Seeber · Satellite Geodesy

Günter Seeber

Satellite Geodesy

2nd completely revised and extended edition

Author

Günter Seeber, Univ. Prof. Dr.-Ing.
Institut für Erdmessung
Universität Hannover
Schneiderberg 50
30167 Hannover
Germany

1st edition 1993

This book contains 281 figures and 64 tables.

To the memory of my grandson

Johannes

Preface

Methods of satellite geodesy are increasingly used in geodesy, surveying engineering, and related disciplines. In particular, the modern development of precise and operational satellite based positioning and navigation techniques have entered all fields of geosciences and engineering. A growing demand is also evident for fine-structured gravity field models from new and forthcoming satellite missions and for the monitoring of Earth's rotation in space. For many years I have had the feeling that there is a definite need for a systematic textbook covering the whole subject, including both its foundations and its applications. It is my intention that this book should, at least in part, help to fulfill this requirement.

The material presented here is partly based on courses taught at the University of Hannover since 1973 and on guest lectures given abroad. It is my hope that this material can be used at other universities for similar courses. This book is intended to serve as a text for advanced undergraduates and for graduates, mainly in geodesy, surveying engineering, photogrammetry, cartography and geomatics. It is also intended as a source of information for professionals who have an interest in the methods and results of satellite geodesy and who need to acquaint themselves with new developments. In addition, this book is aimed at students, teachers, professionals and scientists from related fields of engineering and geosciences, such as terrestrial and space navigation, hydrography, civil engineering, traffic control, GIS technology, geography, geology, geophysics and oceanography. In line with this objective, the character of the book falls somewhere between that of a textbook and that of a handbook. The background required is an undergraduate level of mathematics and elementary mathematical statistics. Because of rapid and continuous developments in this field, it has been necessary to be selective, and to give greater weight to some topics than to others. Particular importance has been attached to the fundamentals and to the applications, especially to the use of artificial satellites for the determination of precise positions. A comprehensive list of references has been added for further reading to facilitate deeper and advanced studies.

The first edition of this book was published in 1993 as an English translation and update of the book "Satellitengeodäsie", that was printed in the German language in 1989. The present edition has been completely revised and significantly extended. The fundamental structure of the first edition has been maintained to facilitate continuity of teaching; however, outdated material has been removed and new material has been included. All chapters have been updated and some have been re-written. The overall status is autumn 2002 but some of the most recent technological developments to March 2003 have been included.

Extensions and updates mainly pertain to reference coordinate systems and reference frames [2.2], signal propagation [2.3], directions with CCD technology [5.2], the Global Positioning System (GPS) and GNSS [7], satellite laser ranging [8], satellite

altimetry [9], gravity field missions [10] and applications [12]. In particular, the chapter on GPS and GNSS [7] has been almost completely re-written and now covers about 200 pages. Together with chapters [2], [3], and [12], it forms a comprehensive GPS manual on its own. New technological developments of the space and user segment are included, as is the current state of data analysis and error budget. Differential GPS and permanent reference networks are now treated in a comprehensive section of their own [7.5]. GLONASS and the forthcoming GALILEO are included in a new section on GNSS [7.7].

Gravity field missions like CHAMP, GRACE and GOCE, because of their increasing importance, are dealt with in a new chapter [10]. VLBI, together with the new inclusion of interferometric SAR, form another new chapter [11]. Coverage of historical techniques like photographic camera observations [5] and Transit Doppler [6] has been considerably reduced. The basic principles, however, are still included because of their historical importance and because they are shared by new technologies like CCD cameras [5.2] and DORIS [6.7]. The geodetic history of Transit Doppler techniques, in addition, is an excellent source for understanding the evolution and basic concepts of the GPS. The chapter on applications, now renumbered [12], has been updated to include modern developments and a new section on the combination of geodetic space techniques [12.5]. International services of interest to satellite geodesy have been included, namely the IGS [7.8.1], the ILRS [8.5.1], the IVS [11.1.3], and the IERS [12.4].

The bibliography has been updated and expanded considerably by adding an increased number of English language references. The total number of references is now reaching 760, about half of which are new in this edition.

Many of the examples within this book are based on field projects and research work carried out in collaboration with my graduate students, doctorate candidates and scientific colleagues at the University of Hannover over more than 20 years. I would like to thank all these individuals for their long standing cooperation and the many fruitful discussions I have had with them. In addition, the help of the staff at the Institut für Erdmessung is gratefully acknowledged. Most figures have been redrawn by cand. geod. Anke Daubner and Dipl.-Ing. Wolfgang Paech.

My sincere thanks for checking and correcting the English language go to Dr. Graeme Eagles of the Alfred Wegener Institut für Polar- und Meeresforschung, Bremerhaven. I should also like to thank the many colleagues from all over the world who helped to improve the book through their comments on the first edition, and the individuals and organizations who provided illustrations.

Finally my gratitude goes to my wife Gisela for her never ending support and understanding. The publisher remained excellently cooperative throughout the preparation of this book. My cordial thanks go to Dr. Manfred Karbe, Dr. Irene Zimmermann, and the staff at Walter de Gruyter.

Hannover, May 2003   Günter Seeber

Contents

Abbreviations

ACP	Area Correction Parameter	CTP	Conventional Terrestrial Pole
ADOS	African Doppler Survey	CTS	Conventional Terrestrial (Reference) System
AI I	Accuracy Improvement Initiative		
APL	Applied Physics Laboratory	DÖDOC	German Austrian Doppler Campaign
ARP	Antenna Reference Point		
AS	Anti Spoofing	DD	Double Difference
ASIC	Application-Specific Integrated Circuit	DEM	Digital Elevation Model
		DGFI	Deutsches Geodätisches Forschungsinstitut
BCRS	Barycentric Celestial Reference System		
		DGPS	Differential GPS
BIH	Bureau International de l'Heure	DOD	Department of Defence
BIPM	Bureau International des poids et Mésures	DOP	Dilution of Precision
		DOY	Day Of the Year
BKG	Bundesamt für Kartographie und Geodäsie	DRMS	Distance Root Mean Square
		EDOC	European Doppler Campaign
BPS	Bits Per Second	EGM96	Earth Gravitational Model 1996
BPSK	Binary Phase Shift Keying	EGNOS	European Geostationary Navigation Overlay System
CACS	Canadian Active Control System		
CAD	Computer Assisted Design	EOP	Earth Orientation Parameter
CBIS	Central Bureau (IGS) Information System	EOS	Earth Observing System
		EPS	Real-Time Positioning Service (SAPOS)
CCD	Charge Coupled Device		
CDP	Crustal Dynamics Program	ERM	Exact Repeat Mission
CEP	Celestial Ephemeris Pole	ERP	Earth Rotation Parameter
CEP	Circular Error Probable	ESA	European Space Agency
CIO	Conventional International Origin	ESNP	European Satellite Navigation Program
CIP	Celestial Intermediate Pole		
CIS	Conventional Inertial (Reference) System	EU	European Union
		FAA	Federal Aviation Administration
CNES	Centre National d'Études Spatiales	FAGS	Federation of Astronomical and Geophysical Data Analysis Services
CONUS	Continental U.S.		
CORS	Continuously Operating Reference Station	FIG	Fédération Internationale des Géomètres
CPU	Central Processing Unit	FK5	Fifth Fundamental Catalogue
CRF	Celestial Reference Frame	FOC	Full Operational Capability
CRS	Celestial Reference System	FRNP	Federal Radio Navigation Plan

GAST	Greenwich Apparent Sidereal Time
GCRS	Geocentric Celestial Reference System
GDR	Geophysical Data Record
GEM	Goddard Earth Model
GEO	Geostationary Orbit
GFO	GEOSAT Follow On
GFZ	GeoForschungsZentrum Potsdam
GIC	GPS Integrity Channel
GIS	Geo Information System
GLAS	Geoscience Laser Altimeter System
GM	Geodetic Mission
GMST	Greenwich Mean Sidereal Time
GNSS	Global Navigation Satellite System
GRGS	Groupe de Recherche de Géodésie Spatiale
GRS80	Geodetic Reference System 1980
GSFC	Goddard Space Flight Center
HEPS	High Precision Real-Time Positioning Service (SAPOS)
IAU	International Astronomical Union
ICD	Interface Control Document
ICO	Intermediate Circular Orbit
ICRF	International Celestial Reference Frame
ICRS	International Celestial Reference System
IDS	International DORIS Service
IERS	International Earth Rotation and Reference Systems Service
IF	Intermediate Frequency
IGEB	Interagency GPS Executive Board
IGN	Institut Géographique National
IGS	International GPS Service
IGSO	Inclined Geo-synchronous Orbit
ILRS	International Laser Ranging Service
ILS	International Latitude Service
INSAR	Interferometric SAR

ION	Institute of Navigation
IPMS	International Polar Motion Service
IRIS	International Radio Interferometric Surveying
IRM	IERS Reference Meridian
IRP	IERS Reference Pole
IRV	Inter-Range Vector
ITRF	International Terrestrial Reference Frame
IUGG	International Union of Geodesy and Geophysics
IVS	International VLBI Service
JD	Julian Date
JGM	Joint Gravity Model
JGR	Journal of Geophysical Research
JPL	Jet Propulsion Laboratory
LADGPS	Local Area Differential GPS
LAN	Longitude of Ascending Node
LBS	Location Based Service
LEO	Low Earth Orbiter
LLR	Lunar Laser Ranging
LOD	Length of Day
MAS	Milli Arc Second
MEO	Medium Earth Orbit
MERIT	Monitoring Earth Rotation and Intercomparison of Techniques
MJD	Modified Julian Date
MSAS	Multifunctional Satellite-based Augmentation System
NAD	North American Datum
NASA	National Aeronautics and Space Administration
Nd:YAG	Neodymium Yttrium Aluminium Garnet
NDGPS	Nationwide Differential Global Positioning System
NEOS	National Earth Orientation Service
NGS	National Geodetic Survey
NIMA	National Imagery and Mapping Agency

NIST	National Institute of Standards	SNR	Signal-to-Noise Ratio
NOAA	National Oceanic and Atmospheric Administration	SPAD	Single Photon Avalanche Diode
OCS	Operational Control Segment	SPS	Standard Positioning Service
PCV	Phase Center Variation	SST	Satellite-to-Satellite Tracking
PDA	Personal Digital Assistant	SST	Sea Surface Topography
PDGPS	Precise Differential GPS	SV	Space Vehicle
PDOP	Position Dilution of Precision	SVN	Space Vehicle Number
PE	Precise Ephemerides	SWH	Significant Wave Height
POD	Precise Orbit Determination	T/P	TOPEX/POSEIDON
PPP	Precise Point Positioning	TAI	International Atomic Time
PPS	Precise Positioning Service	TCB	Barycentric Coordinate Time
PRN	Pseudo Random Noise	TCG	Geocentric Coordinate Time
PTB	Physikalisch Technische Bundesanstalt	TDB	Barycentric Dynamical Time
RA	Radar Altimeter	TDT	Terrestrial Dynamical Time
RDS	Radio Data System	TEC	Total Electron Content
RF	Radio Frequency	TECU	Total Electron Content Unit
RMS	Root Mean Square Error	TID	Travelling Ionospheric Disturbances
RNAAC	Regional Network Associate Analysis Center	TIGO	Transportable Integrated Geodetic Observatory
RTCM	Radio Technical Commission for Marine Sciences	TRF	Terrestrial Reference Frame
RTK	Real Time Kinematic	TT	Terrestrial Time
SA	Selective Availability	TTFA	Time To Fix Ambiguities
SAD	South American Datum	UEE	User Equipment Error
SAO	Smithsonian Astrophysical Observatory	UERE	User Equivalent Range Error
SAPOS	Satellite Positioning Service	URE	User Range Error
SAR	Search And Rescue	USCG	U.S. Coast Guard
SAR	Synthetic Aperture Radar	USNO	U.S. Naval Observatory
SBAS	Satellite Based Augmentation System	UT	Universal Time
SEP	Spherical Error Probable	UTC	Universal Time Coordinated
SGG	Satellite Gravity Gradiometry	VLBA	Very Long Baseline Array
SI	International System of Units	VLBI	Very Long Baseline Interferometry
SIR	Shuttle Imaging Radar	VRS	Virtual Reference Station
SIS	Signal in Space	VSOP	VLBI Space Observatory Program
SISRE	Signal in Space Range Error	VTEC	Vertical Electron Content
SLR	Satellite Laser Ranging	WAAS	Wide Area Augmentation System
		WADGPS	Wide Area Differential GPS

1 Introduction

1.1 Subject of Satellite Geodesy

Following the classical definition of Helmert (1880/1884), geodesy is the *science of the measurement and mapping of the Earth's surface*. This definition includes the determination of the terrestrial external gravity field, as well as the surface of the ocean floor, cf. (Torge, 2001). *Satellite Geodesy* comprises the observational and computational techniques which allow the solution of geodetic problems by the use of precise measurements to, from, or between artificial, mostly near-Earth, satellites. Further to *Helmert*'s definition, which is basically still valid, the objectives of satellite geodesy are today mainly considered in a functional way. They also include, because of the increasing observational accuracy, time-dependent variations.

The basic problems are

1. determination of precise global, regional and local three-dimensional positions (e.g. the establishment of geodetic control)
2. determination of Earth's gravity field and linear functions of this field (e.g. a precise geoid)
3. measurement and modeling of geodynamical phenomena (e.g. polar motion, Earth rotation, crustal deformation).

The use of artificial satellites in geodesy has some prerequisites; these are basically a comprehensive knowledge of the satellite motion under the influence of all acting forces as well as the description of the positions of satellites and ground stations in suitable reference frames. Consequently satellite geodesy belongs to the domain of *basic sciences*. On the other hand, when satellite observations are used for solving various problems satellite geodesy can be assigned to the field of *applied sciences*. Considering the nature of the problems, satellite geodesy belongs equally to *geosciences* and to *engineering sciences*.

By virtue of their increasing accuracy and speed, the methods and results of satellite geodesy are used more and more in other disciplines like e.g. *geophysics, oceanography* and *navigation*, and they form an integral part of *geoinformatics*.

Since the launch of the first artificial satellite, SPUTNIK-1, on October 4, 1957, satellite geodesy has developed into a self-contained field in geodetic teaching and research, with close relations and interactions with adjacent fields (Fig. 1.1). The assignments and contents are due to historical development.

In *Geodetic Astronomy*, based on the rules of Spherical Astronomy, the orientation of the local gravity vector (geographical longitude Λ, geographical latitude Φ), and the astronomical azimuth A of a terrestrial mark are determined from the observation of natural celestial bodies, particularly fixed stars. By *Gravimetry* we mean the measurement of gravity (gravity intensity g) which is the magnitude of the gravity acceleration vector $\boldsymbol{g}$ (Torge, 1989). With *Terrestrial Geodetic Measurements* horizontal angles,
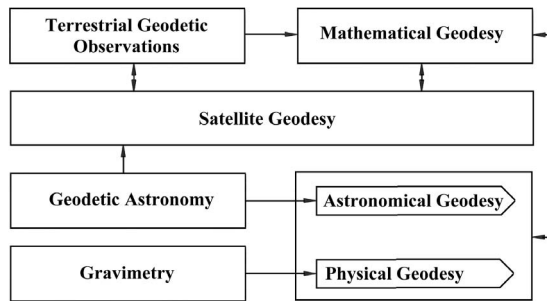
Figure 1.1. Main relations between geodetic fields of teaching and research

distances, zenith angles, and levelled height differences are provided, and serve for the determination of surface point locations. *Satellite Geodesy*, finally, is based on the observation of artificial celestial bodies. Directions, ranges, and range-rates are determined between Earth surface locations and satellites or between satellites. Some measurements, for instance accelerations, are taken within the satellites themselves.

The results of geodetic-astronomic or gravimetric observations are used within the field of *Astronomical and Physical Geodesy* for the determination of the figure and gravity field of Earth (Torge, 2001). In German, this classical domain is called *Erdmessung* (Torge, 2003) and corresponds to the concept of *Global Geodesy* in the English language. The main problems are the determination of a mean Earth ellipsoid and a precise geoid (cf. [2.1.5]).

The determination of coordinates in ellipsoidal or three-dimensional coordinate systems, mainly based on terrestrial geodetic measurements, is treated within the field of *Mathematical Geodesy*. Alternate expressions for this domain are *Geometrical Geodesy* or, in German, *Landesvermessung*, e.g. Großmann (1976). The separate classification of observation- and computation techniques, as developed within the classical fields of geodetic teaching and practice, has not occured to the same extent in satellite geodesy. Here, observation, computation, and analysis are usually treated together. As far as global problems are concerned, satellite geodesy contributes to global geodesy, for example to the establishment of a global reference frame. In regional and local problems, satellite geodesy forms part of surveying and geoinformatics.

Conversely, the fields of mathematical geodesy and geodetic astronomy provide important foundations in satellite geodesy with respect to reference systems. The same is true for the field of astronomical and physical geodesy, which provides information on Earth's gravity field. Due to these close interactions, a sharp separation of the different fields in geodesy becomes more and more difficult, and it is no longer significant.

A combined consideration of all geodetic observables in a unified concept was developed rather early within the field of *Integrated Geodesy*, e.g. Hein (1983). It

finds a modern realization in the establishment of integrated geodetic-geodynamic observatories (see [12.5], Rummel et al. (2000))

The term *Satellite Geodesy* is more restrictive than the French denomination *Géodésie Spatiale* or the more general expression *Geodetic Space Techniques*. The latter term includes the geodetic observation of the Moon, as well as the use of planets and objects outside the solar system, for instance in radio interferometry. Occasionally the term *Global Geodesy* is used, where global stands for both global measurement techniques and for global applications.

In this book the term *Satellite Geodesy* is employed, because it is in common usage, and because artificial near-Earth satellites are almost exclusively utilized for the observations which are of interest in applied geodesy. Where necessary, other space techniques are dealt with.

1.2 Classification and Basic Concepts of Satellite Geodesy

The importance of artificial satellites in geodesy becomes evident from the following basic considerations.

(1) Satellites can be used as high orbiting *targets*, which are visible over large distances. Following the classical concepts of Earth-bound trigonometric networks, the satellites may be regarded as "fixed" control points within large-scale or global three-dimensional networks. If the satellites are observed simultaneously from different ground stations, it is of no importance that the orbits of artificial satellites are governed by gravitational forces. Only the property that they are targets at high altitudes is used. This purely geometric consideration leads to the *geometrical method* of satellite geodesy. The concept is illustrated in Fig. 1.2. It has been realized in its purest form through the BC4 World Network (see [5.1.5]).

Compared with classical techniques, the main advantage of the satellite methods is that they can bridge large distances, and thus establish geodetic ties between continents and islands. All ground stations belonging to the network

Figure 1.2. Geometrical method; the satellite is a high target

can be determined within a uniform, three-dimensional, global coordinate reference frame. They form a polyhedron which goes around Earth.

As early as 1878 H. Bruns proposed such a concept, later known as the *Cage of Bruns*. Bruns regarded this objective to be one of the basic problems of scientific geodesy. The idea, however, could not be realized with classical methods, and was forgotten. The geometrical method of satellite geodesy is also called the *direct method*,

because the particular position of the satellite enters directly into the solution.

(2) Satellites can be considered to be a *probe* or a *sensor* in the gravity field of Earth. The orbital motion, and the variation of the parameters describing the orbit, are observed in order to draw conclusions about the forces acting. Of particular interest is the relation between the features of the terrestrial gravity field and the resulting deviations of the true satellite orbit from an undisturbed Keplerian motion [3.1.1]. The essential value of the satellite is that it is a moving body within Earth's gravity field. This view leads to the *dynamical method* of satellite geodesy.

The main advantage of satellite observations, when compared with classical techniques, is that the results refer to the planet Earth as a whole, and that they have a global character by nature. Data gaps play a minor role. Among the first spectacular results were a reasonably accurate value of Earth's flattening, and the proof that the figure of Earth is non-symmetrical with respect to the equatorial plane (i.e. the *pear-shape* of Earth, cf. [12.2], Fig. 12.5, p. 517).

In dynamical satellite geodesy orbital arcs of different lengths are considered. When arc lengths between a few minutes and up to several revolutions around Earth are used, we speak of *short arc techniques*; the term for the use of longer arcs, up to around 30 days and more, is *long arc techniques*. The orbits are described in suitable geocentric reference frames. The satellite can thus be considered to be the "bearer of its own coordinates". The geocentric coordinates of the observing ground stations can be derived from the known satellite orbits. This so-called *orbital method* of coordinate determination is illustrated in Fig. 1.3.

