  SNR (signal-to-noise ratio), 112
base station, 202–203
  coordinates, 213
  corrections, 206

    DGPS, 214
    RTK, 214
    setting up, 213
baselines, 49
    drawing, 194–202, *196*
    integer fixing, 255
    zero baseline test, 231
batteries, receiver, 102–103, 232
beat frequency, 95, 97, 98–99
Beidou, 288–289, 292–293
Bell, Alexander Graham, 94
benchmark elevations, 167, 185
Bester, O.B., 145
between-receivers difference, 49–51, *50*, *52*
between-satellites difference, 51, *53*, 57, 252–253, 255
biases
    correlation between, 205, 250
    distance-dependent, 216–217
    ionospheric effects, 37–40, 216–217, 278
    orbital, 40–42
    receiver clock, 40, 51, *53*
    receiver noise, 45–46
    satellite clock, 35–37, 49, 51, 202–203, 217–218
    sources of, 249
    tropospheric effects, 42–46, 216–217
    UERE, 35, *36*
    unavoidable, 35, 57
biaxial ellipsoidal model, Earth, 127–130, *129*
BIH (Bureau International de l'Heure), 123, 136
Block I satellites, 78–79, 80, 267–268
Block II satellites, 79, 80, *80*, 82, 267–268
Block IIA satellites, 82, 267
Block IIF satellites, 5, 79–80, 276
Block IIR-M satellites, 269, 271–274
Block IIR satellites, 4, 5, 79, 267–268
blue-booking, 138, 199
BOC (Binary Offset Carrier) modulation, 271, 279, 287
BPSK (binary phase shift keying), 16–17, 271, 273, 287
brass caps, 135
broadcast clock correction, 5, 35–37, 40, 267
broadcast ephemeris, 267

C

C&GS (Coast and Geodetic Survey), 135
C/A (coarse/acquisition) code, 3–4, 272
    antenna bandwidth, 93
    bandpass filter, 95
    chipping rate, 17, 268
    correlating, 96
    generation rate, 8
    maximum precision, 23
    multichannel receivers, 105
    PRN, 8
    replica, 18, 95
    selective availability of, 80
    signal power, 268
    spread spectrum, 15–16
    SPS, 8–9, 72
    squaring, 100
cameras, ballistic, 65
Canadian Coast Guard, 206
carrier beat phase, 15, 25–26, *25*
carrier phase
    approximation, 26–29
    components, 55
    cycle ambiguity, 29–30
    cycle slip, 253–255
    Doppler effect, 26
    multipath in, 44–45
    OTF ambiguity resolution, 209–210
    phase difference, 24–25
    receivers, 231
    SA, 24
    signal squaring, 99–100
carrier phase measurement, 97
    wide laning, 113–114
carrier phase ranging, 12–14
carrier phase receivers, kinematic positioning, 111
carrier tracking loop, 96, 97
carrier wave
    frequency, 3
    modulation, 2, 9–15
    phase delay, 39
    replica, 95, 97
Cartesian coordinates, 3-D, 122–124, *123*, *125*, 131
    site calibration, 248–249
Cassini, Jean Dominique, 127–128
C-band, 63
CBN (Cooperative Base Network), 140, 179
CDMA (Code Division Multiple Access), 282–283
CDU (control and display unit), 100, 101–102
cell phones, RTK link, 211
centering, antenna, 231, 242–243
Central Meridian, *161*
cesium clocks, 5, 79, 267, 276
Chandler period, 123
channels, RF section, 95
chi–square test, 256–257
chip-by-chip time multiplexing, 273
choke ring antenna, 45, 240
CIGNET (Cooperative International GPS Network), 83
CIO (Conventional International Origin), 123
CL (Civil-Long) code, 272
Claire, Charles, 146
Clarke 1866 spheroid, 128, 130, 135, 154