The combined determination of gravity field parameters and geocentric coordinates within the domain of dynamical satellite geodesy leads to the general problem of *parameter determination* or *parameter estimation*. This may include the determination of the rotational parameters of Earth (Earth rotation, polar motion) as well as other geodynamical phenomena (cf. [4.1]). The dynamical method of satellite geodesy is

Figure 1.3. Orbital method; the satellite is a sensor in Earth's gravity field

also characterized as the *indirect method*, because the required parameters are determined implicitly from the orbital behavior of the satellites.

The distinction *geometric–dynamic* has, for many years, characterized the development of satellite geodesy. Today, most of the current techniques have to be considered as combinations of both viewpoints.

A further classification of the observation techniques refers to the relation between the observation platform and the target platform. We distinguish the following groups:

(1) *Earth to Space* methods
  − directions from camera observations,
  − satellite laser ranging (SLR),
  − Doppler positioning (TRANSIT, DORIS), and
  − geodetic use of the Global Positioning System (GPS, GLONASS, future GNSS).
(2) *Space to Earth* methods
  − radar altimetry,
  − spaceborne laser, and
  − satellite gradiometry.
(3) *Space to Space* methods
  − satellite-to-satellite tracking (SST).

Earth-bound methods are the most advanced, because the observation process is better under control. With the exception of radar altimetry, the methods mentioned in (2) and (3) are still under development or in their initial operational phase.

1.3 Historical Development of Satellite Geodesy

The proper development of satellite geodesy started with the launch of the first artificial satellite, SPUTNIK-1, on October 4, 1957. The roots of this development can, however, be identified much earlier. If we include the use of the natural Earth satellite, the Moon, then dynamical satellite geodesy has existed since the early 19th century. In 1802, *Laplace* used lunar nodal motion to determine the flattening of Earth to be $f = 1/303$. Other solutions came, for example, from *Hansen* (1864) with $f = 1/296$, *Helmert* (1884) with $f = 1/297.8$, and *Hill* (1884) with $f = 1/297.2$ (see Wolf (1985), Torge (2001)).

The geometrical approach in satellite geodesy also has some forerunners in the *lunar methods*. These methods have undergone comprehensive developments since the beginning of the last century. In this context, the Moon is regarded as a geometric target, where the geocentric coordinates are known from orbital theory. The directions between the observer and the Moon are determined from relative measurements of nearby stars, or from occultation of stars by the Moon. Geocentric coordinates are thereby received. Within the framework of the *International Geophysical Year* 1957/58 a first outcome from a global program was obtained with the *Dual Rate Moon Camera*, developed by Markovitz (1954). The methods of this so-called *Cosmic Geodesy* were treated comprehensively in 1960 by Berroth, Hofmann. They also form a considerable part of the classical book of Mueller (1964) "Introduction to Satellite Geodesy".

Further foundations to satellite geodesy before the year 1957 were given by the work of Väisälä (1946), Brouwer (1959), King-Hele (1958) and O'Keefe (1958). Therefore, it was possible to obtain important geodetic results very soon after the launch of the first rockets and satellites. One of the first outstanding results was the determination of Earth's flattening as $f = 1/298.3$ from observations of EXPLORER-1 and SPUTNIK-2 by O'Keefe (1958), King-Hele, Merson (1958). Some significant

events during the years following 1957 are

<div style="margin-left: 2em;">

1957   Launch of SPUTNIK-1,
1958   Earth's Flattening from Satellite Data ($f = 1/298.3$),
1958   Launch of EXPLORER-1,
1959   Third Zonal Harmonic (*Pear Shape* of Earth),
1959   Theory of the Motion of Artificial Satellites (*Brouwer*),
1960   Launch of TRANSIT-1B,
1960   Launch of ECHO-1,
1960   Theory of Satellite Orbits (*Kaula*),
1962   Launch of ANNA-1B, and
1962   Geodetic Connection between France and Algeria (*IGN*).

</div>

By the year 1964, many basic geodetic problems had been successfully tackled, namely the

− determination of a precise numerical value of Earth's flattening
− determination of the general shape of the global geoid
− determination of connections between the most important geodetic datums (to ±50 m).

With hindsight, the development of satellite geodesy can be divided into several phases of about one decade each.

1. 1958 to around 1970. Development of basic methods for satellite observations, and for the computation and analysis of satellite orbits. This phase is characterized by the optical-photographic determination of directions with cameras. The main results were the determination of the leading harmonic coefficients of the geopotential, and the publication of the first Earth models, for instance the Standard Earth models of the Smithsonian Astrophysical Observatory (SAO SE I to SAO SE III), and the Goddard Earth Models (GEM) of the NASA Goddard Space Flight Center. The only purely geometrical and worldwide satellite network was established by observations with BC4 cameras of the satellite PAGEOS.

2. 1970 to around 1980. Phase of the scientific projects. New observation techniques were developed and refined, in particular laser ranging to satellites and to the Moon, as well as satellite altimetry. The TRANSIT system was used for geodetic Doppler positioning. Refined global geoid and coordinate determinations were carried out, and led to improved Earth models (e.g. GEM 10, GRIM). The increased accuracy of the observations made possible the measurement of geodynamical phenomena (Earth rotation, polar motion, crustal deformation). Doppler surveying was used worldwide for the installation and maintenance of geodetic control networks (e.g. EDOC, DÖDOC, ADOS).

3. 1980 to around 1990. Phase of the operational use of satellite techniques in geodesy, geodynamics, and surveying. Two aspects in particular are remarkable. Satellite methods were increasingly used by the surveying community, replacing conventional methods. This process started with the first results obtained with the NAVSTAR Global Positioning System (GPS) and resulted in completely new perspectives in surveying

and mapping. The second aspect concerned the increased observation accuracy. One outcome was the nearly complete replacement of the classical astrometric techniques for monitoring polar motion and Earth rotation by satellite methods. Projects for the measurement of crustal deformation gave remarkable results worldwide.

4. 1990 to around 2000. Phase of the international and national permanent services. In particular two large international services have evolved. The International Earth Rotation Service IERS, initiated in 1987 and exclusively based on space techniques, provides highly accurate Earth orientation parameters with high temporal resolution, and maintains and constantly refines two basic reference frames. These are the International Celestial Reference Frame ICRF, based on interferometric radio observations, and the International Terrestrial Reference Frame ITRF, based on different space techniques. The International GPS Service IGS, started in 1994 and evolved to be the main source for precise GPS orbits as well as for coordinates and observations from a global set of more than 300 permanently observing reference stations. At the national level permanent services for GPS reference data have been established and are still growing, e.g. CORS in the USA, CACS in Canada and SAPOS in Germany.

5. 2000 onwards. After more than 40 years of satellite geodesy the development of geodetic space techniques is continuing. We have significant improvements in accuracy as well as in temporal and spatial resolution. New fields of application evolve in science and practice. For the first decade of the new millennium development will focus on several aspects:

- launch of dedicated gravity field probes like CHAMP, GRACE, and GOCE for the determination of a high resolution terrestrial gravity field,
- establishment of a next generation Global Navigation Satellite System GNSS with GPS Block IIR and Block IIF satellites and new components like the European Galileo,
- refinement in Earth observation, e.g. with high resolution radar sensors like interferometric SAR on various platforms,
- further establishment of permanent arrays for disaster prevention and environmental monitoring, and
- unification of different geodetic space techniques in mobile integrated geodetic-geodynamic monitoring systems.

1.4   Applications of Satellite Geodesy

The applications of geodetic satellite methods are determined by the achievable accuracy, the necessary effort and expense of equipment and computation, and finally by the observation time and the ease of equipment handling. A very extensive catalogue of applications can be compiled given the current developments in precise methods with real-time or near real-time capabilities.

Starting with the three basic tasks in satellite geodesy introduced in [1.1], we can give a short summary of possible applications:

Global Geodesy
- general shape of Earth's figure and gravity field,
- dimensions of a mean Earth ellipsoid,
- establishment of a global terrestrial reference frame,
- detailed geoid as a reference surface on land and at sea,
- connection between different existing geodetic datums, and
- connection of national datums with a global geodetic datum.

Geodetic Control
- establishment of geodetic control for national networks,
- installation of three-dimensional homogeneous networks,
- analysis and improvement of existing terrestrial networks,
- establishment of geodetic connections between islands or with the mainland,
- densification of existing networks up to short interstation distances.

Geodynamics
- control points for crustal motion,
- permanent arrays for 3D-control in active areas,
- polar motion, Earth rotation, and
- solid Earth tides.

Applied and Plane Geodesy
- detailed plane surveying (land register, urban and rural surveying, geographic information systems (GIS), town planning, boundary demarcation etc.),
- installation of special networks and control for engineering tasks,
- terrestrial control points in photogrammetry and remote sensing,
- position and orientation of airborne sensors like photogrammetric cameras,
- control and position information at different accuracy levels in forestry, agriculture, archaeology, expedition cartography etc.

Navigation and Marine Geodesy
- precise navigation of land-, sea-, and air-vehicles,
- precise positioning for marine mapping, exploration, hydrography, oceanography, marine geology, and geophysics,
- connection and control of tide gauges (unification of height systems).

Related Fields
- position and velocity determination for geophysical observations (gravimetric, magnetic, seismic surveys), also at sea and in the air,
- determination of ice motion in glaciology, Antarctic research, oceanography,
- determination of satellite orbits, and
- tomography of the atmosphere (ionosphere, troposphere).

With more satellite systems becoming operational, there is almost no limit to the possible applications. This aspect will be discussed together with the respective techniques. A summarizing discussion of possible applications is given in chapter [12].

1.5   Structure and Objective of the Book

Satellite geodesy belongs equally to fundamental and applied sciences. Both aspects are dealt with; however, the main emphasis of this book is on the observation methods and on the applications.

Geodetic fundamentals are addressed in chapter [2], in order to help readers from neighboring disciplines. In addition, some useful information is provided concerning fundamental astronomy and signal propagation. The motion of near-Earth satellites, including the main perturbations and the basic methods of orbit determination, are discussed in chapter [3], as far as they are required for an understanding of modern observation techniques and applications.

The increasing observational accuracy requires a corresponding higher accuracy in the determination of orbits. In practice, particularly for today's applications, the user must be capable to assess in each case the required orbital accuracy, and the influence of disturbing effects. This is only possible with a sufficient knowledge of the basic relations in celestial mechanics and perturbation theory. For further studies, fundamental textbooks e.g. Schneider (1981), Taff (1985), or Montenbruck, Gill (2000) are recommended. Special references are given in the relevant sections.

The different observation methods of satellite geodesy are discussed in chapters [4]–[11]. The grouping into currently important observation methods is not without problems, because common aspects have to be taken up in different sections, and because the topical methods develop very quickly. This classification is nevertheless preferred because the user is, in general, working with a particular observation technique, and is looking for all related information. Also a student prefers this type of grouping, because strategies for solving problems can be best studied together with the individual technique. Cross-references are given to avoid unnecessary repetitions.

The possible applications are presented together with the particular observation technique, and illustrated with examples. In chapter [12], a problem-orientated summary of applications is given.

The implications of satellite geodesy affect nearly all parts of geodesy and surveying. Considering the immense amount of related information, it is often only possible to explain the basic principle, and to give the main guidelines. Recommendations for further reading are given where relevant. For example, an exhaustive treatment of satellite motion (chapter [3]), or of the Global Positioning System GPS (chapter [7]) could fill several volumes of textbooks on their own. As far as possible, references are selected from easily accessible literature in the English language. In addition, some basic references are taken from German and French literature.

2 Fundamentals

2.1 Reference Coordinate Systems

Appropriate, well defined and reproducible reference coordinate systems are essential for the description of satellite motion, the modeling of observables, and the representation and interpretation of results. The increasing accuracy of many satellite observation techniques requires a corresponding increase in the accuracy of the reference systems.

Reference coordinate systems in satellite geodesy are global and geocentric by nature, because the satellite motion refers to the center of mass of Earth [3]. Terrestrial measurements are by nature local in character and are usually described in local reference coordinate systems. The relationship between all systems in use must be known with sufficient accuracy. Since the relative position and orientation changes with time, the recording and modeling of the observation time also plays an important role.

It should be noted that the results of different observation methods in satellite geodesy refer to particular reference coordinate systems which are related to the individual methods. These particular systems are not necessarily identical because they may be based on different data and different definitions. Often the relationship between these particular systems is known with an accuracy lower than the accuracy of the individual observation techniques. The establishment of precise transformation formulas between systems is one of the most important tasks in satellite geodesy.

2.1.1 Cartesian Coordinate Systems and Coordinate Transformations

In a Cartesian coordinate system with the axes x, y, z the position of a point P is determined by its position vector

$$\boldsymbol{x}_P = \begin{pmatrix} x_P \\ y_P \\ z_P \end{pmatrix}, \qquad (2.1)$$

where x_P, y_P, z_P are real numbers (Fig. 2.1).

The transformation to a second Cartesian coordinate system with identical origin and the axes x', y', z', which is generated from the first one by a rotation around the z-axis by the angle γ, can be realized through the matrix operation

$$\boldsymbol{x}'_P = \boldsymbol{R}_3(\gamma)\boldsymbol{x}_P \qquad (2.2)$$
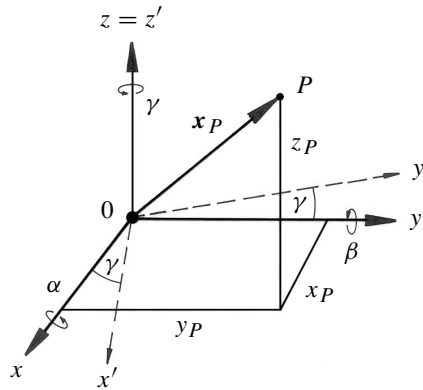
Figure 2.1. Cartesian coordinate system

with

$$R_3(\gamma) = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.3}$$

Equivalent rotations R_1 around the x-axis and R_2 around the y-axis are

$$R_1(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix} \qquad R_2(\beta) = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix}.$$

The representation is valid for a right-handed coordinate system. When viewed towards the origin, a counter-clockwise rotation is positive. Any coordinate transformation can be realized through a combination of rotations. The complete transformation is

$$x_P''' = R_1(\alpha)R_2(\beta)R_3(\gamma)x_P. \tag{2.4}$$

The mathematical properties of rotation matrices are described using linear algebra. The following rules are of importance

(1) Rotation does not change the length of a position vector.

(2) Matrix multiplication is not commutative

$$R_i(\mu)R_j(\nu) \neq R_j(\nu)R_i(\mu). \tag{2.5}$$

(3) Matrix multiplication is associative

$$R_i(R_j R_k) = (R_i R_j)R_k. \tag{2.6}$$

(4) Rotations around the same axis are additive

$$R_i(\mu)R_i(\nu) = R_i(\mu + \nu). \tag{2.7}$$

(5) Inverse and transpose are related by

$$R_i^{-1}(\mu) = R_i^T(\mu) = R_i(-\mu). \tag{2.8}$$

(6) The following relationship also holds

$$(R_i R_j)^{-1} = R_j^{-1}R_i^{-1}. \tag{2.9}$$

The polarity of coordinate axes can be changed with reflectionmatrices

$$S_1 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad S_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad S_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{2.10}$$

Finally, the matrix for a general rotation by the angles α, β, γ is

$$
\boldsymbol{R} = \begin{pmatrix} \cos\beta\cos\gamma & \cos\beta\sin\gamma & -\sin\beta \\ \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\cos\beta \\ \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma & \cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\cos\beta \end{pmatrix}.
$$
(2.11)

The relation between the position vectors in two arbitrarily rotated coordinate systems is then

$$
\boldsymbol{x}_P''' = \boldsymbol{R}\boldsymbol{x}_P; \quad \boldsymbol{x}_P = \boldsymbol{R}^T\boldsymbol{x}_P'''.
$$
(2.12)

In satellite geodesy the rotation angles are often very small, thus allowing the use of the linearized form for $\boldsymbol{R}$. With cos $\alpha \cong 1$ and sin $\alpha \cong \alpha$ (in radians), neglecting higher order terms, it follows that

$$
\boldsymbol{R}(\alpha, \beta, \gamma) = \begin{pmatrix} 1 & \gamma & -\beta \\ -\gamma & 1 & \alpha \\ \beta & -\alpha & 1 \end{pmatrix}.
$$
(2.13)

Although matrix multiplication does not commute (cf. 2.5) the infinitesimal rotation matrix (2.13) does commute.

2.1.2 Reference Coordinate Systems and Frames in Satellite Geodesy

In modern terminology it is distinguished between

- reference systems,
- reference frames, and
- conventional reference systems and frames.

A *reference system* is the complete conceptual definition of how a coordinate system is formed. It defines the origin and the orientation of fundamental planes or axes of the system. It also includes the underlying fundamental mathematical and physical models. A *conventional reference system* is a reference system where all models, numerical constants and algorithms are explicitly specified. A *reference frame* means the practical realization of a reference system through observations. It consists of a set of identifiable fiducial points on the sky (e.g. stars, quasars) or on Earth's surface (e.g. fundamental stations). It is described by a catalogue of precise positions and motions (if measurable) at a specific epoch. In satellite geodesy two fundamental systems are required:

- a space-fixed, *conventional inertial reference system* (CIS) for the description of satellite motion, and
- an Earth-fixed, *conventional terrestrial reference system* (CTS) for the positions of the observation stations and for the description of results from satellite geodesy.

2.1.2.1 Conventional Inertial Systems and Frames

Newton's laws of motion [3.1.2] are only valid in an inertial reference system, i.e. a coordinate system at rest or in a state of uniform rectilinear motion without any acceleration. The theory of motion for artificial satellites is developed with respect to such a system [3].

Space fixed inertial system are usually related to extraterrestrial objects like stars, quasars (extragalactic radio sources), planets, or the Moon. They are therefore also named *celestial reference systems* (CRS). The definition of a CRS can be based on kinematic or dynamic considerations. A *kinematic CRS* is defined by stars or quasars with well known positions and, if measurable, proper motions. A *dynamical CRS* is based on the motion of planets, the Moon, or artificial satellites.

The establishment of conventional celestial reference systems is under the responsibility of the *International Astronomical Union* (IAU). From January 1, 1988, until December 31, 1997, the conventional celestial reference system was based on the orientation of the equator and the equinox for the standard epoch J2000.0 (cf. [2.2.2]), determined from observations of planetary motions in agreement with the IAU (1976) system of astronomical constants as well as related algorithms (cf. Seidelmann (ed.) (1992)). The corresponding reference frame was the *Fifth Fundamental Catalogue* (FK5) (Fricke et al., 1988).

The equatorial system at a given epoch T_0 which has been used in spherical astronomy (Fig. 2.2) for many years yields a rather good approximation to a conventional inertial reference system. The origin of the system is supposed to coincide with the geocenter M. The positive Z-axis is oriented towards the north pole and the positive X-axis to the *First Point of Aries* ♈. The Y-axis completes a right-handed system. Since Earth's center of mass undergoes small accelerations because of the annual motion around the Sun, the term *quasi-inertial system* is also used.
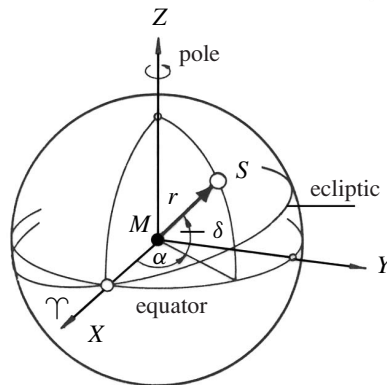
Figure 2.2. Equatorial system in spherical astronomy

The traditional materialization of the above definition for practical purposes is through a catalogue of the positions and proper motions of a given number of fundamental stars. The FK5 is a catalogue of 1535 bright stars, compiled from a large number of meridian observations. The formal uncertainties of the FK5 star positions were about 20 to 30 milliarcseconds at the time of publication (1988). The quality of the FK5 frame is time dependent and is continuously getting worse (de Vegt, 1999; Walter, Sovers, 2000).

Star positions are usually given as spherical coordinates *right ascension* α and *declination* δ. The transformation of spherical coordinates α, δ, r into Cartesian

coordinates X, Y, Z is

$$X = r \cos \delta \cos \alpha, \quad Y = r \cos \delta \sin \alpha, \quad Z = r \sin \delta. \qquad (2.14)$$

The reverse formulas are

$$r = \sqrt{X^2 + Y^2 + Z^2}, \quad \alpha = \arctan \frac{Y}{X}, \quad \delta = \arctan \frac{Z}{\sqrt{X^2 + Y^2}}. \qquad (2.15)$$

In spherical astronomy r is usually defined as the unit radius. We may consider the celestial sphere in Fig. 2.2 as the unit sphere and apply the basic formulas of spherical geometry. Detailed information on spherical astronomy can be found in Green (1985) or in textbooks on geodetic astronomy (e.g. Mackie, 1985; Schödlbauer, 2000).

The accuracy of the celestial reference system, realized through the FK5 catalogue, is by far insufficient for modern needs. A considerable improvement, by several orders of magnitude, was achieved with the astrometric satellite mission *HIPPARCOS* (Kovalevsky et al., 1997), and with extragalactic radio sources (*quasars*) via the technique of *Very Long Baseline Interferometry* (VLBI) which uses radio telescopes [11.1].

In 1991 the IAU decided to establish a new celestial reference system based on a kinematic rather than on a dynamic definition (McCarthy, 2000). The system is called the *International Celestial Reference System* (ICRS) and officially replaced the FK5 fundamental system on January 1, 1998. The axes of the ICRS are no longer fixed to the orientation of the equator and the vernal equinox, but with respect to distant matter in the universe. The system is realized by a celestial reference frame, defined by the precise coordinates of extragalactic objects (mainly quasars) with no measurable transverse motion. The origin of the ICRS is either the barycenter of the solar system, or the geocenter. The ICRS, hence, consists of the

  − *Barycentric Celestial Reference System* (BCRS), and the
  − *Geocentric Celestial Reference System* (GCRS).

The relation between them makes use of general relativity (geodesic precession, Lense-Thirring precession), see McCarthy (2000); Capitaine, et al. (2002).

The *International Celestial Reference Frame* (ICRF) is a catalogue of the adopted positions of 608 extragalactic radio sources observed via the technique of VLBI. 212 of these objects are *defining sources* (Fig. 2.3). They establish the orientation of the ICRS axes. The typical position uncertainty for a defining radio source is about 0.5 milliarcseconds (mas). The resulting accuracy for the orientation of the axes is about 0.02 mas (Ma et al., 1997)

In order to maintain continuity in the conventional celestial reference systems the orientations of the ICRS axes are consistent with the equator and equinox at J 2000.0, as represented by the FK5. Since the accuracy of the FK5 is significantly worse than the new realizations of the ICRS, the ICRS can be regarded as a refinement of the FK5 system.

The *Hipparcos Catalogue* is a realization of the ICRS at optical wavelengths. This catalogue contains 118 218 stars for the epoch J1991.25. The typical uncertainties at catalogue epoch are 1 mas in position and 1 mas / year in proper motion. For the epoch
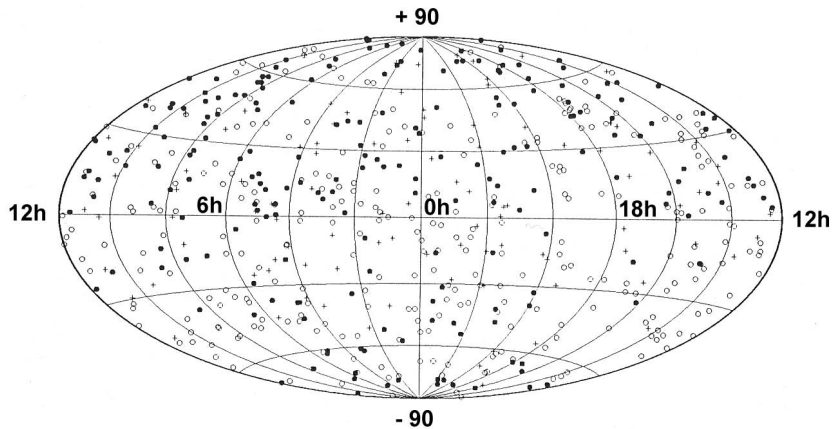
Figure 2.3. International Celestial Reference Frame ICRF; distribution of the 212 best-observed extragalactic sources (after Ma et al. (1998))

J 2000.0 typical Hipparcos star positions can be estimated in the range of 5 to 10 mas (Kovalevsky et al., 1997; Walter, Sovers, 2000).

With forthcoming astrometric space missions like FAME and GAIA (Walter, Sovers, 2000), see [5.3.3], further improvement of the optical realization of the ICRS to the level of 10 microarcseconds (μas) is expected. Also the link between the ICRF based on radio stars and frames at optical wavelengths will be improved.

For more information on conventional inertial reference systems and frames see e.g. Moritz, Mueller (1987, chap. 9), Seidelmann (ed.) (1992, chap. 2), Walter, Sovers (2000), Schödlbauer (2000, chap. 3), Capitaine, et al. (2002) and [12.4.2].

2.1.2.2 Conventional Terrestrial Systems and Frames

A suitable Earth-fixed reference system must be connected in a well defined way to Earth's crust. Such a *Conventional Terrestrial System* (CTS) can be realized through a set of Cartesian coordinates of fundamental stations or markers within a global network.

The origin of an ideal conventional terrestrial reference system should be fixed to the geocenter, including the mass of the oceans and the atmosphere. The z-axis should coincide with the rotational axis of Earth. Since the geocenter and the rotational axis are not directly accessible for observations the ideal system is approximated by conventions. The classical convention for the orientation of axes was based on astronomical observations and has been developed and maintained since 1895 by the *International Latitude Service* (ILS), and since 1962 by the *International Polar Motion Service* (IPMS) (Moritz, Mueller, 1987). It is established through the conventional direction to the mean orientation of the polar axis over the period 1900–1905 (*Conven-*

tional Terrestrial Pole (CTP), also named *Conventional International Origin* (CIO))
and a zero longitude on the equator (*Greenwich Mean Observatory* (GMO)). GMO is
defined through the nominal longitudes of all observatories which contributed to the
former international time service BIH (*Bureau International de l'Heure*).

In 1988 the responsibility for establishing and maintaining both the conventional
celestial and terrestrial reference systems and frames, was shifted to the *International
Earth Rotation Service* (IERS), cf. [12.4.2]. Although the IERS results are based on
modern space techniques like *SLR* [8], *VLBI* [11.1], *GPS* [7], and *Doppler* [6], the
traditional convention has been maintained within the accuracy range of the classical
astronomical techniques in order to provide continuity.

The conventional terrestrial reference system, established and maintained by the
IERS, and nearly exclusively used for today's scientific and practical purposes is the
International Terrestrial Reference System (ITRS); its realization is the *International
Terrestrial Reference Frame* (ITRF). The ITRS is defined as follows (Boucher et al.,
1990; McCarthy, 2000):
  - it is geocentric, the center of mass being defined for the whole Earth, including
    oceans and atmosphere,
  - the length unit is the SI meter; the scale is in context with the relativistic theory
    of gravitation,
  - the orientation of axes is given by the initial BIH orientation at epoch 1984.0,
    and
  - the time evolution of the orientation will create no residual global rotation with
    regard to Earth's crust (no-net-rotation condition).

These specifications correspond with the IUGG resolution no. 2 adopted at the 20th
IUGG General Assembly of Vienna in 1991. The orientation of axes is also called
IERS Reference Pole (IRP) and *IERS Reference Meridian* (IRM).

The realization of the ITRS, the *International Terrestrial Reference Frame* (ITRF)
is formed through Cartesian coordinates and linear velocities of a global set of sites
equipped with various space geodetic observing systems. If geographical coordi-
nates (ellipsoidal latitude, longitude, and height) are required instead of Cartesian
coordinates (X, Y, Z), use of the GRS80 ellipsoid is recommended (cf. [2.1.4]). The
ensemble of coordinates implicitly define the CTP (Z-axis) and the GMO (X-axis).

Nearly every year a new ITRF is realized based on new observations with geodetic
space techniques (e.g. Doppler [6], GPS [7], SLR [8], VLBI [11.1]). The result is
published under the denomination ITRFxx, where xx means the last digits of the
year whose data were used in the formation of the frame. The most recent solution
is ITRF2000 (Fig. 2.4), Altamimi et al. (2001). Each particular ITRF is assembled
by combining sets of results from independent techniques as analyzed by a number
of separate groups. The use of as many different techniques as possible provides a
significant decrease of systematic errors.

The establishment of a terrestrial reference frame is not an easy task because
Earth's crust continuously undergoes various deformations. Since today's geodetic
space techniques provide station coordinates at the 1 cm or subcentimeter level, it is
necessary to model the various deformations at the mm-level. The main influences are
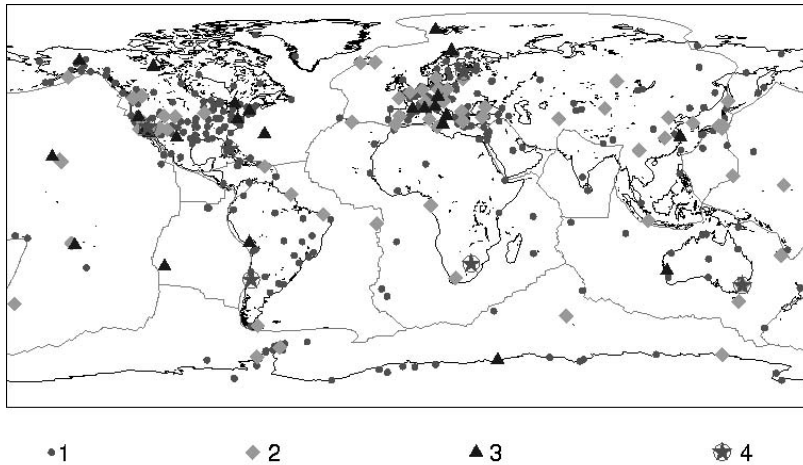
Figure 2.4. International Terrestrial Reference Frame ITRF2000; symbols indicate the number
of different space techniques collocated at the particular site; source: IERS

- − global plate tectonics [12.4.1],
- − solid Earth tides [3.2.3.2],
- − ocean and atmospheric loading effects,
- − polar tides,
- − regional and local effects.

Detailed models and algorithms for these effects are given in the *IERS Conventions*
(McCarthy, 2000).

The largest effect comes from global plate tectonics (cf. Fig. 12.13, p. 529). In
order to maintain the orientation of the coordinate axes stable on the dynamic Earth,
the orientation rate of the ITRF is defined, by convention, so that there is no rotation of
the frame with respect to Earth's lithosphere. In practice, the ITRF orientation rate is
aligned to the plate tectonic model NNR-NUVEL-1A (Argus, Gordon, 1991; DeMets
et al., 1994). This procedure is based on the assumption, that the model fulfills the
condition of *no-net-rotation*, i.e. the integral of model velocities over the entire surface
of Earth becomes zero (Drewes, 1999), see also [12.4.1].

Regional realizations of the ITRS are e.g. the *ETRF89* for Europe and *SIRGAS*
for South America (cf. [12.1]). A particular global realization of a terrestrial reference
system is the *World Geodetic System 1984* (WGS84) [2.1.6].

2.1.2.3   Relationship between CIS and CTS

The transition from the space-fixed equatorial system (CIS) to the conventional terres-
trial system (CTS) is realized through a sequence of rotations that account for
- − precession

    − nutation
    − Earth rotation including polar motion.

These can be described with matrix operations. For a point on the celestial sphere, described through its position vector $\boldsymbol{r}$, we can write

$$\boldsymbol{r}_{\text{CTS}} = \boldsymbol{S}\boldsymbol{N}\boldsymbol{P}\boldsymbol{r}_{\text{CIS}}. \tag{2.16}$$

The elements of the rotation matrices must be known with sufficient accuracy for each observation epoch. These rotations are now considered in more detail.

(a) *Precession and Nutation*

Earth's axis of rotation and its equatorial plane are not fixed in space, but rotate with respect to an inertial system. This results from the gravitational attraction of the Moon and the Sun on the equatorial bulge of Earth. The total motion is composed of a mean secular component (precession) and a periodic component (nutation) (Fig. 2.5).
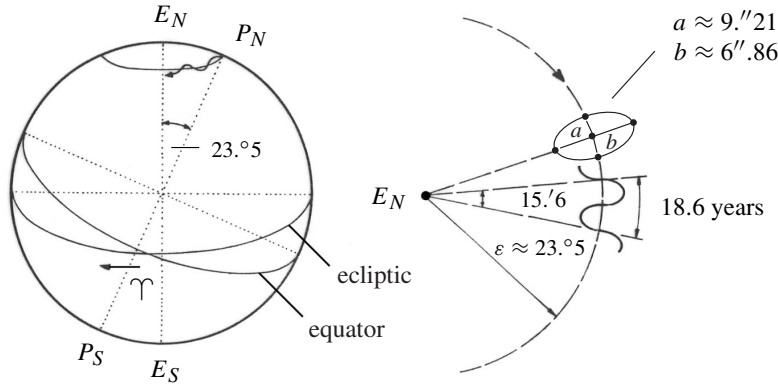
Figure 2.5. Precession and nutation; Earth's rotation axis $P_S P_N$ describes a conic about the ecliptic poles E_S, E_N

    The position and orientation of the equatorial plane and the first point of Aries, ♈, is called *mean equator* and *mean equinox*, respectively, when only the influence of precession is considered. When nutation is taken into account, they are called *true equator* and *true equinox*. The respective star coordinates are termed *mean positions* or *true positions*. Mean positions can be transformed from the reference epoch t_0 (J2000.0) to the required observation epoch t using the precession matrix

$$\boldsymbol{P} = \boldsymbol{R}_3(-z)\,\boldsymbol{R}_2(\theta)\,\boldsymbol{R}_3(-\zeta) \tag{2.17}$$

with three rotations (2.3) by the angles $-z$, θ, $-\zeta$

$$
\begin{aligned}
z &= 0°.6406161\,T + 0°.0003041\,T^2 + 0°.0000051\,T^3 \\
\theta &= 0°.5567530\,T - 0°.0001185\,T^2 - 0°.0000116\,T^3 \\
\zeta &= 0°.6406161\,T + 0°.0000839\,T^2 + 0°.0000050\,T^3.
\end{aligned} \tag{2.18}
$$

$T = (t - t_0)$ is counted in Julian centuries of 36525 days.

The transformation from the mean equator and equinox to the instantaneous *true equator and equinox* for a given observation epoch is performed with the nutation matrix

$$N = R_1(-\varepsilon - \Delta\varepsilon)\, R_3(-\Delta\psi)\, R_1(\varepsilon) \qquad (2.19)$$

where

$$
\begin{aligned}
&\varepsilon \qquad \text{obliquity of the ecliptic,} \\
&\Delta\varepsilon \qquad \text{nutation in obliquity,} \\
&\Delta\psi \qquad \text{nutation in longitude (counted in the ecliptic),}
\end{aligned}
$$

and

$$\varepsilon = 23°26'21''.448 - 46''.815\, T - 0''.00059\, T^2 + 0''.001813\, T^3. \qquad (2.20)$$

In 1980 the International Astronomical Union (IAU) adopted a nutation theory (Wahr, 1981) based on an elastic Earth model. $\Delta\psi$ is computed using a series expansion involving 106 coefficients and $\Delta\varepsilon$ using one of 64 coefficients. The principal terms are

$$\Delta\psi = -17''.1996 \sin\Omega - 1''.3187 \sin(2F - 2D + 2\Omega) - 0''.2274 \sin(2F + 2\Omega) \qquad (2.21)$$

$$\Delta\varepsilon = 9''.2025 \cos\Omega + 0''.5736 \cos(2F - 2D + 2\Omega) + 0''.0977 \cos(2F + 2\Omega) \qquad (2.22)$$

with

$$
\begin{aligned}
&\Omega \qquad \text{mean ecliptic longitude of the lunar ascending node,} \\
&D \qquad \text{mean elongation of the Moon from the Sun,}
\end{aligned}
$$

$$F = \lambda_M - \Omega$$

with λ_M the mean ecliptic longitude of the Moon. By applying the transformations (2.17) and (2.19) we obtain *true coordinates*

$$r_T = (X_T, Y_T, Z_T) \qquad (2.23)$$

in the instantaneous true equatorial system.

More details can be found in Seidelmann (ed.) (1992) and McCarthy (2000). The IAU decided at its 24th General Assembly in 2000 to replace the IAU 1976 Precession Model and the IAU 1980 Theory of Nutation by the *Precession-Nutation Model IAU 2000*, beginning on January 1, 2003. Two versions of the model exist (Capitaine, et al., 2002):

The *IAU 2000A model* contains 678 luni-solar terms and 687 planetary terms and provides directions of the celestial pole in the geocentric celestial reference system (GCRS) with an accuracy of 0.2 mas. The abridged model *IAU 2000B* includes 80

luni-solar terms and a planetary bias. The difference between both models is not greater than 1 mas after about 50 years.

(b) *Earth Rotation and Polar Motion*

For the transition from an instantaneous space-fixed equatorial system to a conventional terrestrial reference system we need three further parameters. They are called *Earth Rotation Parameters* (ERP) or *Earth Orientation Parameters* (EOP), namely

GAST   the Greenwich apparent sidereal time
x_p, y_p the pole coordinates.

GAST is also expressed as the difference UT1-UTC, cf. [2.2.2]. Unlike precession (2.17) and nutation (2.19), Earth rotation parameters cannot be described through theory but must be determined through actual observations by an international time and latitude service. Since the beginning of the last century until about 1980, this service was based mainly on astronomical observations (see [12.4.2]). On January 1, 1988 the *International Earth Rotation Service* (IERS) (Boucher et al., 1988) took over this task. The principle observation techniques now used are laser ranging to satellites and to the Moon [8.5.5] and Very Long Baseline Interferometry (VLBI) [11.1.2].

Fig. 2.6 shows the geometric situation for the transformation. The Earth-fixed system is realized through the conventional orientation of a Cartesian $(X, Y, Z)_{CT}$ system. The Z_{CT}-axis is directed toward the conventional terrestrial pole CTP, and the X_{CT}-axis toward the mean Greenwich meridian. The relative position of the instantaneous true pole with respect to the conventional terrestrial pole CTP is usually described through the pole coordinates x_p, y_p (e.g. Mueller, 1969; Schödlbauer, 2000).
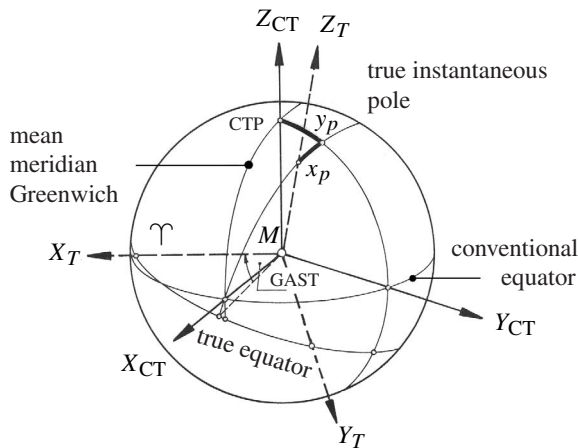
Figure 2.6. True instantaneous and mean conventional terrestrial system

The relative orientation of the X_{CT}-axis depends directly on Earth rotation and is determined through the apparent ($=$ true) *Greenwich Sidereal Time* GAST (cf. [2.2.2]). The symbol θ is often used to denote GAST. The matrix which transforms the instantaneous space-fixed system to the conventional terrestrial system is

$$\boldsymbol{S} = \boldsymbol{R}_2(-x_p)\boldsymbol{R}_1(-y_p)\boldsymbol{R}_3(\text{GAST}) \tag{2.24}$$

where

$$\boldsymbol{R}_3(\text{GAST}) = \begin{pmatrix} \cos(\text{GAST}) & \sin(\text{GAST}) & 0 \\ -\sin(\text{GAST}) & \cos(\text{GAST}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.25}$$

and, because of small angles,

$$\boldsymbol{R}_2(-x_p)\boldsymbol{R}_1(-y_p) = \begin{pmatrix} 1 & 0 & x_p \\ 0 & 1 & 0 \\ -x_p & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -y_p \\ 0 & y_p & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & x_p \\ 0 & 1 & -y_p \\ -x_p & y_p & 1 \end{pmatrix}. \tag{2.26}$$

For most practical purposes, the pole of the instantaneous true space-fixed equatorial system can be considered to be identical to the so-called *Celestial Ephemeris Pole* (CEP). The CEP is defined to be the reference pole for the computation of polar motion and nutation and is free of the quasi diurnal nutation terms with respect to Earth's crust and inertial space (Seidelmann (ed.), 1992). The observed differences between the CEP and the conventional precession-nutation model are named *celestial pole offsets $d\psi$, $d\varepsilon$*. They reach a few milliarcseconds and are published by the IERS, see [11.1.2].

The CEP will be replaced by the *Celestial Intermediate Pole* (CIP) along with the introduction of the IAU 2000 precession-nutation model. Accordingly, for the rigorous definition of sidereal rotation of Earth, based on the early concept of the "non-rotating origin" (Guinot, 1979) the *Terrestrial Ephemeris Origin* (TEO), and the *Celestial Ephemeris Origin* (CEO), defined on the equator of the CIP, will be introduced. This implies that UT1 be linearly proportional to the Earth rotation angle. The CIP coincides with the CEP in the low-frequency domain (periods larger than two days). The reason for the adoption of the CIP is to clarify the difference between nutation and polar motion at high frequencies (periods less than two days). For details see e.g. Capitaine et al. (2000); Capitaine, et al. (2002).