clock correction, 5, 35–37, 40, 267
clock correlation, Navigation message, 4–5
clock errors, eliminating, 49, 51
clocks; *See also* oscillators; receiver clocks; satellite clocks; time measurement
  atomic, 21, 22
  breakdown, 7
  drift, 36–37
  GPS oscillators, 11–12
  offsets, 21, 22
  SA, 9
CM (Civil-Moderate) code, 272
CNAV, 272–273
CNAV-2, 279
CNSS (Compass Navigation Satellite System), 288
code chips, 16–17, 18
code correlation
  receivers, 15, 17–21
  and time shift, 18–19, *19*, 20–21, *20*
code-phase measurement, 96, 99
code states, 18, 100
code tracking loop, 96, 97
codeless receivers, 104, 105
codes, 3–4
coherent modulated carrier waves, 12
combined factor, 157
Commercial Service, 285
communications
  equipment, 230–231
  satellite, 91
COMPASS, 288–289
Computed, 179
cone projections, 141, *142*, 144, *146*, *149*
conformality, 147, 148
constellation
  DGPS, 207
  GALILEO, 292–293, *293*
  GLONASS, 281–282, 292–293, *293*
  GPS, 292–293, *293*
  RTK, 212
constellation, GPS, 292–293, *293*
  dilution of precision, 74–78
  orbital period, 73–74
  satellite names, 78–81
continuous phase observable, 95
control points, 2
Control Segment, 81–83
  broadcast clock corrections, 37
  constant tracking, 82
  ionospheric delay, 7
  Kalman filtering, 81, 82
  Master Control Station, 41, 81, 82
  monitoring stations, 81
  postcomputed ephemerides, 82–83
  satellite tracking, 40–42, *42*
control stations, 2
  access, 187–188
  age and location, 182, 185, *186*
  CORS, 251
  horizontal, 179, *180*, 181, 184–185, *186*, *187*, *195*
  marks, 182, 185
  PID, 179, *180*
  as project point, 186–187, *186*, *187*, *195*
  redundancy, 199–202
  spacing, 139
  vertical, 179, *180*, 181, 185–186, *187*, *195*
CONUS (coterminous United States), 168
coordinates
  astronomic, 65
  3-D Cartesian, 122–124, *123*, *125*, 131, 248–249
  base station, 213
  ECEF, 122, 123, 131, 181
  plane, 140–141, 181
  polar, *77*
  site calibration and, 248–249
  UTM, 140–141, 158, 181
correction signal, 204
    real-time, 215
correlation channel, 21
correlation peak, 19–21, *20*
CORS (Continuously Operating Reference Stations), 140, 179, 251
    NGS, 183–184
    network modeling, 216–217
    and RTN, 216–217
    static GPS, 182–184
Costas loop, 273
cross-correlation, 273
CSOC (Consolidated Space Operations Center), 81
CSRS-PPP (CSRS-Precise Point Positioning), 218
CTP (Conventional Terrestrial Pole), 122
CTRS (Conventional Terrestrial Reference System), 122
CTS (Conventional Terrestrial System), 67, 122, 130
  geocenter, 122
  in GPS, 123
cycle ambiguity, 14–15, 51, 55–56
  buggy wheel method, 29–30
  fixed solution, 57, 253
  float solution, 57, 253
  methods for solving, 99
  modulated carrier waves, 14–15
  OTF resolution of, 114, 209
  in RTK, 208–210, 212
  static GPS, 108
cycle slip, 55–56, 245
  causes, 254
  detection, 253–254

  repairing, 254–255
cylinder projections, 141, *142*, 144, *146*, *149*

D

data
  downloading, 102
  file naming, 250
  noisy, 249
  organizing, 250
  postprocessing; *See* postprocessing
  rate, 211
  simultaneity of, 242
  storage, 251
data collection, DGPS for, 207–208
data interval, 109
data processing; *See* postprocessing
data radio transmitter, 211
datums, 121–122, 127
  coordinate systems, 122–124, 131
  elements of, 124–125
  ellipsoid as, 127–130, *129*
  global geocentric system, 133–134
  NAD27; *See* NAD27
  NAD83; *See* NAD83
  NAVD88, 167
  NGVD29, 166
  SLD29, 164, 166
  regional, 130–131
  WGS84, 6, 122, 134, 136–137
decibel, 94
declination calculator, 246
deflection of the vertical, 124–125, *126*
demodulator, 212
dependent lines, *196*, 197–198, *200*
designation, 179
deterioration
  signal, 80
  SNR, 100
  zero reference point, 167
DGPS (differential GPS), 47, 91, 100–101, 110
  applications, 207–208
  base station, 214
  correction signals, 215
  data collection, 207–208
  dynamic lines, 247
  error sources, 202–203
  initialization, 214
  LADGPS, 205
  maritime, 206
  multipath in, 213–214, 247–248
  point offsets, 246–247, *246*
  radial GPS in, 203–204, *203*
  real-time, 110, 204–205, *205*, 215
  WAAS, 206, *207*
  WADGPS, 205
differencing, 49, 56, 249
  double difference, 51, *53*, 57, 252–253, 255
  single difference, 49–51, *50*, *52*
  triple difference, 51, *54*, 55–56, 252, 254–255
dipole antennas, 93
directions, to-reach description, **233**, 235–236
dish antennas, 91
distance measurement, 1–2, 11–12
distortion, map projections, 143–144, *145*
diurnal tide, 165–167
DLL (delay lock loops), 21, 96
DOP (dilution of precision), 74–78
Doppler, Christian, 26
Doppler shift, 18, 97–100
  carrier phase, 26
  Doppler count, 98–99
  GPS, 26
  satellite technology and, 68
DORIS (Doppler Orbitography and Radiopositioning Integrated by Satellite), 135
double difference, 51, *53*, 57, 252–253, 255
double occupation, 112
dual-frequency receivers, 39–40, 95, 105, 211, 213
DX, DY, and DZ, 123, *124*
dynamic lines, 247

E

E5a, GALILEO, 285–288, *286*, 288
E5b, GALILEO, 283, 285–288, *286*
early (E) correlators, 19–21
Earth
  ellipsoidal model, 125, 127–130, *129*
  geocentric model, 131–132
  geoidal surface, 132–133, *133*
eastings, 160
ECEF (Earth-Centered-Earth-Fixed) coordinates, 122, 123, 131, 181
EDM (electronic distance measuring), 1, 212
  atmospheric effects, 2
  phase shift measurements, 12, *13*, 14, *14*
  two-way ranging, 9, *10*, 14–15
EGNOS (European Geostationary Navigation Overlay Service), 206
Einstein, Albert, 36
elevation factor, 154–157
elevations, 93, 94, 130
  benchmark, 167, 185
  determining, 162
ellipsoid factor, 154–157
ellipsoid-geoid-mean sea level, *168*
ellipsoid height, 156, 162–163, *162*, 212
ellipsoidal model, Earth, 125, 127–130, *129*
ellipsoids, map projections, 141–143
EOP (Earth Orientation Parameters), 136

ephemerides
  accuracy, 41
  broadcast, 6
  codeless receivers, 104, 105
  postcomputed, 82–83, 217
  truncated, 7
epoch, 48, 247
equipment
  auxiliary, 232
  communications, 230–231
  conventional, 229–230
  GPS, 231–232
  papers, 232
  safety, 230
equipotential surfaces, 133, 167, *168*, *170*
Erastosthenes, 127
error budget, 35
errors; *See* biases
European Commission, 285
European Datum 1950, 64, 130, *131*
European Space Agency (ESA), 285
exosphere, 37
extraterrestrial radio positioning, 66, *67*

F

FAA (Federal Aviation Administration), 206
false easting, 150
false northing, 150
fast-static positioning, **107**, 113–114
fast-switching receiver, 96
FBN (Federal Base Network), 140, 179
FCC (Federal Communications Commission), 210
FDMA (Frequency Division Multiple Access), 283–284, *283*
FEC (Forward Error Correction), 272–273
Federal Radionavigation Plan of 1984, 72
FGCC (Federal Geodetic Control Committee), 139
  accuracy standards, 194, 235, 241, 245
  positioning standards, 177–178, 181
  redundancy standards, 199–202
field notes, 232
file naming, 250
FIPS (Federal Information Processing Standard), 149
FIPS zones, 149–150
fixed solution, 57, 253
flagging, 229, 236
flattening ratio, 127, 134, 141
float solution, 57, 253
FM radios, 15, 230
FOC (full operational capability), 276, 278
4 minute difference, 74
fractional initial phase, 55
frequency
  carrier waves, 3
  harmonic, 95
frequency shift
  and code correlation, 18
  Doppler effect, 26
fundamental clock rate, 16–17

G

gain pattern, antennas, 93
GALILEO, 272, 278
  constellations, 292–293, *293*
  data processing, 210
  GIOVE-A, 284–285
  GIOVE-B, 284–285
  interoperability, 286–288, *286*, 292–293
  PRS code, 287, 288
  signals and services, 285–288, *286*
Galileo Galilei, 285
"garbage in, garbage out," 57
GBAS (ground-based augmentation system), 280
GDOP (geometric dilution of precision), 75
geocentric datum, 133–134
geocentric model, Earth, 131–132
geodesy, and surveyors, 121–122
geodetic distances, 150–157, *152*, *155*
geodetic height, 162
geodetic latitude, 125
geodetic longitude, 125
Geodetic Survey of Canada, 135
geoid, 132–133, *133*, 167–169, 181
    height, 130, 162, 168–169, 212
    models, 169–170
GEOID90, 169
GEOID93, 169
GEOID99, 169–170
geometric range, 23
GEONET, 182
GIS (Geographical Information Systems), 100
GJU (Galileo Joint Undertaking), 285
GLONASS, 210, 215, 272
  L bands, 283–284
  accuracy of, 284
  CDMA, 282–283
  constellations, 281–282, 292–293, *293*
  FDMA, 283–284, *283*
  interoperability, 292–293
  signals, 282
  time, 284
  Uragan M and K, 280–281, *281*
GNSS (Global Navigation Satellite System), 73, 215, 279–284, *281*
  accuracy, 291–292
  Beidou, 288–289
  constellations, 292–293, *293*
  future of, 289–292

  - GALILEO, 284–288
  - interoperability, 292–293
  - QZSS, 289
- Goddard Space Flight Center, 131
- GPS
  - accuracy, **70**, **290–291**
  - carrier phase ranging, 12–14
  - civilian applications, 83
  - constellation, 292–293, *293*
  - control, 2
  - Doppler effect, 26
  - error sources, 202–203, 217–218
  - interoperability, 292–293
  - observables, 15
  - one-way ranging, 9–10, 14–15
  - passive system, 1
  - real-time, 91, 101
  - time, 1
  - *vs.* trilateration, 1–2, 12–14
- GPS message; *See* Navigation message
- GPS Space Segment, 267
- GPS space vehicles, 65
- GPS time
  - carrier phase measurement, 97
  - clock drift, 36–37
  - rate of, 4–5
  - solar units, 74
  - standard, 5
- GPS week, 5–6
- gravity, 167
- grid factor, 157
- ground distance, 154–157, *155*
- Ground Earth Stations, 206
- ground plane, 45, *46*, 93
- group delay, 37–40
- GRS80 (Geodetic Reference System 1980), 134, 168
- GSI (Geographical Survey Institute), 182
- Gunter's chain, 12

H

- handover word (HOW), 3, 7–8, 97
- HARN (High Accuracy Reference Network), 138–139, 179
- HDOP (horizontal dilution of precision), 75
- height
  - conversion, 169, *170*
  - ellipsoidal, 156, 162–163, *162*, 212
  - geoidal, 130, 162, 168–169, 212
  - orthometric, 162, 163–164, 169, 185–186, 212, 248
- height of instrument (H.I.), 231, 243, 245, 248
- helix antennas, 93
- Helmert transformation, 137
- high gain antenna, 92
- Hiran (high-precision Shoran), 64
- Hopfield model, 43
- horizontal control, 179, *180*, 181
  - drawing baselines, 194–202, *196*
  - survey design, 179, *180*, 181, 184–185, *186*, *187*, *195*
- horizontal distance at mean elevation, 154–157, *155*
- HPGN (High Precision Geodetic Networks), 139
- HTDP (Horizontal Time Dependent Positioning), 137, 183
- HTMOD (National Height Modernization Program), 139
- Huygens, Christian, 127
- hydrogen masers, 67, 69