The IERS will introduce the new system in 2003. The old system, however, will continue to be used, and the IERS will continue to provide all necessary data until further notice.

2.1.3 Reference Coordinate Systems in the Gravity Field of Earth

Terrestrial geodetic observations, with the exception of the slant ranges s, are related to the local gravity vector $\boldsymbol{g}$. They can therefore easily be described in a local reference coordinate system which is tied to the direction of the plumb line, $\boldsymbol{n}$, at the observation

point, P. The orientation of the vector, $\boldsymbol{n}$, is usually determined from astronomical observations and described as
  — the astronomical latitude Φ and
  — the astronomical longitude Λ

$$\boldsymbol{n} = \begin{pmatrix} \cos\Phi\cos\Lambda \\ \cos\Phi\sin\Lambda \\ \sin\Phi \end{pmatrix}. \tag{2.27}$$

The relationship between the *local astronomical system*, defined as

  — origin at the observation point P,
  — Z'-axis directed to the astronomical zenith,
  — X'-axis directed to the north (astronomical meridian),
  — Y'-axis directed to the east,

and the global conventional terrestrial system (CTS) is described in Fig. 2.7 (Torge (1980, 2001)). The location of a point, P_i, within the local astronomical system is derived from terrestrial observations

Figure 2.7.   Local astronomical system and global terrestrial system

  — astronomical azimuth A,
  — horizontal directions (azimuth differences),
  — slant ranges s, and
  — zenith angles z,

and may be written as

$$X' = \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = s \begin{pmatrix} \cos A \sin z \\ \sin A \sin z \\ \cos z \end{pmatrix}. \tag{2.28}$$

Observed coordinate differences may be transformed from the local system into the global system (CTS) using

$$\Delta X = A \Delta X' \tag{2.29}$$

with

$$A = R_3(180^0 - \Lambda)R_2(90^0 - \Phi)S_2.$$

The matrix S_2 changes the orientation of the Y-axis and converts a left-handed into a right-handed coordinate system. The explicit form of A is

$$A = \begin{pmatrix} -\sin\Phi\cos\Lambda & -\sin\Lambda & \cos\Phi\cos\Lambda \\ -\sin\Phi\sin\Lambda & \cos\Lambda & \cos\Phi\sin\Lambda \\ \cos\Phi & 0 & \sin\Phi \end{pmatrix}. \tag{2.30}$$

The inverse formula reads (Torge, 2001)

$$\Delta X' = A^{-1} \Delta X = A^T \Delta X \tag{2.31}$$

with

$$A^{-1} = \begin{pmatrix} -\sin \Phi \cos \Lambda & -\sin \Phi \sin \Lambda & \cos \Phi \\ -\sin \Lambda & \cos \Lambda & 0 \\ \cos \Phi \cos \Lambda & \cos \Phi \sin \Lambda & \sin \Phi \end{pmatrix}.$$

The formulas (2.29) and (2.31) are used in the combination of results from local terrestrial observations and from satellite techniques, either in the global Cartesian system or in the local astronomical system.

2.1.4 Ellipsoidal Reference Coordinate Systems

For most practical applications ellipsoidal coordinate systems are preferred because they closely approximate Earth's surface, and they facilitate a separation of horizontal position and height. Usually a rotational ellipsoid is selected which is flattened at the poles and which is created by rotating the meridian ellipse about its minor axis b. The geometric parameters are

$$\text{semi-major axis} \quad a \quad \text{and} \quad \text{flattening} \quad f = \frac{a-b}{a}. \tag{2.32}$$

Alternatively the *first numerical eccentricity e* is used

$$e^2 = \frac{a^2 - b^2}{a^2}. \tag{2.33}$$

Further suitable relations between these quantities are

$$e^2 = 2f - f^2;$$
$$1 - e^2 = (1-f)^2. \tag{2.34}$$

A best possible approximation to the figure of the whole Earth is a *global ellipsoidal system* (Fig. 2.8). The *ellipsoidal geographic coordinates* are

    φ ellipsoidal latitude

    λ ellipsoidal longitude

    h ellipsoidal height.
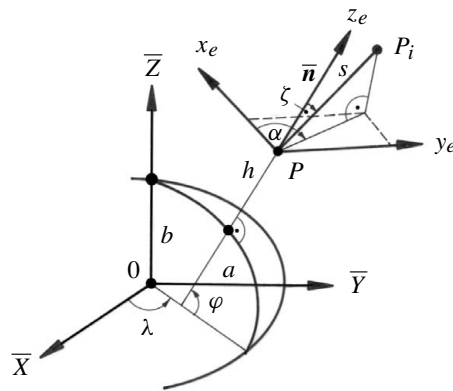
Figure 2.8. Global and local ellipsoidalsystems

A concentric Cartesian coordinate system $\overline{X}, \overline{Y}, \overline{Z}$ can be defined within the ellipsoid

- origin at the center 0 of the ellipsoid,
- $\overline{Z}$-axis directed to the northern ellipsoidal pole (along the minor axis),
- $\overline{X}$-axis directed to the ellipsoidal zero meridian, and
- $\overline{Y}$-axis completing a right-handed system.

The transformation equation between the geographic ellipsoidal coordinates φ, λ, h and the Cartesian coordinates $\overline{X}$, $\overline{Y}$, $\overline{Z}$ is:

$$\overline{X} = \begin{pmatrix} \overline{X} \\ \overline{Y} \\ \overline{Z} \end{pmatrix} = \begin{pmatrix} (\overline{N} + h) \cos \varphi \cos \lambda \\ (\overline{N} + h) \cos \varphi \sin \lambda \\ ((1 - e^2)\,\overline{N} + h) \sin \varphi \end{pmatrix}, \tag{2.35}$$

$\overline{N}$ is the *radius of curvature in the prime vertical*:

$$\overline{N} = \frac{a}{\sqrt{1 - e^2 \sin^2 \varphi}} = \frac{a}{\sqrt{1 - f(2 - f) \sin^2 \varphi}}. \tag{2.36}$$

One solution of the inverse problem is, cf. Heiskanen, Moritz (1967, p. 183), Ehlert (1991); Torge (2001):

$$h = \frac{\sqrt{\overline{X}^2 + \overline{Y}^2}}{\cos \varphi} - \overline{N} \tag{2.37}$$

$$\varphi = \arctan \frac{\overline{Z}}{\sqrt{\overline{X}^2 + \overline{Y}^2}} \left(1 - e^2 \frac{\overline{N}}{\overline{N} + h} \right)^{-1}, \quad \lambda = \arctan \frac{\overline{Y}}{\overline{X}}.$$

The equation can only be solved by iteration, because φ and h are also present on the right-hand side of (2.37). The convergence is fast since $h \ll N$. Another efficient solution which also works for large heights is by Sjöberg (1999). The transformation (2.37) has become a standard routine in modern satellite positioning.

A *local ellipsoidal system*, tied to the ellipsoidal vertical $\overline{n}$ at the observation point, P, can be defined as (Fig. 2.8)

$$\overline{n} = \begin{pmatrix} \cos \varphi \cos \lambda \\ \cos \varphi \sin \lambda \\ \sin \varphi \end{pmatrix} \tag{2.38}$$

with the specifications:

- origin at the observation point, P,
- z_e-axis in the direction of the ellipsoidal vertical,
- x_e-axis directed to the north (geodetic meridian), and
- y_e-axis directed to the east, completing a left-handed system.

The location of a second point, P_i, in the local ellipsoidal system, can be determined via the following quantities (polar coordinates):

- slant ranges s,
- ellipsoidal azimuths α,
- ellipsoidal directions or horizontal angles $\Delta\alpha$, and
- ellipsoidal zenith angles ζ.

The local spherical coordinates s, α, ζ are related to the local Cartesian coordinates

$$\boldsymbol{x}_e = \begin{pmatrix} x_e \\ y_e \\ z_e \end{pmatrix} = s \begin{pmatrix} \cos\alpha\sin\zeta \\ \sin\alpha\sin\zeta \\ \cos\zeta \end{pmatrix}. \tag{2.39}$$

For the transformation of coordinate differences from the local to the global ellipsoidal system we obtain the relation

$$\Delta\overline{\boldsymbol{X}} = \boldsymbol{R}_3(180° - \lambda)\boldsymbol{R}_2(90° - \varphi)\boldsymbol{S}_2\Delta\boldsymbol{x}_e = \boldsymbol{A}\Delta\boldsymbol{x}_e \tag{2.40}$$

where

$$\boldsymbol{A} = \begin{pmatrix} -\sin\varphi\cos\lambda & -\sin\lambda & \cos\varphi\cos\lambda \\ -\sin\varphi\sin\lambda & \cos\lambda & \cos\varphi\sin\lambda \\ \cos\varphi & 0 & \sin\varphi \end{pmatrix}. \tag{2.41}$$

The inverse formula is the same as (2.31).

For further reading on ellipsoidal computation and on the use of plane coordinate systems (e.g. UTM) see textbooks on mathematical geodesy, such as Großmann (1976); Vaníček, Krakiwsky (1986); Torge (2001).

2.1.5 Ellipsoid, Geoid and Geodetic Datum

The physical shape of the real Earth is closely approximated by the mathematical surface of the rotational ellipsoid. The ellipsoidal surface is smooth and convenient for mathematical operations. This is why the ellipsoid is widely used as the *reference surface for horizontal coordinates* in geodetic networks.

On the other hand, the ellipsoid is much less suitable as a *reference surface for vertical coordinates* (heights). Instead, the *geoid* is used. It is defined as that level surface of the gravity field which best fits the mean sea level, and may extend inside the solid body of Earth (Torge, 2001). The relationship between geoid and ellipsoid is illustrated with Fig. 2.9.

The vertical separation between the geoid and a particular reference ellipsoid is called *geoid undulation N*. The numerical values of the undulations evidently depend on the particular ellipsoid used. For a global reference ellipsoid they can reach up to 100 m. The geometrical relation between the geoid undulation, N, the ellipsoidal height, h, and the orthometric height, H (obtained from spirit levelling), is approximately (see also Fig. 7.82, p. 366)

$$h = N + H. \tag{2.42}$$
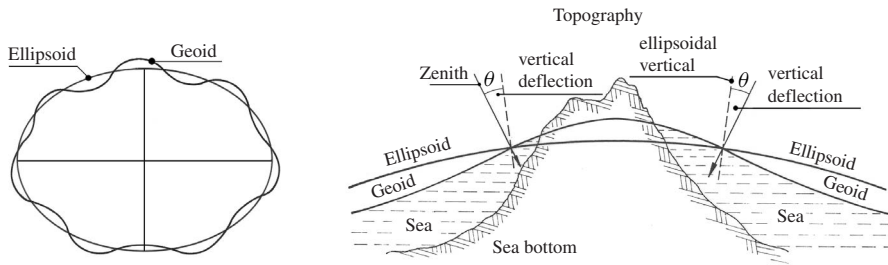
Figure 2.9. Relationship between ellipsoid and geoid

It is evident that the geoid undulation N must be known when observations from satellite geodesy (leading to ellipsoidal heights) and from terrestrial geodesy (leading to heights defined in the gravity field) are used in a combined adjustment. This aspect will be treated in more detail in chapter [7.6.2.3].

Note that in many countries *normal heights*, H^N, are used rather than orthometric heights, H, in order to avoid any hypothesis on the density and the distribution of topographic masses. The reference surface for the normal heights is the *quasi-geoid* which is close to the geoid. The vertical separation between ellipsoid and quasi-geoid is called the *height anomaly* ζ. The relation (2.42) for normal heights hence is written

$$h = \zeta + H^N. \tag{2.43}$$

Geoid and quasi-geoid only deviate on the mm to cm-order at low elevations and may reach one-meter deviation in the high mountains (Torge, 2001, p. 82).

The angle θ between the direction of the ellipsoidal normal and of the plumb line at point P is called the *deflection of the vertical*, cf. Fig. 2.9. Usually, θ is divided into two components and defined as

$$\xi = \Phi - \varphi \quad \text{and} \quad \eta = (\Lambda - \lambda)\cos\varphi. \tag{2.44}$$

Φ, Λ are obtained from astronomical and φ, λ from geodetic observations. The deflections of the vertical are hence also named *astrogeodetic deflections*. The parameters defining a local ellipsoid were usually estimated in such a way that the distribution of known deflections of the vertical fulfilled some minimum condition in the adjustment process. The related local ellipsoids are hence *best fitting ellipsoids*.

A global ellipsoidal system is related to a reference ellipsoid that best fits the figure of Earth as a whole. The origin of the ellipsoid is supposed to coincide with Earth's center of mass. Furthermore, the directions of the ellipsoidal axes are defined so as to be parallel to the conventional terrestrial reference system (CTS). The set of parameters that describe the relationship between a particular local ellipsoid and a global geodetic reference system is called a *geodetic datum*. A geodetic datum is defined by a set of

at least five parameters:

a	semi-major axis of the reference ellipsoid,
f	flattening, and
$\Delta X, \ \Delta Y, \ \Delta Z$	coordinates of the ellipsoid origin with respect to the geocenter.

For $\Delta X = \Delta Y = \Delta Z = 0$ the geodetic datum is called an *absolute datum*. The *Geodetic Reference System 1980* (GRS80), adopted by the IUGG General Assembly in Canberra (1979), belongs to this group

$$a = 6\,378\,137\,\text{m}$$
$$f = \frac{1}{298.2572}. \tag{2.45}$$

Further constants of the GRS80 are (Moritz, 2000) the geocentric gravitational constant of Earth (including the atmosphere)

$$GM = 398\,600.5\,\text{km}^3\,\text{s}^{-2},$$

the dynamical form factor of Earth (related to f)

$$J_2 = 0.00108263$$

and the mean angular velocity of Earth

$$\omega = 7.292115 \times 10^{-5}\text{rad s}^{-1}.$$

For a large number of particular local reference systems the so-called *datum shift constants* or *datum shift parameters* $\Delta X, \Delta Y, \Delta Z$ can be derived from satellite observations. They represent, however, only a mean position of the particular local system with respect to the geocentric system (cf. [12.1]).

In practice, the establishment of a local geodetic datum did not always achieve the objective of axes parallel to the CTS. This is in particular the case for many existing national datums. For this reason a transition from one ellipsoidal reference coordinate system to another also includes rotations. Usually such datum transformations are established between Cartesian systems. When a formulation with ellipsoidal coordinates is required the equations (2.35) to (2.37) have also to be applied.

A complete datum transformation equation between two Cartesian systems requires seven parameters (Fig. 2.10):

3 translations $\Delta X, \Delta Y, \Delta Z$,   3 rotations $\varepsilon_x, \varepsilon_y, \varepsilon_z$ and   1 scale factor m.

In most cases the rotation angles are very small, hence (cf. (2.13))

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_2 = \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix}_{1,2} + (1+m) \begin{pmatrix} 1 & \varepsilon_z & -\varepsilon_y \\ -\varepsilon_z & 1 & \varepsilon_x \\ \varepsilon_y & -\varepsilon_x & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_1. \tag{2.46}$$

For limited areas only three local or regional translation parameters may be sufficient (cf. [12.1]).

   The number of datum parameters increases to nine when the parameters of the ellipsoid have to be considered. The number is further increased when specific rotations or deformations are allowed for parts of the terrestrial network (cf. [7.6.2.1]), and when the datum information is derived from satellite orbits. In the latter case the potential coefficients of Earth's gravity field as well as some fundamental constants, like Earth rotation, velocity of light and geocentric

Figure 2.10. Datum transformation between two Cartesian systems

gravitational constant, form parts of the datum definition. One example of the latter group is the *World Geodetic System* WGS 84.

2.1.6   World Geodetic System 1984 (WGS 84)

WGS has been developed by the U.S. Department of Defense (DoD) since about 1960 in order to define and establish a geocentric terrestrial reference system. WGS 60 was followed by WGS 66, then by WGS 72, and finally by WGS 84. Each realization of the reference system incorporated more data, better computational techniques, a better knowledge of Earth, and improved accuracy (Malys, Slater, 1994; Slater, Malys, 1997; Merrigan et al., 2002). WGS 72 and WGS 84 have been used to compute the operational broadcast ephemeris of Transit Doppler [6.2] and GPS [7.1.5] satellites. As a consequence, coordinates derived from the broadcast ephemeris with Transit or GPS refer to WGS. This is the main reason for the high acceptance of WGS 84 as a primary reference coordinate system. The major parameters of WGS 72 and of the latest version of WGS 84 are given in Table 2.1 (cf. NIMA, 2000).

Table 2.1. Main parameters of WGS 72 and WGS 84

Parameter	Name	WGS 72	WGS 84
semi-major axis	a	6 378 135 m	6 378 137 m
flattening	f	1/298.26	1/298.257223563
angular velocity	ω	7.292115147 $\times 10^{-5}$ rad s^{-1}	7.292115 $\times 10^{-5}$ rad s^{-1}
geocentric gravitational constant	GM	398 600.8 km^3 s^{-2}	398 600.4418 km^3 s^{-2}
2nd zonal harmonic	$\overline{C}_{2,0}$	-484.1605×10^{-6}	$-484.16685 \times 10^{-6}$

WGS 84 practically coincides with the Geodetic Reference System 1980. The associated gravity field is the *Earth Gravitational Model 1996* (EGM96), complete to degree and order 360 (Lemoine et al., 1998), cf. [3.2.2.1], [12.2.2].

The WGS 84 coordinate system is a Conventional Terrestrial Reference System (CTRS) and follows the criteria outlined in the IERS Conventions (cf. [2.1.2.2], McCarthy (2000)). The Z-axis is in the direction of the *IERS Reference Pole* (IRP); the X-axis is in the intersection of the *IERS Reference Meridian* (IRM) and the plane passing through the origin and normal to the Z-axis. The Y-axis completes a right-handed Earth-centered orthogonal coordinate system (Fig. 2.11).

The realizations of the WGS 84 after several refinements now coincide with the ITRF at the level of 1 cm (Merrigan et al., 2002). Hence, for most practical purposes WGS 84 and ITRF can be considered as identical.
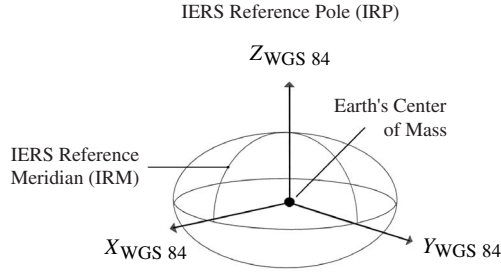
Figure 2.11. World Geodetic System 1984

The practical realization happens through a set of station coordinates. These are the Air Force and NIMA permanent GPS tracking stations (cf. [7.1.3]). Until 1994 Transit Doppler derived WGS 84 coordinates [6.2] were used for the realization of the WGS 84 reference frame. In 1994, in 1996 and again in 2002 a refined set of station coordinates based on GPS observations were introduced into the realization of the WGS 84 frame. These frames are designated as WGS 84 (G730), WGS 84 (G873) and WGS 84 (G1150). The letter G stands for GPS and the number indicates the GPS week number [7.1.5.3] when these coordinates were implemented into the ephemeris estimation process. It should be noted that the original definition of the WGS 84 did not change along with these refinements in the realization of the WGS 84 frames.

The responsible organization for the establishment and maintenance of the WGS 84 is the *National Imagery and Mapping Agency* NIMA (formerly the Defense Mapping Agency DMA). A comprehensive report (NIMA, 2000) containing all necessary information on WGS 84, including its definition and relationship with local geodetic systems, is freely available.

For coordinate transformations between WGS 72 and WGS 84 the following equations can be used

$$\begin{pmatrix} \varphi \\ \lambda \\ h \end{pmatrix}_{\text{WGS 84}} = \begin{pmatrix} \varphi \\ \lambda \\ h \end{pmatrix}_{\text{WGS 72}} + \begin{pmatrix} \Delta\varphi \\ \Delta\lambda \\ \Delta h \end{pmatrix} \tag{2.47}$$

$\Delta\varphi \; [''] = (4.5 \cos\varphi)\,(a \sin 1'') + (\Delta f \sin 1'') + (\Delta f \sin 2\varphi / \sin 1'')$

$\Delta\lambda \; [''] = 0.554$

$\Delta h \; [\text{m}] = 4.5 \sin\varphi + a\,\Delta f \sin^2\varphi - \Delta a + \Delta r$

with

$$\Delta f = 0.3121057 \times 10^{-7}, \quad a = 6\,378\,135 \text{ m}, \quad \Delta a = 2.0 \text{ m}, \quad \Delta r = 1.4 \text{ m}.$$

The latitude is counted positive to the north, and the longitude positive to the east.

Absolute geocentric coordinates of an isolated single observation station which have been derived from satellite observations with TRANSIT [6.6.1] or GPS [7.6.2] have usually standard deviations in the order of several meters or even tens of meters. It is evident that a datum transformation with (2.47) cannot improve the coordinate accuracy. The worth of a general transformation formula like (2.47) must not be over-estimated. For more details on datum transformation see [6.6.1], [7.6.2.1], [12.1], and the cited document NIMA (2000).

2.1.7 Three-dimensional Eccentricity Computation

When satellite observations are made from eccentric station marks it is often necessary to provide the eccentricity in a geocentric Cartesian coordinate system. The observed or previously available ellipsoidal elements $\Delta\varphi$, $\Delta\lambda$, Δh have to be transformed into Cartesian elements ΔX, ΔY, ΔZ. Starting with approximate ellipsoidal coordinates φ, λ, h of the station center, and the ellipsoid parameters a and e, it follows from (2.35) that

$$dX = -(\overline{M} + h)\sin\varphi\cos\lambda\,d\varphi - (\overline{N} + h)\cos\varphi\sin\lambda\,d\lambda + \cos\varphi\cos\lambda\,dh$$
$$dY = -(\overline{M} + h)\sin\varphi\sin\lambda\,d\varphi + (\overline{N} + h)\cos\varphi\cos\lambda\,d\lambda + \cos\varphi\sin\lambda\,dh \quad (2.48)$$
$$dZ = (\overline{M} + h)\cos\varphi\,d\varphi + \sin\varphi\,dh$$

where

$$\overline{M} = \frac{a(1 - e^2)}{(1 - e^2\sin^2\varphi)^{\frac{3}{2}}} \quad \text{and} \quad \overline{N} = \frac{a}{(1 - e^2\sin^2\varphi)^{\frac{1}{2}}}$$

are the meridian radius of curvature and the radius of curvature in the prime vertical, respectively. For practical purposes the differential expressions can be replaced by small finite quantities. The algorithm is only valid for small eccentricities. The inverse formula is (White, 1980; Ehlert, 1991)

$$\begin{pmatrix} d\varphi \\ d\lambda \\ dh \end{pmatrix} = \begin{pmatrix} \dfrac{-\sin\varphi\cos\lambda}{\overline{M} + h} & \dfrac{-\sin\varphi\sin\lambda}{\overline{M} + h} & \dfrac{\cos\varphi}{\overline{M} + h} \\ \dfrac{-\sin\lambda}{(\overline{N} + h)\cos\varphi} & \dfrac{\cos\lambda}{(\overline{N} + h)\cos\varphi} & 0 \\ \cos\varphi\cos\lambda & \cos\varphi\sin\lambda & \sin\varphi \end{pmatrix} \cdot \begin{pmatrix} dX \\ dY \\ dZ \end{pmatrix}. \quad (2.49)$$

2.2 Time

2.2.1 Basic Considerations

Three basic groups of time scales are of importance in satellite geodesy:

(1) The time-dependent orientation of Earth with respect to the inertial space is required in order to relate the Earth-based observations to a space-fixed reference frame. The appropriate time scale is connected with the diurnal rotation of Earth, and is called *Sidereal Time* or *Universal Time*.

(2) For the description of the satellite motion we need a strictly uniform time measure which can be used as the independent variable in the equations of motion. An appropriate time scale can be derived from the orbital motion of celestial bodies around the Sun. It is called *Ephemeris Time*, *Dynamical Time*, or *Terrestrial Time*.

(3) The precise measurement of signal travel times, e.g. in satellite laser ranging, requires a uniform and easily accessible time scale with high resolution. The appropriate measure is related to phenomena in nuclear physics and is called *Atomic Time*.

All these time scales are based on the observation of uniform and repetitive astronomical or physical phenomena. The time interval between two consecutive phenomena forms the scale measure of the particular time scale. A certain multiple or fraction of the scale measure is called the *time unit*. In general, the *second* (s) is used as the basic time unit. Larger time units, such as days or years, are derived from the second.

Within the time scale a starting point or origin has to be fixed. This may be achieved through a certain astronomical event, such as the particular position of a star, or the meridian transit of a particular celestial object.

The instant of the occurrence of some phenomena or observations can be related to a certain reading of the particular time scale, and gives the *datation* of the event. In astronomy such an event is called the *epoch* of the observation. With respect to the particular time scale the epoch determination reflects an *absolute time measurement*. For many purposes, e.g. for the determination of signal travel times, a *relative time measurement*, i.e. the determination of the time interval between two epochs, is sufficient. In many cases the relative time measurement can be done much more accurately than the absolute time measurement. In satellite geodesy the datation of an event is often called *time-tag* or *time-tagging*, e.g. when the instant of transmission or reception of a signal is considered.

Strictly speaking, we have to distinguish between the ideal conception of a time scale and the practical realization through observations. This becomes particularly evident with the atomic time, when we compare the definition of the atomic time second with its practical realization through a group of individual atomic clocks. A time scale may be regarded as an approximation to the particular time concept. In the following we will not use this distinction. For further reading see e.g. Seidelmann et al.

(1992), Guinot (1995). A useful source on time in relation to GPS is Langley (1991b), the *GPS World* Supplement on Precise Timing (December 1998) or Lombardi et al. (2001) .

In order to meet the various requirements, stemming from science and technology, the relationship between the different time scales have to be established with the highest possible accuracy. Fig. 2.12 illustrates how timing errors in satellite geodesy are related to a position error of 1 cm:

1 cm motion of a point on the equator caused by Earth's rotation corresponds to about 2×10^{-5} s,

1 cm motion of a near-Earth satellite in the orbit corresponds to about 1×10^{-6} s,

1 cm in the satellite range derived from signal travel time (e.g. laser ranging) corresponds to about 1×10^{-10} s.
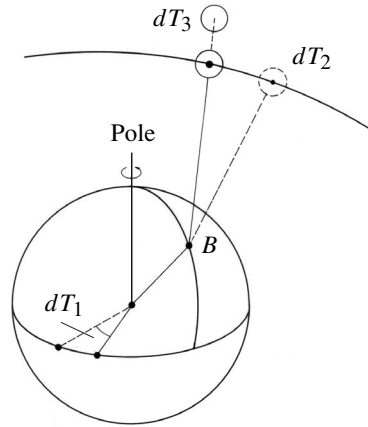
Figure 2.12. Effect of timing errors in satellite geodesy

The related requirements for the accuracy of time determination dT_i are as follows:

$$
\begin{aligned}
dT_1[\text{s}] &\leq 2 \times 10^{-5} && \text{for Earth rotation,} \\
dT_2[\text{s}] &\leq 1 \times 10^{-6} && \text{for orbital motion, and} \\
dT_3[\text{s}] &\leq 1 \times 10^{-10} && \text{for signal travel time.}
\end{aligned}
\tag{2.50}
$$

2.2.2 Sidereal Time and Universal Time

Sidereal time and universal time are directly related to the rotation of Earth, and they are thus equivalent time scales. Sidereal time equals the hour angle of the vernal equinox ♈, and consequently depends on the geographical longitude of the particular observation station. From Fig. 2.13 we may easily derive the following relations. The *Local Apparent* (or *True*) *Sidereal Time* (LAST), referred to the true vernal equinox, is

LAST = Local hour angle of the true vernal equinox.

For Greenwich we obtain the *Greenwich Apparent Sidereal Time* (GAST)

GAST = Greenwich hour angle of the true vernal equinox.

The vernal equinox is subject to the nutation in longitude (cf. [2.1.2]). Removing the nutation term, we obtain the *Local Mean Sidereal Time* (LMST), and the *Greenwich*

Mean Sidereal Time (GMST), respectively

$$\text{LMST} = \text{Local hour angle of the mean vernal equinox}$$
$$\text{GMST} = \text{Greenwich hour angle of the mean vernal equinox.}$$

The difference between the apparent and mean sidereal times is termed *Equation of Equinoxes* (Eq.E)

$$\text{GMST} - \text{GAST} = \Delta\psi \cos\varepsilon, \tag{2.51}$$

with $\Delta\psi$ (2.21) the nutation in longitude. For the east longitude Λ of the local meridian

$$\text{LMST} - \text{GMST} = \text{LAST} - \text{GAST} = \Lambda. \tag{2.52}$$

The apparent sidereal time is used for the evaluation of astronomical observations. However, for the construction of a time scale, only the mean sidereal time is used. The fundamental unit is the *Mean Sidereal Day*, defined as the interval between two consecutive upper transits of the mean vernal equinox across the meridian. The mean sidereal day does not correspond exactly to a complete revolution of Earth on its spin axis with respect to inertial space, because the position of the vernal equinox is af-

Figure 2.13. Definition of sidereal time

fected by precession. The daily difference is $0.^{\text{s}}0084$, with the sidereal day being shorter.

For practical purposes a time scale is required which corresponds to the apparent diurnal motion of the Sun. The hour angle of the true Sun experiences rather large variations during the year, caused by the changing declination of the Sun and the ellipticity of Earth's orbit. Consequently, this measure is not suitable for a uniform time scale. It is therefore substituted by a fictitious *Mean Sun*, which moves in the plane of the equator with constant velocity. The *Mean Solar Day* is thus defined as the interval between two successive transits of the mean fictitious Sun across the meridian. *Mean Solar Time* is measured by the hour angle of the mean Sun. The Greenwich hour angle of the mean Sun is called *Universal Time* (UT). For practical purposes the day starts at midnight, hence

$$\text{UT} = 12^{\text{h}} + \text{Greenwich hour angle of the mean Sun.} \tag{2.53}$$

Both concepts of time are based on Earth's rotation and they are closely connected to each other. Universal time can be considered as a special form of sidereal time. The difference in the length of the day for both definitions is around 4 minutes, because

the diurnal motion of Earth in its orbit amounts to $360°/365 \approx 1°$. The approximative relation is

$$1 \text{ mean sidereal day} = 1 \text{ mean solar day} - 3^{m}55.^{s}909. \qquad (2.54)$$

The raw universal time $UT0_B$, which is obtained from observations at a particular station B, is still affected by the location-dependent influences of the actual true pole position. The reduction to the conventional terrestrial pole (CTP) (cf. [2.1.2.3]) causes a change, $\Delta\Lambda_P$, in longitude, and hence in time. The universal time, which is referred to CTP, is termed

$$UT1 = UT0_B + \Delta\Lambda_P. \qquad (2.55)$$

UT1 is the fundamental time scale in geodetic astronomy and satellite geodesy, because it defines the actual orientation of the conventional terrestrial system with respect to space. UT1 is also the basic time scale for navigation. UT1 contains, however, all variations of Earth's rotation rate, and is thus not a uniform time scale.

The fundamental relation between UT1 and GMST was reformulated by the International Astronomical Union (IAU), and since 01.01.1984 has been defined as

$$\begin{aligned} GMST \text{ at } 0^{h}UT1 = 6^{h}41^{m}50.^{s}54841 + 8\,640\,184.^{s}812\,866\ T_u \\ + 0.^{s}093\,104\ T_u^2 - 6.^{s}2 \times 10^{-6}\ T_u^3. \end{aligned} \qquad (2.56)$$

T_u is the time interval from the standard epoch

$$J2000.0 = 01.01.2000,\ 12^{h}UT1,$$

counted in Julian centuries of 36 525 days. For more information see e.g. Aoki et al. (1982), Green (1985, chap. 10), Schödlbauer (2000, chap. 5) or the "Explanatory Supplement to the Astronomical Almanac" Seidelmann (ed.) (1992, chap. 2).

In order to achieve a rigorous definition of the sidereal rotation of Earth at the microarcsecond level, the IAU recommended at its 24th General Assembly 2000 (Resolution B 1.8) to reformulate the relation (2.56) with respect to a *non-rotating origin*, Guinot (1979), and to implement this by January 1, 2003. In this concept UT1 is linearly proportional to the *Earth rotation angle θ* defined as the angle measured along the equator of the CIP between the unit vectors directed toward the space-fixed CEO and Earth-fixed TEO (cf. [2.1.2.3]). Details are given in McCarthy (2000); Capitaine et al. (2000); Capitaine, et al. (2002).

For many purposes it is convenient to use a continuous count of days instead of the civil calendar in order to determine time intervals. A frequently used solution is the *Julian Date* (JD). The *Julian Day Number* is the number of days that has elapsed since Greenwich mean noon on January 1, 4713 B.C. The *Julian Date* of a given instant is the Julian day number followed by the fraction of the day elapsed since the preceding noon. The *Modified Julian Date* (MJD) is an abbreviated version of JD

$$MJD = JD - 2\,400\,000.5.$$

The half day is subtracted so that the day starts at midnight, as is the case with civil time reckoning. The MJD has been recommended by various international bodies such as IAU as a decimal day count which is independent of the civil calendar. MJD is usually reckoned in universal time (UT). The modified Julian day number has to be distinguished from the *Day of the year* (DOY). DOY is counted from the beginning of the respective year. Thus

$$\text{for 2002:} \quad \text{MJD} = 52275 + \text{DOY}$$
$$\text{for 2003:} \quad \text{MJD} = 52640 + \text{DOY}$$
$$\text{for 2004:} \quad \text{MJD} = 53005 + \text{DOY}$$
$$\text{for 2005:} \quad \text{MJD} = 53371 + \text{DOY}$$

2.2.3   Atomic Time

The international *atomic time scale* TAI (Temps Atomique International) was introduced to meet the requirements for an easily accessible and strictly uniform time scale. The unit of the atomic time was selected in such a way that it equals the duration of the ephemeris second [2.2.4]. The definition of the second of the atomic time scale has been worded by the 13th Conference of the International Committee of Weights and Measures in Paris, 1967, as follows:

> *The second is 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.*

This is also the definition of the unit of time of the *International System of Units* (SI).

The international atomic time scale is maintained by the Time Section of the International Bureau of Weights and Measures (*Bureau International des Poids et Mesures*, BIPM) in Paris, based on the readings of a large number of the most accurate atomic clocks in various laboratories. The Bureau International de l'Heure (BIH) was responsible for maintaining the atomic time scales until the 31st of   December 1987.

In practice, atomic time scales are derived from groups of commercial and laboratory cesium standards [2.2.5] which generate time intervals, based on the definition of the SI second. The readings refer to non-moving clocks at sea level. TAI is computed as the weighted mean of individual clocks (about 250 clocks in 2002). TAI is hence a statistically formed common time scale for international use. Each laboratory time scale can be regarded as a particular realization of the atomic time scale. The differences between TAI and the time scales of the participating laboratories are distributed on a monthly basis in the Circular T of the BIPM Time Section.

The epoch of TAI agreed with the epoch of UT1 on January 1, 1958. Due to the deceleration of Earth's rotation the difference between the time scales is increasing.

The difference, for some selected dates, amounts to

$$\begin{aligned}
\text{TAI} - \text{UT1} = +\ 6.^{\text{s}}1 \quad &\text{on January 1, 1968} \\
= +16.^{\text{s}}4 \quad &\text{on January 1, 1978} \\
= +23.^{\text{s}}6 \quad &\text{on January 1, 1988} \\
= +30.^{\text{s}}8 \quad &\text{on January 1, 1998} \\
= +31.^{\text{s}}9 \quad &\text{on January 1, 2001} \\
= +32.^{\text{s}}3 \quad &\text{on January 1, 2003.}
\end{aligned}$$

The rather large size of the differences stems from the fact that the unit of the SI-second was adopted from the length of the ephemeris second; and the ephemeris second was derived from the mean duration of the solar day between 1756 and 1895, when Earth's rotation was faster than today.

For many applications, navigation in particular, a time scale is required which provides both a highly uniform time unit and the best possible adaptation to UT1, and hence to Earth rotation. This is why, in 1972, a compromise time scale, *Universal Time Coordinated* (UTC), was introduced. UTC and TAI differ by an integer number n of seconds

$$\text{UTC} = \text{TAI} - n \cdot (1\,\text{s}). \tag{2.57}$$

Depending on the prevailing situation, n can be changed at given dates, namely on January 1 and/or July 1. Thus the epoch of UTC is adapted to UT1 by inserting or removing so-called *leap seconds*. The unit of UTC remains the SI second. The difference, DUT1, between both times should not exceed 0.9 s in absolute value

$$|\text{DUT1}| = |\text{UT1} - \text{UTC}| \overset{!}{\le} 0.9\,\text{s}. \tag{2.58}$$

DUT1 is distributed through the bulletins of the IERS, and it must be taken into account with all calculations related to Earth-fixed reference systems. In most countries the disseminated time signals refer to UTC. On January 1, 2003 the difference between TAI and UTC was

$$\text{TAI} - \text{UTC}_{2003} = +32\,\text{s}. \tag{2.59}$$

The Global Positioning System (GPS) uses its own particular time scale *GPS time*. It differs from UTC by a nearly integer number of seconds. Both time scales had identical epochs on January 5, 1980. Because GPS time is not incremented by leap seconds the difference between UTC and GPS time is increasing. The unit of GPS time is the SI second. However, GPS time is only derived from atomic clocks which form part of the GPS system. It is hence a "free" atomic time scale and may show slight differences when compared to TAI. The relation between UTC and GPS time is included in time bulletins of the USNO and the BIPM, and it is also disseminated within the "GPS satellite message" [7.1.3]. In 2003 the difference was approximately

$$\text{GPS time} - \text{UTC}_{2003} = +13\,\text{s}, \tag{2.60}$$

The exact relation is (e.g. BIPM (2002))

$$\text{GPS time} - \text{UTC} = n\,\text{s} - C_0,$$

n is an integer number, and the correction term C_0 is in the order of several nanoseconds. Thus the reception of GPS signals provides real-time access to TAI and UTC with uncertainties below 1 microsecond.

A similar relationship holds for GLONASS time [7.7] and UTC. Note that UTC and GPS time, as well as GLONASS time are atomic time scales.

2.2.4   Ephemeris Time, Dynamical Time, Terrestrial Time

A strictly uniform time scale can be found in the independent arguments of the theories of dynamics and of the ephemerides, i.e. the time-dependent positions of celestial bodies, described in adequate reference frames. Time scales which are based on such concepts fulfill at best the conceptional idea of *Inertial Time*.

In 1952 the IAU introduced *Ephemeris Time* (ET) as a theoretically uniform time scale for use with ephemeris. The *Ephemeris Second* was defined as a certain fraction of the *Tropical Year* 1900, and hence it was strictly uniform. In practice, the ephemeris time was derived from lunar observations, and it depended on a theory of the Sun and the system of astronomical constants. Its reading accuracy was only about 0.1 s on yearly averages. ET has never been disseminated by time signals. It was made available only through the publication of differences with respect to UT1, and later to TAI (Guinot, 1995).

In 1977 the IAU adopted the so-called *Dynamical Time Scales* in order to meet the arising requirements for a relativistic formulation of orbital motion. *Barycentric Dynamical Time* (TDB) was defined to be the time-like argument for the barycenter of the solar system, and *Terrestrial Dynamic Time* (TDT) was referred to geocentric ephemerides.

In the concept of *General Relativity* a clock, moving with Earth, experiences periodic variations up to 1.6 milliseconds, caused by the annual motion within the gravity field of the Sun. This effect, however, must not be considered in the computation of near-Earth satellite orbits, because the satellites move together with Earth. This is why *Terrestrial Dynamical Time* (TDT) was the appropriate time scale for geocentric calculations in satellite geodesy. A further advantage is, that compared with the *Barycentric Dynamical Time* (TDB), TDT is independent of various forms of relativistic theories (Seidelmann et al., 1992).

Dynamical time has been used as the argument for astronomical ephemerides since January 1, 1984. The SI second was formally introduced as the fundamental time unit in the TDT scale. It corresponds to the time which an atomic clock would measure on the rotating geoid. For the sake of continuity TDT was set equal to ET at the beginning of January 1, 1984. This is why a constant difference of $32.^{\text{s}}184$ exists between the TAI time scale, and the TDT (or ET) time scale.

In 1991 the IAU has defined new time scales in the framework of the general theory of relativity to clarify the relationships between *space-time coordinates*. In

this concept a time scale is regarded as one of the coordinate axes of a space-time reference frame (Guinot, 1995). The new time scales are the *Barycentric Coordinate Time* (TCB), the *Geocentric Coordinate Time* (TCG), and the *Terrestrial Time* (TT). Explicit formulas relating TDB, TDT, TCB, and TCG are given in Seidelmann et al. (1992, p. 42ff).

Terrestrial Time (TT) in essence is a new denomination for TDT. The word "dynamical" has been omitted because TT as an idealized time scale is no longer based on dynamical theories. TT is the time reference for geocentric ephemerides and hence the primary time scale for the relativistic treatment of near-Earth satellite orbits. TT differs from TCG only by a constant rate

$$dTT/dTCG = 1 - L_G, \tag{2.61}$$

where $L_G = 6.969290134 \times 10^{-10}$ is a defining constant (resolution B1.9, IAU 24th General Assembly 2000). The unit of TT is the SI second, hence TT is realized through the atomic time scale TAI (cf. [2.2.3]) with a constant offset of $32.^s184$ between both scales

$$TT \equiv TDT \equiv ET = TAI + 32.^s184. \tag{2.62}$$

Consequently, there is only a conceptional, not a practical, difference between both time scales.