I

- IERS (International Earth Rotation Service), 123, 136
- IF (intermediate frequency), 95
- IGLD85 (International Great Lakes Datum of 1985), 166, 167
- IGGS (International GPS Geodynamics Service), 83
- IGS (International GNSS Service), 84, 182, 218, 280
- in-fill surveys, 248
- in phase, 12
- independent lines, *196*, 197–198, *200*
- initialization, 214, 290
- input/output (I/O) port, 204
- integer ambiguity; *See* cycle ambiguity
- integer cycle ambiguity; *See* cycle ambiguity
- integer fixing; *See* cycle ambiguity
- interferometry, 66, 72
- intermittent-static positioning, **107**, 112–113
- International Date Line, 158
- International Ellipsoid, 161
- interoperability, 280, 286–288, *286*, 292–293
- Interrange Operation Number, 79
- IOC (initial operational capability), 276
- IODC (Issue of Data Clock), 5
- IODE (Issue of Date Ephemeris), 6
- ionosphere
  - effects of, 37–40
  - gradients, 38
  - sun and, 38
- ionospheric biases, 216–217, 278
- ionospheric delay, 39, 192, 194
  - clocks and, 69
  - Control Segment, 7
- isostatic rebound, 167
- isotropic antenna, 94
- ITRF (International Earth Rotation Service Terrestrial Reference Frame), 83
- ITRF00, 134, 136, 183

ITRS (International Terrestrial Reference System), 136, 182
ITU (International Telecommunications Union), 289
IUGG (International Union of Geodesy and Geophysics), 134, 140

J

Joint Program Office (JPO), 70
JPL (Jet Propulsion Laboratory), 218
Julian date, navigation code, 6
Julian day, 195, *196*, 197

K

Kalman, R.E., 82
Kalman filtering, 81, 82
K-band, 63
Kepler, Johann, 6
K factor, 150–157, *152*, *155*
kinematic point positioning, 47
kinematic positioning, 46–47, 56, 101
  accuracy, 110–111
  applications, 111
  carrier phase receivers, 111
  leapfrog, 110
  lock, 111
  receivers, 97, 110–112
  reconnaissance, 111
  signal width, 111–112
kinematic relative positioning, 47
K satellites, 286

L

L1, 3, 8
  with M-code, *270*
  atmospheric bias, 278
  code modulation, *16*
  correlation protection, 278
  GALILEO, 285–288, *286*
  GLONASS, 283–284
  interoperability of, 288
  ionospheric effects, 39–40
  phase center coordinates, 183–184
  power of, 269
  tropospheric effects, 42–46
  wavelength of, 23
L1C, 278–279
L2, 3, 8
  atmospheric bias, 278
  correlation protection, 278
  GLONASS, 283–284
  ionospheric effects, 39–40
  phase center coordinates, 184
  power of, 268, 269, *270*, *271*
  tropospheric effects, 42–46
  wavelength of, 24
L2C, 8, 271–274
  chipping rate, 273, 274
L5, 8, 80, 274–276, *275*
  atmospheric bias, 278
  chipping rate, 275
  correlation protection, 278
LADGPS (Local Area Differential GPS), 205
Lambert conic map projection, 121, 140, 146–147, 149, 150, 151
land ownership, 236
Laplace correction, 181, 188
lasers, 246
late (L) correlators, 19–21
latitude, 123–124, 132; *See also* astronomic latitude; geocentric latitude; geodetic latitude
  converting to UTM coordinates, 161
  NGS control data sheets, 178–179, *180*, 181
  on station visibility diagrams, 238
L-band, 2, 63, 283–284
leap seconds, 37, 284
leapfrog kinematic positioning, 110
least squares adjustment, 256–257
LEFT1, 138
Legacy Accuracy Improvement Initiative (L-AII), 41, *42*
LF (low-frequency range), 64–65
LHCP (Left Hand Circular Polarized), 45
LHGS (Local Geodetic Horizon System), 163
line offset, *247*
lines of exact scale, 144, *161*
Listing, J.B., 132
lithosphere, 127
LLR (lunar laser ranging), 65, 135
local coordinate system, 142, *143*
Local Mean Sea Level, 166
localization, 248–249
lock, lost, 55
longitude, 123–124, 132; *See also* geodetic longitude
  converting to UTM coordinates, 161
  NGS control data sheets, 178–179, *180*, 181
  on station visibility diagrams, 239
loop closure, 197–198
loops, forming, 199
Loran-C (long-range navigation-C), 64, **70**
Loverro, Colonel Douglas, 272
lunar day, 164