For more information on this subject see e.g. Seidelmann et al. (1992), Seidelmann, Fukushima (1992), Guinot (1995), McCarthy (2000). In Fig. 2.14 an overview is given for time scales of interest to satellite geodesy.
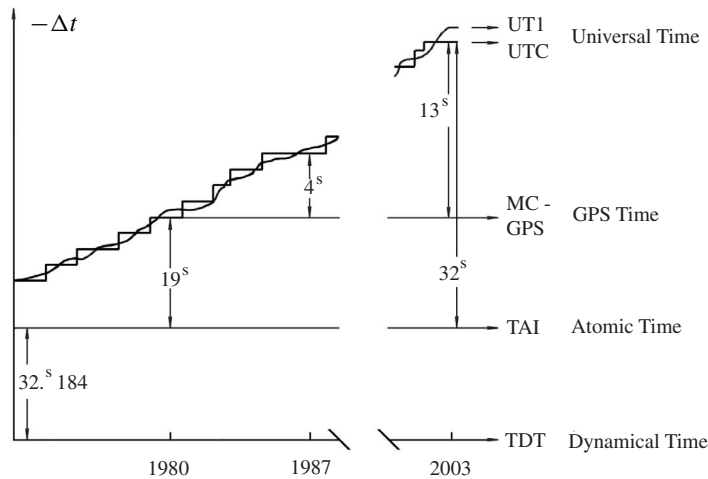
Figure 2.14. Time scales in satellite geodesy

2.2.5 Clocks and Frequency Standards

In satellite geodesy precise information is required on time and frequency. In many cases it is necessary to relate the epochs of some events which are observed at different stations, separated by large distances, with an accuracy of ± 1 microsecond ($1 \mu s$). The performance of frequency standards must reach a stability of up to $1 \cdot 10^{-15}$ over several hours. These high demands can only be fulfilled with atomic clocks. The most important component of a clock is an oscillating system (oscillator). The periodic motion of this system has to be generated, maintained and read out by suitable means. In modern clocks, e.g. in atomic clocks, the conversion of the oscillator cycles to the scale unit "one second" is realized via electronic counters or divisors.

For an ideal clock C_I the relation between the cycle period T_I and the oscillator frequency f_I is defined as

$$T_I = \frac{1}{f_I}. \tag{2.63}$$

Counting N_I cycles over a given time interval $(t - t_0)$ yields the ideal (strictly uniform) time scale

$$(t - t_0) = N_I T_I = \frac{N_I}{f_I}. \tag{2.64}$$

Here N_I equals the integral

$$N_I = \int_{t_0}^{t} f_I \, dt = f_I (t - t_0), \tag{2.65}$$

which is the total number of cycles since the starting epoch t_0.

For an atomic clock C_i which exists in reality, like for every other clock, the frequency is not strictly constant. The behavior of the frequency is usually described as e.g. Fell (1980); Wübbena (1991); Hahn (1999)

$$f_i(t) = f_I + \Delta f_i + \dot{f}_i(t - t_0) + \tilde{f}_i(t). \tag{2.66}$$

The individual terms are:

$$\Delta f_i \quad \text{constant frequency bias of the oscillator } C_i,$$
$$\dot{f}_i \quad \text{frequency drift, and}$$
$$\tilde{f}_i \quad \text{random frequency error.}$$

Counting the oscillations of this real clock C_i yields

$$N_i = \int_{t_0}^{t} f_i(t) \, dt = f_I(t - t_0) + \Delta f_i(t - t_0) + \frac{\dot{f}_i(t - t_0)^2}{2} + \int_{t_0}^{t} \tilde{f}_i(t) \, dt. \tag{2.67}$$

The related epoch is

$$(t_i - t_0) = N_I T_I(t - t_0) + \frac{\Delta f_i}{f_I}(t - t_0) + \frac{\dot{f}_i}{2 f_I}(t - t_0)^2 + \int_{t_0}^{t} \frac{\tilde{f}_i(t)}{f_I} \, dt. \tag{2.68}$$

When

$$\Delta t_i(t_0) = N_0 T_I \tag{2.69}$$

is the synchronization error at the first epoch t_0 we obtain for some later epoch t the total timing error of the clock C_i as

$$\Delta t_i(t) = t_i - t = \Delta t_i(t_0) + \frac{\Delta f_i}{f_I}(t - t_0) + \frac{\dot{f}_i}{2 f_I}(t - t_0)^2 + \int_{t_0}^{t} \frac{\tilde{f}_i(t)}{f_I} \, dt. \tag{2.70}$$

After renaming the expressions in (2.70) we get a frequently-used description of the timing error $\Delta t_i(t)$:

$$\Delta t_i = T_i(t_0) + R_i(t - t_0) + \frac{D_i}{2}(t - t_0)^2 + \int_{t_0}^{t} y(t) \, dt \tag{2.71}$$

with

$T_i(t_0)$	constant time bias,
R_i	time drift,
D_i	quadratic term (drift rate, ageing), and
$y(t)$	random relative frequency error.

For a particular clock the first three terms have to be estimated. Consequently the timing error of the clock depends on the uncertainty of the estimation, and on the integral of the random frequency error from the start epoch to the epoch of estimation. The particular estimation can be obtained through comparison with other clocks. This is why time laboratories and fundamental observation stations may operate several atomic clocks which are compared among each other or with clocks at other institutions on a regular basis.

The relative frequency errors show a typical behavior for different types of atomic clocks. These errors can be characterized either in the time domain or in the frequency domain. A suitable measure for relative frequency errors in the time domain is the so-called *Allan variance* (Allan, 1987).

It should be noted that, because of the extremely high accuracy requirements in some parts of satellite geodesy, the behavior of clocks in fundamental observation stations and in the satellites must be carefully studied. This is particularly true for the clocks in navigation satellites like GPS [7.1.2]. The typical frequency performance of clocks is demonstrated in Fig. 2.15.

In satellite geodesy the following classes of oscillators are in use:
− precision quartz crystal oscillator,
− rubidium standard,
− cesium standard, and
− hydrogen maser.

Precision quartz crystal oscillators are completely sufficient as time generators in satellite receivers when they are continuously controlled and updated by external signals, for example by the time and frequency signals from satellites. This is e.g.
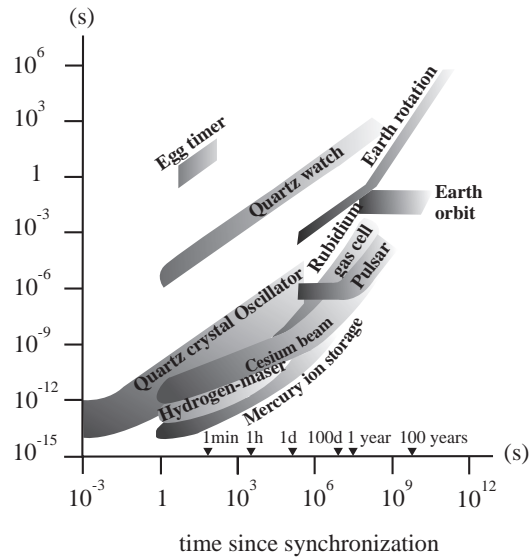
Figure 2.15. Stability of frequency standards; cf. Allan et al. (1997), also GPS World Supplement on Precise Timing, Dec. 1998

the case with the GPS satellites. Characteristic features of the quartz oscillators are that they are quite sensitive to temperature variations and that they are prone to a rather strong ageing process. In practice it is of importance that the quartz runs in stable temperature conditions and without interruptions or other disturbances. The frequency stability per day may range from 10^{-9} to 10^{-13}.

The characteristic feature of the *rubidium frequency standard* is its excellent long term stability. A rubidium standard can be used as an external oscillator for GPS observations, in particular to bridge periods with insufficient satellite coverage. A rubidium clock can reach a stability of 1×10^{-13} per day under the best conditions.

The *cesium frequency standards*, because of their high short- and long-term stability, can be regarded as the atomic clocks "par excellence". Assembled in groups, they form the core of time laboratories, and they are also present in fundamental satellite observation stations, in tracking stations for orbit control, or in laser ranging systems. The time base in the GPS satellites is realized through cesium and rubidium standards. Cesium standards are transportable, and commercially available. Laboratory cesium beam standards can realize the second with an accuracy of 1.5×10^{-14}. Commercially available standards are less accurate, but may equal the stability of laboratory standards for periods up to about 1 year (Seidelmann et al., 1992).

Hydrogen masers are necessary to meet the highest accuracy demands, such as those required by Very Long Baseline Interferometry (VLBI) [11.1]. A frequency stability $\sigma_{(\Delta f/f)}$ of 10^{-15} is required over time periods of 10^2 to 10^5 seconds. Hydrogen

masers are very sensitive, and to date have only been operational under laboratory conditions.

For a deeper treatment of the subject see the special literature, such as Bauch et al. (2000), PTTI (2000), or the *Supplement on Precise Timing* from GPS World (December 1998). From modern developments for improved time and frequency standards an uncertainty of about 1×10^{-15}, or even 10^{-16}, can be expected. Technology developments with *laser cooling of atoms*, *mercury ion chambers*, and *cesium fountains* are but some of the efforts. Their optimum use would require global time comparison with uncertainties of down to 10 ps, i.e. the time taken by a photon to travel 3 mm. A correct modeling of time comparison then has to include, among other effects, solid Earth tides and plate tectonics (Guinot, 1995). Detailed information on new developments can be taken from publications of leading time laboratories like the *National Institute of Standards* (NIST) in the U.S. or the *Physikalisch Technische Bundesanstalt* (PTB) in Germany.

A completely new development is *pulsar time*. Fast rotating neutron stars, so-called *pulsars* can be used as stable cosmic frequency generators. Pulsar PSR1937+21, discovered in 1982, is rotating with a period of 1.6 ms and emitting a beam of electromagnetic radiation sweeping Earth at each revolution. The arrival of each pulse can be dated with uncertainty of 0.3 μs. A number of other millisecond pulsars have subsequently be found.

After correction for the orbital motion of Earth and of the pulsars, and for other effects like deceleration of pulsar rotation, pulsars can serve as clocks at least as stable as the best atomic clocks. In future the combination of pulsar data with the readings of atomic clocks may generate stable long term time scales (Guinot, 1995).

2.3   Signal Propagation

Signals, on their path between satellites and ground stations, propagate through atmospheric regions of different nature and variable state, and thus experience different kinds of influences. Perturbations may occur to the direction of propagation, to the velocity of propagation and to the signal strength. For the user who is interested in the undisturbed signal the atmosphere introduces unwanted perturbations. The impacts on the observational results are, in many cases, much larger than the accuracy required in satellite geodesy. Consequently, atmospheric influences have to be determined directly by measurements and/or by modeling, and they have to be considered within the adjustment process.

On the other hand, information on the state of the upper atmosphere can be obtained when the received satellite signals are compared with signals that would be observed under atmospheric free conditions (e.g. Coco, 1991; Wanninger, 1992; Wild, 1994; Schüler, 2001). This latter aspect is, however, not discussed here.

In this chapter some elementary fundamentals of wave propagation are given [2.3.1], [2.3.2], and the characteristics of signal propagation through the troposphere and the ionosphere are presented [2.3.3]. For a full treatment of the subject see the

special literature (e.g. Maral, Bousquet, 1986; Davies, 1990; DeMunck et al., 1992; Parkinson et al., 1996; Langley, 1998b). The explicit correction formulas for a particular observation technique, for example Doppler, GPS or SLR are given in the relevant chapters ([6.4.2], [7.4.4], [8.4.1]).

2.3.1 Some Fundamentals of Wave Propagation

2.3.1.1 Basic Relations and Definitions

The relation between the *wavelength*, λ, the *frequency*, f, and the *propagation velocity*, v, is

$$v = \lambda \cdot f. \tag{2.72}$$

Herein λ has units of meters (m), f has units of Hertz (Hz, oscillations per second), and v units of meters per second. In the context of observation methods in satellite geodesy only electromagnetic waves are considered here. An electromagnetic wave is self-propagating with both electric and magnetic field components generated by the rapid oscillations of a charged particle. Its characteristics are described by *Maxwell's* equations, see Lorrain et al. (1988); Langley (1998b), or textbooks on physics. In the electromagnetic metrology waves can be regarded as disturbances of the electromagnetic field in time and space (Wells, 1974). For a periodic wave the disturbance is repeated at a fixed point after a lapse of time known as the period, P, and/or at a fixed time after the addition of a distance known as the wavelength, λ. The relation between frequency and period is

$$f = \frac{1}{P}. \tag{2.73}$$

The *phase*, Φ, of a periodic wave is the fractional part t/T of the period, P, through which the time t has advanced with respect to an arbitrary time origin t_0. Furthermore

$$\omega = 2\pi f, \tag{2.74}$$

the *angular frequency*, and

$$k = \frac{2\pi}{\lambda}, \tag{2.75}$$

the *phase constant* or *wave number*. It follows for the propagation velocity v, that

$$v = \lambda \cdot f = \frac{\lambda}{P} = \frac{\omega}{k}. \tag{2.76}$$

A periodic wave which can be modeled by a sinusoidal function in space and time is a *sinusoidal wave*. In what follows only waves tat are periodic functions in time are considered:

$$y = A \sin 2\pi \left(\frac{t}{P} + \Phi_0 \right), \tag{2.77}$$

where y is the magnitude of the disturbance at time t; Φ_0 is the phase of the wave at $t = 0$, and A is the maximum magnitude or the *amplitude* of the wave. The phase at time t is then

$$\Phi = \frac{t}{P} + \Phi_0. \tag{2.78}$$

$2\pi\,\Phi$ is called the *phase angle* φ. With (2.74) it follows from (2.77) that

$$y = A\sin(\omega t + \varphi_0). \tag{2.79}$$

Fig. 2.16 shows the geometrical interpretation of equation (2.79) (cf. Kahmen, 1978).
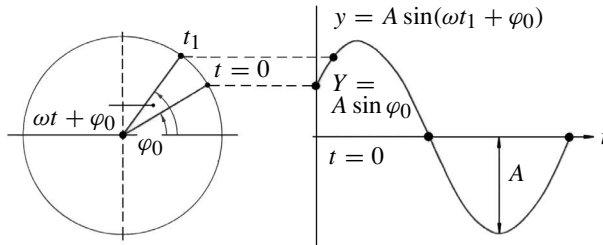
Figure 2.16. Representation of wave propagation

If we substitute the number of cycles N in equation (2.64) by the total phase Φ, referred to some starting epoch t_0, we obtain the relation between time, phase, and frequency

$$t = \frac{\Phi}{f}. \tag{2.80}$$

Equation (2.80) is identical with (2.73), and can be considered to be the defining equation of a clock. It gives the fundamental relation between the phase of a periodic process and the corresponding time reading at the clock. Equation (2.80) is essential for deriving the observation equation of GPS phase measurements (cf. [7.3.1.2]).

A further property of electromagnetic waves is *polarization*. *Linear polarization* means that the electric field component has a constant orientation in space. With *circular polarization* the electric field vector is rotating in space. In satellite geodesy circular polarized signals are mainly used because of the satellite's orbital motion, and because of the *Faraday rotation* of the plane of polarization due to Earth's magnetic field in the upper atmosphere [7.4.4.1].

In general, an electromagnetic wave can be characterized by these parameters: frequency, amplitude, phase, and polarization. In order to transmit information, one of these parameters can be varied in a controlled manner. This is called *modulation*, and can be achieved by analogue or digital means. We distinguish amplitude-, phase-, and frequency modulation. Data transmission via signal modulation is used in operational satellite systems like GNSS [7].

The wavelengths of electromagnetic waves, and hence their propagation velocity, depend on certain properties of the medium in which the waves are propagating. In a vacuum the velocity is

$$c = \frac{\lambda_{vac}}{P} = f\lambda_{vac} = \frac{\omega}{k_{vac}}. \tag{2.81}$$

The numerical value c, for the propagation velocity in a vacuum, is adopted by international scientific bodies. The value currently in use in satellite geodesy is (McCarthy, 2000)

$$c = 2.997\,924\,58 \cdot 10^8 \text{ms}^{-1}. \tag{2.82}$$

For propagation media other than a vacuum the propagation velocity is characterized by the *index of refraction n*

$$n = \frac{c}{v} = \frac{\lambda_{vac}}{\lambda} = \frac{k}{k_{vac}}. \tag{2.83}$$

Instead of n, which is near to 1, the refractivity

$$N = (n - 1) \cdot 10^6 \tag{2.84}$$

is preferred.

The appropriate determination of the refractivity N along the signal propagation path is essential in satellite geodesy because travel times of electromagnetic signals, or phase differences between different electromagnetic waves, are measured, and they are scaled into distances (measured in meters) with the adopted or modeled propagation velocity.

2.3.1.2   Dispersion, Phase Velocity and Group Velocity

A medium in which the propagation velocity of electromagnetic waves depends on the frequency is called a *dispersive medium*. In such a medium the refractivity depends on the frequency or the wavelength. The dispersion effect is caused by electromagnetic interactions between the electrically charged field of the medium and the external field of the penetrating wave. When the atomic frequency of the medium and the frequency of the penetrating wave are close together resonance occurs which generates a frequency-dependent influence on the propagation velocity (see e.g. Wells, 1974; Davies, 1990; Brunner, 1992; Langley, 1998b). The expression

$$\frac{dv}{d\lambda} \qquad \text{is called } \textit{velocity dispersion.} \tag{2.85}$$

In a medium with velocity dispersion we observe different propagation velocities for sinusoidal waves (phases) and groups of waves. We must distinguish the

   − propagation velocity of the phase of a particular wave with uniform wavelength (*phase velocity v_p*), and the

– propagation velocity of a wave group, generated by a superposition of different waves of different frequencies (*group velocity* v_g).

The relation between group velocity and phase velocity was first described by *Rayleigh* (1881) as

$$v_g = v_p - \lambda \frac{dv_p}{d\lambda}. \tag{2.86}$$

For the derivation of (2.86) see Wells (1974) or textbooks on physics or electromagnetic waves. Corresponding relations are valid for the refraction index

$$n_g = n_p + f \frac{dn}{df}. \tag{2.87}$$

The group velocity characterizes the velocity at which energy, or information, is propagated. Following the theory of *Fourier* such a signal can be regarded as a superposition of many particular periodic waves with different frequencies which all experience a different dispersion.

In satellite geodesy we have to prove carefully whether for a particular observable the group velocity or the phase velocity has to be applied. In GPS technology, for instance, the propagation of code signals is affected by the group velocity v_g, and the propagation of carrier phases by the phase velocity v_p.

The ionosphere is a dispersive medium for microwaves, but the troposphere is not. For frequencies in the optical domain the contrary holds. The phase velocity, in a dispersive medium, can exceed the vacuum velocity c. The group velocity, however, cannot, in accordance with the relativity theory. In non-dispersive media $v_g = v_p$.

2.3.1.3   Frequency Domains

The frequency spectrum of electromagnetic waves spans nearly 20 orders of magnitude (Fig. 2.17). In satellite geodesy only two rather small domains are used, namely the visible light (0.4–$0.8 \cdot 10^{15}$ Hz) and microwave domains (10^7–10^{10} Hz). Some prefixes and symbols which are commonly used for the description of frequencies are explained in Table 2.2.
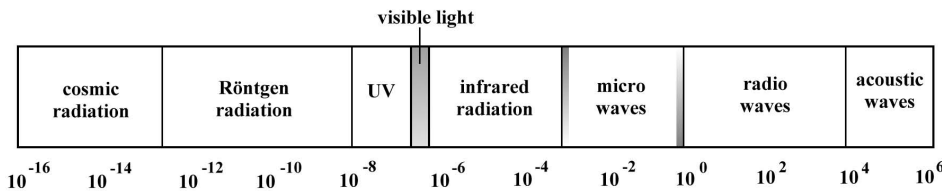
Figure 2.17. Spectrum of electromagnetic waves [m]; acoustic waves are included for information

Table 2.2. Prefixes, symbols, and orders of magnitude

centi	c	10^{-2}	Hecto	H	10^{2}
milli	m	10^{-3}	Kilo	K	10^{3}
micro	μ	10^{-6}	Mega	M	10^{6}
nano	n	10^{-9}	Giga	G	10^{9}
pico	p	10^{-12}	Tera	T	10^{12}
femto	f	10^{-15}	Peta	P	10^{15}

Different kinds of subdivisions and terminology are in use for electromagnetic waves. In information technology a subdivision into *frequency bands* is customary (Table 2.3). In satellite geodesy the subdivision into *radar bands* is also in use (Table 2.4). The particular assignments to capital letters were generated in a random way during World War II.

Table 2.3. Frequency bands

symbol	denomination	wavelength	frequency
VLF	Very Low Frequency	> 10 000 m	< 30 KHz
LF	Low Frequency	1000–10 000 m	30–300 KHz
MF	Medium Frequency	100–1000 m	300–3000 KHz
HF	High Frequency	10–100 m	3–30 MHz
VHF	Very High Frequency	1–10 m	30–300 MHz
UHF	Ultra High Frequency	10 cm–1 m	300–3000 MHz
SHF	Super High Frequency	1 cm–10 cm	3 GHz–30 GHz
EHF	Extremely High Frequency	1 mm–1 cm	30–300 GHz

Table 2.4. Radar bands

denomination	frequency		mean wavelength
P-band	220–300	MHz	115 cm
L-band	1–2	GHz	20 cm
S-band	2–4	GHz	10 cm
C-band	4–8	GHz	5 cm
X-band	8–12.5	GHz	3 cm
Ku-band	12.5–18	GHz	2 cm
K-band	18–26.5	GHz	1.35 cm
Ka-band	26.5–40	GHz	1 cm

2.3.2   Structure and Subdivision of the Atmosphere

The structure of the atmosphere can be described, for most practical purposes, as a set of concentric spherical shells with different physical and chemical properties. Various subdivisions are possible, that in most cases follow the main characteristic feature of interest. Fig. 2.18 gives a simplified schematic representation.

With respect to signal propagation a subdivision into *troposphere* and *ionosphere* is advisable, because the particular propagation conditions are quite different.

— The troposphere is the lower part of Earth's atmosphere which extends from the surface to about 40 km. Signal propagation depends mainly on the water vapor content and on temperature.

— The ionosphere is the upper part of Earth's atmosphere between approximately 70 and 1000 km. Signal propagation is mainly affected by free charged particles.

Altitude [km]	Temperature	Ionisation	Magnetic field	Propagation	Technical
100 000 10 000	Thermo - sphere	Protono - sphere	Magneto - sphere	Iono - sphere	Upper Atmo - sphere
1 000		Iono - sphere			
100	Mesosphere				
	Stratosphere				
10	Troposphere	Neutro - sphere	Dynamo - sphere	Tropo - sphere	Lower Atmo - sphere

Figure 2.18. Possible subdivision schemes of Earth's atmosphere

The *troposphere* is the gaseous atmosphere where the daily weather takes place. The temperature decreases with height by 6.°5 C/km. Horizontal temperature gradients are only a few degrees/100 km. Charged particles are virtually absent. The uncharged atoms and molecules are well mixed, and thus the troposphere is practically a neutral gas. The index of refraction is slightly greater than 1. It decreases with increasing height and becomes nearly 1 at the upper limit of the troposphere, corresponding to the continuously decreasing density of the medium. Nearly 90% of the atmospheric mass is below 16 km altitude, and nearly 99% is below 30 km (Lutgens, Tarbuck, 1998). For electromagnetic waves in the radio-frequency spectrum the troposphere is not a dispersive medium. The index of refraction does not depend on the frequency; it depends on air pressure, temperature, and water vapor pressure. Because of the dynamic behavior of tropospheric conditions it is difficult to model the index of refraction.

The *ionosphere* can be defined as that part of the high atmosphere where sufficient electrons and ions are present to affect the propagation of radio waves (Davies, 1990; Langley, 1998b). The generation of ions and electrons is proportional to the radiation intensity of the sun, and to the gas density. A diagram indicating the number of ions produced as a function of height shows a maximum in ion production rate. Such a diagram is called the *Chapman-profile*; the general behavior of this profile is illustrated in Fig. 2.19. The exact shape of the curve and the related numerical values are not given in the graph because they depend on several parameters, and they are highly variable functions (see later). The spatial distribution of electrons and ions is mainly determined by two processes:

- photo-chemical processes that depend on the insolation of the sun, and govern the production and decomposition rate of ionized particles, and

- transportation processes that cause a motion of the ionized layers.
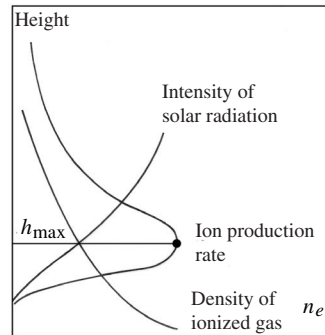
Figure 2.19. Chapman curve of ionization

Both processes create different layers of ionized gas at different heights. The main layers are known as the D-, E-, F_1-, and F_2-layers. In particular, the F_1-layer, located directly below the F_2-layer, shows large variations that correlate with the relative sun spot number. Geomagnetic influences also play an important role. Hence, signal propagation in the ionosphere is severely affected by solar activity, near the geomagnetic equator, and at high latitudes (cf. [7.4.4.1]).

The state of the ionosphere is described by the *electron density n_e* with the unit [number of electrons/m^3] or [number of electrons/cm^3]. The four principal layers are designated in Table 2.5.

Table 2.5. Characteristic features of the main ionospheric layers

layer		D	E	F_1	F_2
height domain	[km]	60–90	85–140	140–200	200–1000
electron density	at day	10^2–10^4	10^5	$5 \cdot 10^5$	10^6
n_e [el/cm^3]	at night	—	$2 \cdot 10^3$	$5 \cdot 10^4$	$3 \cdot 10^5$

Due to variable insolation of the Sun the spatial distribution of the layers varies during the day. The D-layer is only generated over the daylight side of Earth. The

impact of the state of the ionosphere on the propagation of waves is characterized by the *Total Electron Content* TEC, where

$$\text{TEC} = \int_S^R n_e(s)\, ds. \tag{2.88}$$

The integral contains the total number of electrons that are included in a column with a cross-sectional area of $1\ \text{m}^2$, counted along the signal path s between the satellite S and the receiver R. For comparison purposes among sets of TEC data the vertical electron content VTEC is formed as

$$\text{VTEC} = \frac{1}{F} \cdot \text{TEC}, \tag{2.89}$$

where

$$F = \frac{1}{\cos z^I}$$

is called the *obliquity factor* or *mapping function*. z^I is the zenith angle between the signal path and a horizontal plane in the mean altitude h_i. The unit of measurement is the TECU (*Total Electron Content Unit*):

$$1\ \text{TECU} = 1 \cdot 10^{16} \text{el/m}^2. \tag{2.90}$$

A frequently used model for data reduction in satellite geodesy is the *single layer model*. In this the total electron content is represented by a spherical layer at the mean ionospheric height h_I, usually near 400 km (Fig. 2.20). On this layer, P_I is the *ionospheric piercing point* of the signal path to a satellite S, P_S the subionospheric point, r_E Earth's radius, and z the zenith angle of S for an observer R. The zenith angle z^I at P_I then is given by
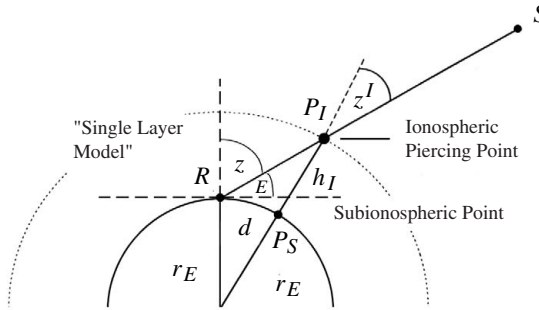
Figure 2.20. Single layer model of the ionosphere

$$z^I = \arcsin\left(\frac{r_E}{r_E + h_I} \sin z\right). \tag{2.91}$$

F increases with increasing zenith angle z to a satellite target. Table 2.6 (Wanninger, 1994) shows that for small elevation angles TEC can reach at most three times the value of VTEC. This is also true for the effect of ionospheric path delay in satellite geodesy (see Table 2.6).

Values of TEC vary between 10^{16} and 10^{19} electrons per m^2 along the radio wave path. The electron density is highly variable and depends mainly on

Table 2.6. Obliquity factor F and distance d between observer and subionospheric point

E [degree]	z [degree]	z^I [degree]	F	d [km]
90	0	0	1.00	0
60	30	28	1.13	215
30	60	55	1.73	603
20	70	62	2.14	873
10	80	68	2.66	1344
5	85	70	2.87	1712

− geographic location,
− time of the day,
− season of the year, and
− solar activity.

Regions of highest TEC are located approximately ± 15 to ± 20 degrees each side of Earth's magnetic equator (cf. Fig. 7.52, p. 313). The day to day variability has a standard deviation of $\pm 20\%$ to 25% of monthly average conditions (Klobuchar, 1996). Short term variations are *travelling ionospheric disturbances* (TID) with a period of minutes to about 1 hour, and *ionospheric scintillation* with a period of seconds. Of particular importance is the variance of the solar UV flux. The sun varies in its energy output over an approximate 11-year cycle (see Fig. 2.21). The last maximum was in the year 2000. In times of solar maximum the signals of operational GNSS systems can be heavily corrupted (cf. [7.4.4.1]).
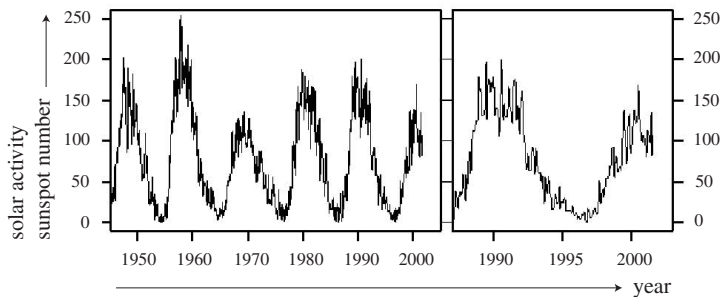
Figure 2.21. 11-year cycles of solar activity

The high variability of the ionosphere makes modeling and prediction difficult. Models of the electron density fall into two types: empirical models, derived from existing data, and physical models, derived from physical principles. Examples of

empirical models are the *International Reference Ionosphere* (IRI), the *Bent model*, and the *Klobuchar model*. Physical models are rather cumbersome and seldom used in satellite geodesy. Within the models the VTEC is described either by two-dimensional polynomials for local and regional applications, or by spherical harmonic expansion for continental and global representation. For details see Davies (1990); Wild (1994); Klobuchar (1996); Wanninger (2000).

Since 1996 the *International GPS Service* (IGS) [7.8.1] has generated, on a regular basis, global TEC models from GPS observations at selected globally distributed stations. Rapid products are available after several hours, and precise products after three days.

The ionosphere is a dispersive medium for radio waves. For an index of refraction n in ionized gas the *formula of dispersion* (e.g. Davies, 1990) is

$$n^2 = 1 - n_e \frac{C^2 e^2}{\pi f^2 m_e} \tag{2.92}$$

with

e elementary mass, and

m_e electron mass.

Rearranging and neglecting higher order terms gives

$$n = 1 - \frac{C \cdot n_e}{f^2}, \tag{2.93}$$

with $C = 40.3$. The coefficient C contains all constant parameters. An explicit derivation of (2.93) can be found in Hartmann, Leitinger (1984) or in textbooks on geophysics. Formula (2.93) indicates that the index of refraction, and thus the time delay of signal propagation, is proportional to the inverse of the squared frequency. Consequently, one part of the ionospheric delay can be modeled when two frequencies are used [2.3.3]. Furthermore (2.93) shows that higher frequencies are less affected by the ionosphere.

2.3.3   Signal Propagation through the Ionosphere and the Troposphere

Fig. 2.22 shows, for the microwave domain, the behavior of the refractivity N as a function of height. For the troposphere, N is positive, and independent of the frequency used. For the ionosphere, N is negative, and depends on the frequency. According to (2.93) the refractivity decreases with increasing frequency. One consequence is that higher accuracy can be obtained in propagation modeling when higher frequencies are used (cf. Table 2.7). Two considerations, however, limit the increase of the selected frequencies:

- Higher frequencies are technically demanding. The frequency domain above 10 GHz cannot easily be utilized with existing technology.
- With higher frequencies the atmospheric absorption in the troposphere increases. Without rainfall, the absorption can be neglected for frequencies between 30 MHz and 30 GHz. With precipitation, however, signals in the frequency domain > 1 GHz experience considerable attenuation.
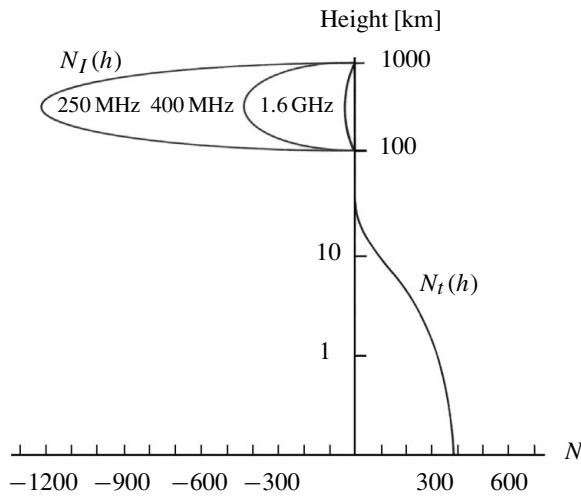
Figure 2.22. Behavior of the refractivity N for microwaves as a function of height; for the troposphere (N_t) and the ionosphere (N_I)

Table 2.7. Effect of the ionospheric propagation delay on range measurements for single-frequency observations, and residual errors for dual-frequency observations (Hieber, 1983)

single-frequency	400 MHz	1600 MHz	2000 MHz	8000 MHz
average effect	50 m	3 m	2 m	0.12 m
for 90% <	250 m	15 m	10 m	0.6 m
maximum effect	500 m	30 m	20 m	1.2 m
dual-frequency	150/400 MHz	400/2000 MHz	1227/1572 MHz	2000/8000 MHz
average effect	0.6 m	0.9 cm	0.3 cm	0.04 cm
for 90% <	10 m	6.6 cm	1.7 cm	0.21 cm
maximum effect	36 m	22 cm	4.5 cm	0.43 cm

The selection of frequencies for a particular satellite system is always a compromise. This was the case with the TRANSIT system [6] when 150/400 MHz were selected reflecting the technological progress of the 1960's. And this is true for the GPS system [7] with the selection of 1.2/1.6 GHz. Table 2.7 gives an impression of how the ionosphere affects the propagation delay at different frequencies, and it indicates the residual errors when measurements on two frequencies are available. It becomes clear that for the GPS system, operating with two frequencies, the residual errors are mostly below 1cm.

Because of the very different behavior of signal propagation in the troposphere and ionosphere their effects are usually discussed and treated as separate topics. For a review see Davies (1990); DeMunck et al. (1992); Wanninger (1994); Schüler (2001).

2.3.3.1 Ionospheric Refraction

The influence of the ionosphere on signal propagation in the radio frequency domain is mainly characterized by the dispersion. The refraction coefficient describing the propagation of phases can be written as a power series

$$n_p = 1 + \frac{c_2}{f^2} + \frac{c_3}{f^3} + \frac{c_4}{f^4} + \cdots . \tag{2.94}$$

The coefficients c_i are independent of the carrier frequency f; however, through the electron density n_e, they depend on the state of the ionosphere. The coefficient c_2 was estimated with (2.93) to be $c_2 = -40.3\,n_e$. Hence we find the approximate relation

$$n_p = 1 - \frac{40.3\,n_e}{f^2}. \tag{2.95}$$

Consequently, with a knowledge of the electron density, an approximate correction can be computed for the delay in signal propagation. Various, some highly sophisticated, models have been developed for estimation of the electron density. For correction of GPS measurements the model of Klobuchar (1987) is usually applied [7.4.4.1]. This model corrects about 50% of the total ionospheric effect. The model is represented through a set of variable coefficients that are valid for a few days. A much better correction is possible when the coefficient c_2 can be determined from simultaneous observations of satellite signals transmitted on two different frequencies. These requirements were fulfilled with the TRANSIT system [6.2] and by modern systems like GPS [7], GLONASS [7.7], PRARE [4.3.3.3], DORIS [6.7], and most of the altimeter sensors [9.2].

The first order term c_2/f^2 can be separated from higher order terms when they are sufficiently different in magnitude. For the frequencies of the TRANSIT system the 3rd order term is 10 times smaller than the 1st order term; the 2nd order term can be neglected (Black, 1980). For GPS the 1st order term is nearly three orders of magnitude larger than the other terms. The remaining residual errors are discussed together with the particular observation method, e.g. [6.4.2], [7.4.4.1], [9.3.3].

The refraction coefficient n_g of the group delay follows from equation (2.87) and with the first derivative of (2.94)

$$\frac{dn}{df} = -\frac{2c_2}{f^3} - \frac{3c_3}{f^4} - \frac{4c_4}{f^5} - \cdots \tag{2.96}$$

as

$$n_g = n_p - \frac{2c_2}{f^2} - \frac{3c_3}{f^3} - \frac{4c_4}{f^4} - \cdots , \tag{2.97}$$

and with (2.94) as

$$n_g = 1 - \frac{c_2}{f^2} - \frac{2c_3}{f^3} - \frac{3c_4}{f^4} - \cdots .$$ (2.98)

The coefficient n_g is greater than 1, i.e. the measured range is too large. The group velocity is smaller than the phase velocity, cf. (2.86). Truncating after the first order term, it follows that

$$n_g = 1 + \frac{40.3\, n_e}{f^2}.$$ (2.99)

A comparison with (2.95) makes it clear that the effects of the ionosphere on the phase velocity and the group velocity is approximately equal in magnitude, but has a different sign. This property can be used in the data reduction, when both types of observations are available, as is the case with GPS (cf. [7.3]). Using (2.83) we find for the phase and the group velocity

$$v_p = \frac{c}{1 - \frac{40.3\, n_e}{f^2}} \quad \text{and} \quad v_g = \frac{c}{1 + \frac{40.3\, n_e}{f^2}}.$$ (2.100)

The range error ΔS_I caused by the ionospheric propagation delay can be easily estimated (Hofmann-Wellenhof et al., 2001; Kaplan, 1996). The measured range is, by integration along the signal path

$$S = \int_{\mathrm{SV}}^{\mathrm{User}} n\, ds,$$ (2.101)

the geometrical range is along the straight line, setting $n = 1$, see Fig. 2.20

$$S_0 = \int_S^R ds_0.$$ (2.102)

The path length difference is then

$$\Delta S_{\mathrm{ION}} = S - S_0 = \int_S^R n\, ds - \int_S^R ds_0.$$ (2.103)

With (2.95) and (2.99) we find for the phase delay

$$\Delta S_{\mathrm{ION},\, p} = \int_S^R \left(1 - \frac{40.3\, n_e}{f^2}\right) ds - \int_S^R ds_0,$$ (2.104)

and for the group delay

$$\Delta S_{\mathrm{ION},\, g} = \int_S^R \left(1 + \frac{40.3\, n_e}{f^2}\right) ds - \int_S^R ds_0.$$ (2.105)

The delay will be small, hence we simplify (2.104) and (2.105) by integrating the first term along the straight line ds_0 and obtain, with

$$\mathrm{TEC} = \int_S^R n_e\, ds_0$$

for the phase and the group delay

$$\Delta S_{\text{ION}, p} = -\frac{40.3 \, \text{TEC}}{f^2} \quad \text{and} \quad \Delta S_{\text{ION}, g} = \frac{40.3 \, \text{TEC}}{f^2}. \tag{2.106}$$

Fig. 2.23 shows, for some current space techniques, the range errors caused by different levels of electron content.
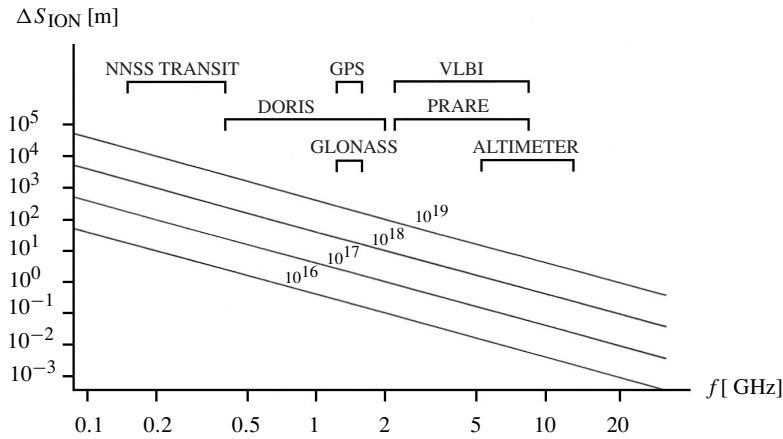
Figure 2.23. Ionospheric induced range error ΔS_{ION} for different TEC levels (cf. Wanninger (1994))

2.3.3.2 Tropospheric Refraction

Refractive effects in non-dispersive media are independent of the frequency used [2.3.1.2]; this is why all electromagnetic waves in the total radio-spectrum up to about 15 GHz are affected in the same way by the troposphere. The refractivity at the base of the atmosphere can be described as a function of meteorological parameters with an empirical formula (Hartmann, Leitinger, 1984):

$$N_T = C_1 \frac{P'}{T} + C_2 \frac{e}{T} + C_3 \frac{e}{T^2} \tag{2.107}$$

with

P,	the air pressure, in Hectopascal [HPa],
e,	partial pressure of the water vapor, in Hectopascal,
$P' = P - e$,	pressure of the dry gas, and
T,	temperature, in Kelvin.