M

magnetic declination, 246
map projections, 121, 127, 140, 141–143

  conformal, 147
  conic, 141, *142*, 144, *146*, *149*
  cylindrical, 141, *142*, 144, *146*, *149*
  distortions, 143–144, *145*
  Lambert, 121, 140, 146–147, 149, 150, 151
  Mercator, 121, 140, 147, *149*
  Mercator Oblique, 147
  Mercator Transverse, 147, *149*, 150, 151
  orthomorphic, 147
  secant, 144–147
  SPCS, 147–148
  Universal Transverse Mercator, 140–141, 147, 158–161, *159*, *160*, 161, *161*, 181
mapping browsers, online, 178
maps, static GPS survey design, 178
Maritime DGPS, 206
marking techniques, 229–230, 236
mask angle, 45, 73
  lowering, 247–248
Master Frame, 3
M-code, 270–271, *270*, *271*
  overlay issue, 287–288
MCS (Master Control Station), 37, 41, 81
mean equator, 124, *132*
Mercator map projection, 121, 140, 147, *149*
  Oblique, 147
  Transverse, 147, *149*, 150, 151
  Universal Transverse, 140–141, 147, 158–161, *159*, *160*, 161, *161*, 181
mesosphere, 37, *38*
metonic cycle, 166
microprocessors, 100–101
microstrip antenna, 93
mixer, 95
MMD (Mean Mission Duration), 79, 80
  Block II satellites, 267
  Block IIF satellites, 275
modulated carrier wave, 9–11
  cycle ambiguity, 14–15
  distance measurement, 11–12
  phase shift, 12, *13*, 14, *14*
monitoring stations, 81
monumentation, 235, 241
moon
  orbital period, 165
  tidal forces, 164–167, *165*
MSAS, 206
MSL (Mean Sea Level), 132, 154, 164, 166
  and geoid, 167–170
  zero point, 167
multichannel receivers, 105
multipath, 43–45, *44*, *46*
  and antenna design, 45, *46*, 93, 240
  in DGPS, 213–214, 247–248
  minimizing with signal squaring, 100
  RTK surveying, 213–214, 247–248
  station visibility diagrams, 240
multiplexing, 273
multiplexing receiver, 96
multireceiver positioning, 49

N

NAD27 (North American Datum of 1927), 64, 130, 135, 137–138, 139
  SPCS, 148–150
NAD83 (North American Datum of 1983), 134, 135, 137–138, 139, 183
  positions, 140–141
  SPCS, 148–150
NADCON, 138
narrow correlators, 20
narrow lane, 113
NASA (National Aeronautics and Space Administration), 136, 140
NATO (North Atlantic Treaty Organization), 288
NAVD88 (North American Vertical Datum of 1988), 167
navigation code, 2–8
  almanac, 7, 18
  antispoofing, 7
  atmospheric correction, 7
  broadcast ephemeris, 6
  frequency, 3–4
  GPS time, 4–5
  GPS week, 5–6
  Julian date, 6
  satellite clocks, 5
  wavelength, 3
Navigation message, 3–4, *4*, 100
  carrier phase measurement, 97
  clock corrections, 37
  clock correlation, 4–5
  group delay, 39
  L1C satellites, 279
  WGS84, 134
navigation solution, 47–48, *48*
Navstar 1, 78
Navstar 2, 78–79
Navstar 7, 79
NAVSTAR GPS, 69–81
  civilian applications, 71, 72–73
  clocks, 69
  military applications, 69–71
  passive system, 70
  program development, 69–73
  satellite orbit, 71, 73–81, *73*
  segment organization, 73–84
Navy Navigational Satellite System (NNSS), 67–69, *68*
NDGPS (Nationwide DGPS), 206
NDU (navigation data units), 275
networks, 49
Newton, Sir Isaac, 127–128, 134

NGA (National Geospatial-Intelligence Agency), 41, *42*, 134, 158
NGS (National Geodetic Survey), 83, 121, 135
  calibration baselines, 232
  control data sheets, 178–179, *180*, 181–182
  CORS, 183–184
  GEOID program, 181
  SPCS83 program, 152
NGS Survey Control Map, 179
NGVD29 (National Geodetic Vertical Datum of 1929), 166
NMEA (National Marine Electronics Association), 217
NOAA (National Oceanic and Atmospheric Administration), 83, 135, 136
nontrivial lines, *196*, 197–198, *200*
northings, 160
NRCan (National Resources Canada), 218
NSRS (National Spatial Reference System), 178, 249

O

Oblique Mercator projection, 147
observables, 15–17
observation schedule, static GPS, 232, **233**
observed cycle count, 95
observing, 242–245
  arrival, 242
  auxiliary equipment, 232
  communications, 230–231
  conventional equipment, 229–230
  daily, 245
  flagging, 236
  GPS equipment, 231–232
  logs, 243–245, *244*
  monumentation, 235–236, 241
  papers, 232
  photographs, 235
  quad sheet name, 235
  rubbings, 235
  safety equipment, 230
  scheduling, 241–242
  setup, 242–243
  station data sheet, 232, 234–236, *234*
  to-reach description, 233, 235–236
  visibility diagrams, 236–240, *237*, 239
  weather, 245
obstructions, visibility diagrams, 236–240, *237*, **239**
OCS (Operational Control Segment), 40–42, *42*, 134
offsets, point, *241*, 244
Omega, 64–65, **70**
one percent rule, 23–24
one-way ranging, 9, *10*, 14–15
Open Service, 285, 288
optical tracking systems, 65–66
OPUS (Online Positioning User Service), 257
orbit
  eccentric, 36
  posigrade, 73–74
orbit calculation, GPS for, 83
orbit errors, 216–217
orbital bias, 40–42
orthometric elevations, 162, 163–164, 169, 185–186, 212, 248
oscillators, 11; *See also* satellite clocks
  imperfect, 21
  standard rate of, 16–17
OTF (on-the-fly), **107**, 114, 209
out of phase, 12
outages, 76

P

parallel receivers, 95, 96
patch antennas, 93, 94
P-code, 3–4, 63
  antenna bandwidth, 93
  antispoofing, 80
  bandpass filter, 95
  chipping rate, 17, 268
  encrypted, 7
  generation rate, 8
  group delay, 39
  maximum precision, 23
  PPS, 8–9
  PRN, 8
  pseudoranges, 114
  replica, 95
  signal power, 268
  spread spectrum, 15–16
  squaring, 100
PDOP (position dilution of precision), 75, 76, *78*, 189–190, 192, 238
perigee, 165
permissions, obtaining, 188–189
phase ambiguity, 51, 55–56, 111
phase center
  antennas, 94, 96, 243
  coordinates, 183–184
phase delay, 37–40
phase differencing, VLBI, 67
phase lock, 12, 96
phase-locked loop, 273
phase measurements, sampling rate, 106
phase modulation, encoding by, 15–17
phase shifted modulated carrier wave, 12, *13*, 14, *14*
Philosophiae naturalis principia mathematica, 127
photogrammetry, 114, 208

photographs
    mark/monument, 235
    triangulation with, 65
Picard, J., 128
PID (Permanent Identifier), 179
piezoelectric effect, 40
plane coordinate system, 121–122, 140–141
plane surveying, 122–123
PLL (phase locking loop), 96, 97
point offsets, RTK surveying, 246–247, *246*
polar motion, 122–123
polar plot, *77*, 190, 192, 236–240, *237*, **239**
polar regions, 158
poles, flattening, 127–128, 134
posigrade orbit, 73–74
positioning, classifications, 46–49
postprocessing, 183–184, 202
    control, 251
    correlation of biases, 250
    cycle slip, 253–255
    daily, 249
    double difference, 252–253, 255
    downloading data, 250
    file naming, 250
    first position, 251–252
    least squares adjustment, 256–257
    memory capacity, 251
    obtaining vectors, 255
    services, 257
    triple difference, 252, 254–255
power, back up, 102–103, 232
PPP (Precise Point Positioning), 217–218
PPS (Precise Positioning Service), 8–9
preamplifier, 95
precise orbit, 184
Prime Minitrack, 66
Principia, 127
PRN (pseudorandom noise)
    C/A code, 8
    codes, 275
    number, 79
    P-code, 8
project points, plotting, 186–187, *186*, *187*, *195*
projection factor, 150–157, *152*, *155*
prompt (P) correlators, 19–21
propagation delay, 17
PRS (Public Regulated Service), 285, 287, 288
PSD (power spectral diagrams), 268–269
pseudokinematic positioning, 47, **107**, 112–113
    applications, 113
    productivity, 112
pseudorange, 15, 17–24, 95, 99
    autocorrelation, 18–19
    correction, 204
    correlation peak, 19–21, *20*
    DGPS, 110, 214
    equations, 21–23
    Kalman filtering, 82
    lock, 21
    multipath in, 44–45
    one-percent rule, 23–24
    P-code, 114
    precise, 209
    propagation delay, 17
    RF section, 96
    time shift, 21
pseudostatic positioning, **107**, 112–113
public–private partnerships (PPP), 285

Q

QPSK (Quad Phase Shift Key), 275
Quad sheet, 235
quadrafilar antenna, 93
quadrature, 17
QZSS (Quasi-Zenith Satellite System), 278, 280, 289
    interoperability, 292–293

R

radar, 63
radial surveys, 113
radio, antenna, 211
radio license, RTK surveying, 210
radio modulator, 211
radio positioning
    extraterrestrial, 66, *67*
    terrestrial, 63–65
radio telescopes, 66–67
radio waves, 2–3, 230
radionavigation, 72
range biases, 35
range delay, 43–45, *44*, *46*
range rate, 97
range rate correction, 206
ranges, GPS, measuring, 1
rapid static positioning, 47, 56, **107**, 113
RDOP (relative dilution of precision), 75
real-time DGPS, 110, 204–205, *205*, 215
real-time GPS, 91, 101
receiver noise, 45–46
receiver-satellite-time triple difference, 51, *54*, 55–56
receivers, 1
    accuracy of, 72–73
    antenna, 92–94
    base station, 202–203
    baselines, *196*, 197–198, *200*
    battery power, 102–103, 232
    block diagram, 91, *92*
    carrier phase, 231
    CDU, 100, 101–102

  choosing, 103–106
  clocks, 21
    errors, 40, 51, *53*
    relativistic effects, 36
  code correlation, 15, 17–21
  codeless, 104, 105
  cost, 106
  development trends, 104–106
  DGPS, 100, 101, 202–214
  dual-frequency, 39–40, 95, 105, 211, 213
  general features, 91, *92*, 106
  handheld, 103
  kinematic positioning, 110–112
  locked on, 21
  memory capacity, 251
  microprocessor, 100–101
  mounting, 231
  multichannel, 105
  preprogramming feature, 230
  pseudokinematic positioning, 112–113
  rapid-static positioning, 113–114
  replica code, 17–18
  RF section, 94–97
  roving, 46–47, 56
  shared antennas, 231
  single frequency, 39
  spacing, 43
  static, 47, 97–99
  types of, 103
reconnaissance, kinematic positioning, 111
reconstructed carrier phase, 25
redundancy, 96
  control stations, 198–199
  GPS networks, 231
  in radial GPS, 203
  in RTK surveying, 248
Reference Meridian Plane, 122
reference wave, 14, *14*
reflected wave, 14, *14*
relative positioning, 47, 49
Remondi, Benjamin, 47, 110, 208
remote sensing, 208
replica code, 17–18
retroprism, 9
RF (Radio Frequency) section, 92, 94–100
  carrier phase measurement, 97
  carrier tracking loop, 97
  channels, 95
  cycle ambiguity, 99
  multiplexing, 96
  preamplifier, 95
  pseudorange, 96
  signal squaring, 99–100
  tracking loops, 96
RHCP (Right Hand Circular Polarized), 45, 92
RINEX (Receiver Independent Exchange Format), 109, 251
route marking, 229–230, 236
rovers, 202–214
RTCM (Radio Technical Commission for Maritime Services), 204, 215
RTK (real-time kinematic) surveying, 114–115, 208, *209*
  applications, 215
  base station, 214
  cycle ambiguity in, 208–210, 212
  dual-frequency receiver, 213
  dynamic lines, 247
  error sources, 202–203
  initialization, 214
  multipath in, 213–214, 247–248
  point offsets, 246–247, *246*
  radial GPS in, 203–204, *203*
  radio link, 210
  redundancy in, 248
  satellite constellations, 212
  site calibration, 248–249
  typical set-up, 211–212
  vertical components, 212
RTN (real-time network), 215–218, *216*, 251
  differential corrections, 216
rubbings, 235
rubidium clocks, 5, 36, 40, 58, 69, 79, 267

S

SA (Selective Availability), 9, 72
  C/A code, 80
  Block II satellites, 267
  carrier phase, 24
Saastamoinen model, 43
safety equipment, 230
safety factor, 201
Safety of Life Service, 285
sampling rate, phase measurements, 106
satellite clocks
  biases, 35–37, 49, 51, 202–203, 217–218
  drift, 5, 36–37
  GPS time standard, 4–5
  imperfect synchronization, 21
  relativistic effects, 36
satellite tracking, 66
satellites
  advantages of, 64–65
  common characteristics, 80–81
  design lives, 73
  Doppler shift and, 68
  elevation, 38–39
  GPS signals, 1
  health of, 7–8
  names, 78–81
  orbit, 1, 35
  orbital bias, 40–42
  position of, 2

    prototypes, 67
    windows, 189–190, 192, 194
S-band, 63
SBAS (space-based augmentation system), 206, 272, 280
scale factor, 150–157, *152*, *155*
Scaled, 179
scheduling, 241–242
SCIGN (Southern California Integrated GPS Network), 182
SCOUT (Scripps Coordinate Update Tool), 257
sea level factor, 154
Search and Rescue Service, 285
secant projections, 144–147
segment organization, GPS
        control segment, 81–83
        space segment, 73–81
        user segment, 83–84
semikinematic positioning, 47
sequencing receiver, 96
sessions, 101, 102
    kinematic positioning, 110–111
    naming, 195, *196*, 197, *200*
    number of, 199, *200*, 201–202
Shoran (short range navigation), 63–64
sideshot, 247
signal processing, RF section, 94–100
signal splitter, 231
signal squaring, 99–100
signals
    control, 2
    Doppler shift, 97–100
    ionospheric effects, 37–40
    large capacity, 70–71
    multipath, 43–45, *44*, *46*
    new, 269, 276, *277*, 278
    sources of inaccuracies, 202–203, 217–218
    spread spectrum, 15–16
    travel time, 1
    tropospheric effects, 42–46
    wideband, 111–112
single difference, 49–51, *50*, *52*
single-point positioning, 47
site calibration, 248–249
site evaluation, virtual, 178
skyplot, 236–240, *237*, **239**
SLD29 (Sea Level Datum of 1929), 164, 166
SLR (satellite laser ranging), 65–66, 123, 135
small circles, 144
SNR (signal-to-noise ratio), 96, 247–248, 275
    bandwidth, 112
software, survey design, 189
SOPAC (Scripps Orbit and Permanent Array Center), 184, 257
South American Datum 1969, 130, *131*, 195
Space Vehicle Number, 79
SPCS (State Plane Coordinate System), 141
    development of, 146–147
    map projections, 147–148
    scale and distance, 150–157, *152*, *155*
    SPCS27 to SPCS83, 148–150
spirit leveling, 162, 163–164
    vertical control positions, 185–186
spread spectrum, 71, 91
    C/A codes, 15–16
    despreading, 111–112
    modulation techniques, 16–17
    P-code, 15–16
SPS (Standard Positioning Service), 8–9, 72
Sputnik, 64, 68
Sputnik 2, 131
Stanford Linear Accelerator, 72
state plane coordinates, 121–122, 140–141, 181
static GPS, 91, *107*, **107**
    azimuth marks, 188
    communications equipment, 230–231
    compatible receivers, 109
    CORS, 182–184
    cycle ambiguity, 108
    data interval, 109
    data processing, 108
    design facts, 189–194
    drawing baselines, 194–202, *196*
    evaluating access, 187–188
    NGS data sheets, 178–179, *180*, 181–182
    observation schedule, 232, **233**
    observation settings, 109
    offsets, 188
    PDOP, 189–190, 192
    permissions, 188–189
    planning, 177–178
    plotting project points, 186–187, *186*, *187*, *195*
    point offsets, 240, *241*
    preprogrammed observations, 108–109
    prerequisites, 106
    productivity, 108
    session length, 108
    survey design, 177–178, 184–202
static point positioning, 47
static positioning, 47
static relative positioning, 47
station coordinates, approximating, 238
station data sheet, 232, 234–236, *234*
station name, 234–235
station visibility diagram, 236–240, *237*, **239**
stations; *See* control stations
sun
    ionosphere and, 38
    tidal forces, 164–167, *165*
Superseded Survey Control, 179
survey design
    azimuth and elevation, 190, **191**, **193**
    choosing the window, 189–190, 192, 194
    class, 181

  coordinates, 181
  drawing baselines, 194–202, *196*
  evaluating access, 187–188
  facts, 189–194
  horizontal control, 179, *180*, 181, 184–185, *186*, *187*, *195*
  NGS control data sheets, 178–179, *180*, 181–182
  order, 181
  PDOP, 189–190, 192
  permissions, 188–189
  planning offsets, 188
  plotting project points, 186–187, *186*, *187*, *195*
  polar plots, 190, 192
  static GPS, 184–202
  variables, 194
  vertical control, 179, *180*, 181, 185–186, *187*, *195*
survey masts, 229
survey methods
  comparison, **107**
  DGPS, 110, 202–208, 213–215
  kinematic GPS, **107**, 110–112
  OTF, **107**, 114
  pseudokinematic positioning, **107**, 112–113
  rapid-static, **107**, 113–114
  real-time network services, 215–218, *216*
  RTK, 114–115, 202–204, 208–215
  static GPS, 106–110, 177–202
survey project identifier, 179
surveyors
  contact between, 230
  and geodesy, 121–122
Syme, George F., 145
synchronization information, 3

T

TDOP (time dilution of precision), 75
telemetry (TLM) word, 3, 7–8
telescopes, radio, 66–67
terrestrial radio positioning, 63–65
thermosphere, 37, *38*
tidal forces, *156*, 164–167
tide station gauges, 164
time dilation, 36
time measurement, 1
  EDM signals, 9
  GPS signals, 9–10
  oscillators, 11, 16–17, 21
  standards, 4–5
time shift, 17
  and code correlation, 18–19, *19*, 20–21, *20*
  pseudorange, 21
time tags, 67
to-reach description, **233**, 235–236
total electron count (TEC), 37
tracking facilities, 41–42, *42*, 81; *See also* Control Segment; control stations
traffic control, 230
TRANSIT, 67–69, *68*, **70**
  and GPS, 69
transmitter, setting up, 213
Transverse Mercator projections, 147, *149*, 150, 151
tri-stations, 135
triangulation, with photographs, 65
tribrach, 23, 231, 243
trilateration survey, *vs.* GPS, 1–2, 12–14
triple difference, 51, *54*, 55–56, 252, 254–255
tripod, 213
trivial lines, *196*, 197–198, *200*
troposphere
  effects, 42–46
  refraction in, 43
tropospheric biases, 216–217
turns, describing, 236
two-way ranging, 9, *10*
  and cycle ambiguity, 14–15

U

UERE (user equivalent range error), 35, *36*
UFCORS (User Friendly CORS), 184
unknowns, 47–48
UPS (Universal Polar Stereographic) projection, 158, 160–161
U.S. Coast & Geodetic Survey, 135
U.S. Coast Guard, 206
U.S. Department of Defense (DoD), 270
  NAVSTAR GPS, 69–73
U.S. Department of Transportation (DOT), 206
USGS (U.S. Geological Survey), 136
  quad sheet, 179, 235
USNO (U.S. Naval Observatory), 4
UT (Universal Time), 74
UTC (Coordinated Universal Time)
    and clock drift, 36–37
    rate of, 4–5, 284
UTM (Universal Transverse Mercator), 147
  converting coordinates, 161
  coordinates, 140–141, 158, 181
  scale factor, 158
  zones, 158–161, *159*, *160*, *161*

V

variables, survey design, 194
vector solutions, cycle ambiguity, 255
vectors, correlation of, 250
VERTCON, 179
vertical control, 179, *180*, 181, 185–186, *187*, *195*, 202
vertical datum

  evolution of, 164–165
  spirit leveling, 162, 163–164
visibility diagrams, 236–240, *237*, **239**
VLBI (very long baseline interferometry), 66–67, *67*, 123, 135
VNIIFTRI, 284
VRS (Virtual Reference Station), 217

W

WADGPS (Wide Are Differential GPS), 205
wavelength, distance measurements, 11–12
waypoint, 103
week, GPS, 5–6
WGS84 (World Geodetic System 1984), 6, 122, 134, 136–137
wheels, 42
wide laning, 113–114
wideband signal, 111–112
windows, satellite, 189–190, 192, 194

X

X-band, 63

Y

Y code, 7, 80, 105, 268
  antenna bandwidth, 93

Z

Z-count, 3, 8
zenith, 125–128, *126*
zenith gain, 93
zero baseline test, 231
zero meridian, 124, *132*
zero point, 167

Environmental Engineering

GPS for Land Surveyors

THIRD EDITION

Since the last edition of this international best seller, global positioning systems (GPSs) have grown to become part of a larger international context, the Global Navigation Satellite System (GNSS). Both GPS and GNSS technologies are becoming ever more important in the everyday practice of surveying and mapping. With **GPS for Land Surveyors, Third Edition**, a book written by a land surveyor for land surveyors, you can stay in the know on the latest GPS techniques, technologies, codes, and signals.

What's New to the Third Edition?

- More than 50 of the 80-plus illustrations now appear in color—adding depth and value
- Sections on Real-Time Network Services, Block IIF, and control segment modernization
- GPS code, such as the M-code, L1C, and L2C
- An entire chapter dedicated to GNSS
- Discussion of the Russian GLONASS system
- The Chinese Beidou system, and the Japanese QZSS

From fundamental theory to practical application and advanced technologies, the book covers GPS without pages of complicated math. It demonstrates the basics of GPS technology, common hardware, surveying methods, survey design, and planning and observation. Additionally, each chapter includes helpful review questions and answers with detailed explanations. GPS and GNSS are revolutionizing the practice of surveying and mapping. This user-friendly manual provides you with all the necessary tools to understand and use these important technologies.

CRC Press
Taylor & Francis Group
an **informa** business

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
270 Madison Avenue
New York, NY 10016
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

9195
ISBN: 978-0-8493-9195-8
90000
9 780849 391958
www.crcpress.com