Rearrangement of the terms, and separation into dry and wet components, gives (Spilker, 1996d; Langley, 1998b; Schüler, 2001)

$$N_T = C_1 \frac{P}{T} + C_4 \frac{e}{T^2} = N_d + N_w. \tag{2.108}$$

Recommended numerical values for the coefficients are

$$C_1 = 77.6 \quad \text{and} \quad C_4 = 3.73 \cdot 10^5.$$

This leads to

$$N_d = 77.6 \frac{P}{T} \quad \text{and} \quad N_w = 3.73 \cdot 10^5 \frac{e}{T^2}. \tag{2.109}$$

The integration of the tropospheric propagation delay along the signal path between the observer R and the top of the effective atmosphere A yields the total influence on the measured range:

$$\Delta s_T = \int_R^A (n - 1)\, ds = 10^{-6} \int_R^A N_T\, ds. \tag{2.110}$$

The integral can be solved when the behavior of the refractivity along the signal path is known. The dry and the wet terms are determined separately because they are formulated as distinct functions of height:

$$\Delta s_T = \Delta s_d + \Delta s_w = 10^{-6} \int_R^{H_d} N_d\, ds + 10^{-6} \int_R^{H_w} N_w\, ds. \tag{2.111}$$

H_d and H_w are the effective altitudes of the dry and the wet terms respectively.

A direct measurement of the refractivity along the signal propagation path is usually not feasible. This is why various models for a description of the height-dependent behavior of the refractivity have been developed. Input parameters are mostly the meteorological surface data near the observation site. The basic research was done by Hopfield (1969, 1971), and the *Hopfield model*, with some modifications and supplements, is still used today.

Theoretical considerations lead to a height-dependent function of the 4th degree for the dry component

$$N_d(h) = N_{d0} \left(\frac{H_d - h}{H_d} \right)^4. \tag{2.112}$$

N_{d0} is the surface refractivity from formula (2.109); h is the height above the surface. The exponent 4 and the parameter H (in meters) were determined by Helen Hopfield empirically from globally distributed balloon data

$$H_d = 40\,136 + 148.72\,(T - 273.16). \tag{2.113}$$

Further assumptions are:
  − the temperature decreases with height as 6.71 C°/km,

- the dry atmosphere behaves as an ideal gas,
- the atmosphere is built up in spherical layers, and
- the refractivity does not change with time.

No equivalent theoretical derivation has yet been developed for the wet term. This is why a corresponding expression for N_w is used in practice, without a theoretical foundation:

$$N_w(h) = N_{w0} \left(\frac{H_w - h}{H_w} \right)^4 . \tag{2.114}$$

For H_w a mean value is adopted

$$H_w = 11\,000 \text{ m.}$$

The integration of (2.111) along the curved signal path is very difficult, and not possible in the closed form. In practice, several simplifications are used; the line of sight is assumed to be a straight line.

In general, the delay in zenith direction, Δs_z, is estimated, and the projection to an arbitrary zenith angle z or elevation angle E is realized through an appropriate *mapping function* $\mathrm{m}(z)$ or $\mathrm{m}(E)$. Over the past 20 years or so, a variety of model profiles and mapping functions have been developed. The simplest mapping function is the cosecant of the elevation angle E (see Fig. 2.20, p. 50),

$$\Delta s_T = \mathrm{m}(E)\,\Delta s_z = \frac{1}{\sin E}\,\Delta s_z = \frac{1}{\cos z}\,\Delta s_z, \tag{2.115}$$

assuming that spherical constant-height surfaces can be approximated as plane surfaces. This is only reasonable for high elevation angles (Langley, 1998b). Many solutions are based on the *fractional expansion approach* of Marini (1972) who showed that the elevation angle dependence can be expressed as a continued fraction form in terms of the sine of the elevation angle E

$$\mathrm{m}(E) = \cfrac{1}{\sin E + \cfrac{a}{\sin E + \cfrac{b}{\sin E + \cfrac{c}{\sin E + \cdots}}}}. \tag{2.116}$$

The coefficients $a, b, c, \ldots$ are constants or linear functions. Note, that in contradiction to theory, $m\,(90°) \neq 1$. Some mapping functions therefore use a normalized form of (2.116). This can be achieved by replacing the second and following sine-terms by the tangent.

The following algorithm stems directly from Hopfield (cf. Hopfield, 1971; Kouba, 1983a)

$$\Delta s_T = \frac{K_d}{\sin(E^2 + 6.25)^{\frac{1}{2}}} + \frac{K_w}{\sin(E^2 + 2.25)^{\frac{1}{2}}} \tag{2.117}$$

with

$$K_d = 155.2 \cdot 10^{-7} \cdot \frac{P}{T} H_d, \quad K_w = 155.2 \cdot 10^{-7} \cdot \frac{4810\,e}{T^2} H_w.$$

K_d and K_w describe the total effect of the tropospheric refraction in the direction to the zenith. E is the elevation angle (in degrees) of the satellite as it is seen by the observer. The following approximation of Black (1978) is frequently used in satellite geodesy

$$
\begin{aligned}
\Delta s_T = K_d &\left\{ \left(1 - \left(\frac{\cos E}{1 + l_c \left(\frac{H_d}{r} \right)} \right)^2 \right)^{-\frac{1}{2}} - b\,(E) \right\} \\
+ K_w &\left\{ \left(1 - \left(\frac{\cos E}{1 + l_c \left(\frac{H_w}{r} \right)} \right)^2 \right)^{-\frac{1}{2}} - b\,(E) \right\}
\end{aligned}
\tag{2.118}
$$

with

$$l_c = 0.167 - (0.076 + 0.00015(T - 273))\exp^{-0.3E} \tag{2.119}$$

and r the geocentric distance of the observation station. The *bending correction* $b\,(E)$ describes the deviation of the real, bent, signal path from the straight geometrical connection between the observer and the satellite:

$$b\,(E) = \frac{1.92}{(E^2 + 0.6)}. \tag{2.120}$$

The empirical expression for l_c, given by Black (1978), contains some further terms of a series expansion. However, the form (2.119) or even a constant value $l_c = 0.15$ is sufficient, in most practical applications, for satellite elevations $> 5° \ldots 10°$. The bending correction is not required in many cases.

For elevations $>30°$ Black (1978) proposes the simple correction formulas

$$\Delta s_d = 2.31\,P\,\mathrm{cosec}\,E, \quad \text{and} \quad \Delta s_w = k_w\,\mathrm{cosec}\,E. \tag{2.121}$$

P is the surface pressure, expressed in *standard atmospheres* (1 atm = 1013.25 HPa), and k_w is a regional empirical constant:

$$
\begin{aligned}
k_w = {} & 0.28 \text{ summer in tropical areas or mean latitudes} \\
& 0.20 \text{ spring or autumn in mean latitudes} \\
& 0.12 \text{ winter in maritime mean latitudes} \\
& 0.06 \text{ winter in continental mean latitudes} \\
& 0.05 \text{ in polar regions.}
\end{aligned}
$$

These coefficients can also be introduced into the explicit formula (2.118).

Further algorithms and mapping functions can be taken from the literature; for example from Saastamoinen (1973); Davis et al. (1985); Herring (1992); Niell (1996). An excellent review of the current status is given in Nothnagel (2000) and Schüler (2001).

When larger portions of the satellite orbit are observed, for example with GPS or GLONASS [7] a *tropospheric scale bias* C_T can be introduced into the adjustment algorithm as a parameter for the total correction term. Black (1980) proposed for the adjustment of Transit observations

$$\Delta s_T = \Delta s_d + \Delta s_w = C_T I_a,$$

with

$$I_a \approx \left(\left(1 - \frac{\cos E}{1 + 0.15 \left(\frac{H_d}{r} \right)} \right)^2 \right)^{-\frac{1}{2}}. \tag{2.122}$$

For modern approaches to GNSS adjustment see [7.3]. The estimation of a scale bias is of particular interest in order to separate tropospheric delay from radial orbit and station height errors.

As a result of tropospheric refraction the optical distance, measured between the observer and the satellite, is longer than the direct geometrical range. For elevations $< 10°$ the influence easily exceeds 10 m. The portion of the wet term reaches only around 10% of the total influence. Table 2.8 gives some numerical values for an average situation ($T = 15°C$; $K_w = 0.20$), and for different elevation angles, that can be used for a rough estimation (Richardus, 1984).

Table 2.8. Influence of the tropospheric refraction on measured ranges (m)

Elevation angle	90°	20°	15°	10°	5°
Δs_d	2.31	6.71	8.81	12.90	23.61
Δs_w	0.20	0.58	0.77	1.14	2.21
Δs_T	2.51	7.29	9.58	14.04	25.82

The dry component in (2.118) has a standard deviation of about 2%, corresponding to 4 cm in the direction of the zenith (Kouba, 1983a). However, this is only true if the atmosphere is regarded as an ideal gas. Near a weather front the error of the model can reach up to 15%. The wet component can be modeled for the zenith direction with an accuracy of about $10\% - 20\%$, corresponding to a standard deviation of 3 to 5 cm. A significant improvement can be expected as soon as there are satisfactory models for the distribution of water vapor in the atmosphere.

Water vapor observing systems may be ground-based or space-based. Ground-based systems include, besides routine surface meteorological observations, radiosondes with balloons or aircraft, GPS observations from continuously operating GPS arrays (cf. [7.8.1]), and *water vapor radiometers*. Space-based systems are still in a research and development stage, and use the *radio occultation method* in conjunction with GPS receivers on low Earth orbiters, cf. [7.6.2.9], Melbourne et al. (1994a).

For recent developments see Nothnagel (2000) or Schüler (2001). Satellites designed for Earth observation, like ERS-1/2, ENVISAT [4.3.2], [9.2] carry, besides other sensors, also a radiometer. The radiometer data can be used for the tropospheric delay correction of altimeter measurements (cf. [9.3.3]).

When the highest accuracy is required for ground-based observations, e.g. in the use of VLBI [11.1.2] or GPS [7.6.2.2] for geodynamic modeling, attempts can be made to measure the water vapor content directly along the signal propagation path with a *water vapor radiometer*. The development of such devices began in about 1980 and has now reached a certain level of maturity (Nothnagel, 2000). Besides stationary instruments, e.g. in connection with VLBI antennas (Elgered, Jarlemark, 1998), systems are also available for field applications (Bürki, Kahle, 1995). Fig. 2.24 shows a portable dual frequency microwave water vapor radiometer developed for geodetic applications at the Technical University (ETH) Zürich (Switzerland). The instrument operates at 23.8 and 31.5 GHz and is capable of automatically tracking space targets like GPS satellites. The accuracy estimate for the determination of the signal path delay is about 2 mm.
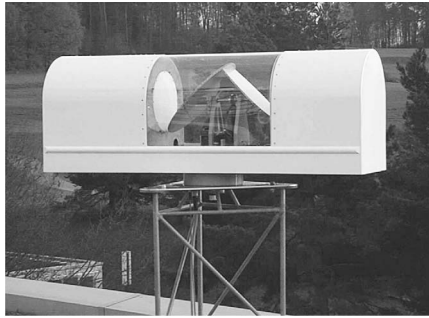
Figure 2.24. Portable dual frequency microwave water vapor radiometer (WVR2000); courtesy ETH Zürich

In stable meteorological conditions the water vapor content of the air shows a high regional correlation over horizontal distances to about 50 km. In such cases the biases are nearly identical at adjacent stations, and they cancel through differencing.

The propagation delay caused by the troposphere is nearly identical for the total spectrum of visible light and for the radio frequency domain. Due to the wet component, however, the absorption is much greater for visible light.

3 Satellite Orbital Motion

Precise time-dependent satellite positions in a suitable reference frame are required for nearly all tasks in satellite geodesy. The computation and prediction of precise satellite orbits, together with appropriate observations and adjustment techniques is, for example, essential for the determination of

- geocentric coordinates of observation stations [12.1],
- field parameters for the description of the terrestrial gravity field as well as for the determination of a precise and high resolution geoid [12.2]
- trajectories of land-, sea-, air-, and space-vehicles in real-time navigation [12.3]
- Earth's orientation parameters in space [12.4].

Essentially, the accuracy of the final results depends on the accuracy of the available satellite orbits. This is increasingly true for tasks in applied geodesy, such as the determination of relative coordinates with the Global Positioning System [7]. The requirement for 1 cm relative accuracy in coordinates implies the requirement for the knowledge of satellite orbits on the few meter accuracy level or even better [7.4.3].

Those who apply satellite methods in geodesy, navigation and adjacent fields, must have a basic knowledge of satellite orbital motion, including the major perturbations, in order to assess the appropriate requirements for orbit determinations. Chapter 3 aims to provide this basic knowledge. Starting with the undisturbed Keplerian motion in a central force field [3.1] the major perturbations, as well as an elementary perturbation theory are discussed [3.2]. The effects of perturbations on satellite orbits are also treated [3.2.4]. A section on the integration and representation of orbits [3.3] follows because algorithms for orbit improvement are included in modern software packages for applied satellite geodesy. The appropriate use of satellite ephemerides is discussed together with the corresponding observation methods (e.g. [7.1.5]).

3.1 Fundamentals of Celestial Mechanics, Two-Body Problem

In celestial mechanics we are concerned with motions of celestial bodies under the influence of mutual mass attraction. The simplest form is the motion of two bodies (*two-body problem*). For artificial satellites the mass of the smaller body (the satellite) usually can be neglected compared with the mass of the central body (Earth). The two-body problem can be formulated in the following way:

> *Given at any time the positions and velocities of two particles of known mass moving under their mutual gravitational force calculate their positions and velocities at any other time.*

Under the assumption that the bodies are homogeneous and thus generate the gravitational field of a point mass the orbital motion in the two-body problem can be

described empirically by *Kepler's laws* [3.1.1]. It can also be derived analytically from *Newtonian* mechanics [3.1.2].

The two-body problem is one of the few problems in celestial mechanics that has a complete solution. Other subjects of celestial mechanics are the three-body and the multi-body problem, i.e. motions of three and more celestial bodies under the influence of their mutual gravitation. These problems have no general solution.

Orbit perturbations [3.2], orbit determination [3.3] and ephemeris computation are also treated in celestial mechanics. Orbit determination refers to orbital parameters derived from observations [3.3.1]. Ephemeris computation refers to geocentric or topocentric positions of celestial bodies or artificial satellites that are derived from orbital elements (e.g. [3.3.3], [7.1.5]).

Modern celestial mechanics has its origin in the year 1687 with the publication of Isaac Newton's *Principia* (Philosophiae naturalis principia mathematica). Herein the law of gravitation and the laws of motion are described for the first time. In the subsequent 300 years there were no major revolutions in celestial mechanics. Only the launch of the first artificial satellite and the development of powerful computers gave an impetus for new ideas. Besides the classical observation of directions, the measurements of ranges and range-rates can now be made.

Also, the influences of Earth's anomalous gravitational field and non-gravitational forces have to be modeled in addition to the classical perturbations caused by the Sun, the Moon and the planets. Through the development of high speed computers large amounts of data can be processed, and numerical integration methods can be used.

Comprehensive textbooks are available for a detailed study of problems and methods in celestial mechanics, such as Stumpff (1959/1965/1974), Brouwer, Clemence (1961), Kovalevsky (1971); Kovalevsky et al. (1989), Schneider (1981, 1993), Taff (1985), Vinti (1998). Easily readable introductions with special regard to satellite and rocket orbits are Escobal (1965), Bate et al. (1971), Roy (1978), Chobotov (1991), Logsdon (1998), and Montenbruck, Gill (2000). Suitable references with particular emphasis on GPS orbits are Rothacher (1992), Yunck (1996), and Beutler et al. (1998).

3.1.1 Keplerian Motion

Johannes Kepler (1571–1630) formulated the three laws of planetary motion associated with his name from an empirical study of observational data collected by *Tycho Brahe* (1546–1601), an astronomer who mainly worked in Denmark. The three laws give a description of the planetary motion but not an explanation. They provide a very good approximation to the real motion within the solar system because the planetary masses can be neglected when compared to the mass of the sun, and because of the fact that the sun can be considered a point mass due to the large distances involved. This is why the undisturbed gravitational motion of point masses is also called *Keplerian motion*. From a historical point of view it may be of interest that Kepler, through his three laws, provided the major breakthrough for Copernicus's heliocentric hypothesis. In the following, Kepler's laws of planetary motion are introduced and explained.

1st Law: *The orbit of each planet is an ellipse with the Sun at one focus*

The orbital geometry is defined by this law. The usual relations and symbols are shown in Fig. 3.1. The major axis in the ellipse, $\overline{A\,\pi}$, is called the *line of apsides*. The orbital point A, farthest from the center of mass of the orbital system, 0, is named the *apocenter*. The point π on the orbit, closest to the center, is named *pericenter*. When 0 is the center of the sun, A and π are called *aphelion* and *perihelion* respectively. When 0 is identical with Earth's center of mass, then A and π are named *apogee* and *perigee*. The angle v is the *true anomaly*.
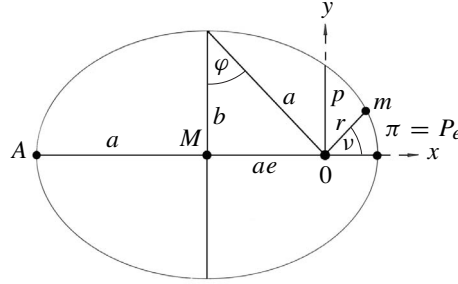
Figure 3.1. Geometry of the orbital ellipse

The Keplerian motion refers to a plane. The orbital plane can be used for the definition of a coordinate system with 0 being the origin. The location of a point mass m may be described with polar coordinates r, v, when $\overline{0\,\pi}$ is selected as one axis of the orbital coordinate system. With

r,   distance of the point mass m from the center of the primary mass,
v,   true anomaly,
a,   semi-major axis,
e,   numerical eccentricity, and
p,   parameter of the ellipse,

we find for the equation of the elliptical curve

$$r = \frac{p}{1 + e \cos v}.$$ (3.1)

Equation (3.1) also gives the mathematical form of Kepler's first law. From the elliptical geometry other relations follow:

$$p = \frac{b^2}{a}; \quad e = \sqrt{1 - \frac{b^2}{a^2}}; \quad a = \frac{p}{1 - e^2}; \quad b = \frac{p}{\sqrt{1 - e^2}}.$$ (3.2)

$a \cdot e$ is the *linear eccentricity* and gives the distance of the focal points from the orbital center. For $e = 0$, it follows

$$a \;=\; b \;=\; p,$$

i.e. the ellipse evolves to a circle. The *eccentricity angle* φ may be used instead of e. The following relations are then valid:

$$
\begin{aligned}
e &= \sin \varphi &\qquad p &= a \cos^2 \varphi \\
\sqrt{1 - e^2} &= \cos \varphi &\qquad b &= a \cos \varphi = p \sec \varphi.
\end{aligned}
$$ (3.3)

2nd Law: *The line from the Sun to any planet sweeps out equal areas of space in equal lengths of time*

Kepler's second law, also called the *Law of Areas*, describes the velocity of a planet in its orbit. With this law it is possible to determine the location of a planet as a function of time with polar coordinates r and v.
According to Fig. 3.2 the formula

$$\Delta F \approx \frac{1}{2} r^2 \Delta v$$

is approximately valid for the area of an infinitesimal triangle 0, P, P'. According to the second law the area ΔF swept by r is proportional to the corresponding time interval Δt, thus

$$r^2 \Delta v \approx c \cdot \Delta t$$

with c being a constant. In terms of differential relations, we find

Figure 3.2. Satellite motion governed by Kepler's second law

$$r^2 \frac{dv}{dt} = c. \tag{3.4}$$

Equation (3.4) is the mathematical expression of the law of areas. Kepler actually found it earlier than his first law (3.1).

Further relations can be deduced. Introducing rectangular coordinates x, y into Fig. 3.1 we get

$$x = r \cdot \cos v, \quad y = r \cdot \sin v, \quad r = \sqrt{x^2 + y^2} \tag{3.5}$$

and

$$\tan v = \frac{y}{x}. \tag{3.6}$$

The derivative of (3.6) with respect to time leads to

$$\frac{\dot{v}}{\cos^2 v} = \frac{x\dot{y} - y\dot{x}}{x^2}. \tag{3.7}$$

Substituting $\cos v$ in (3.7) by (3.5) and introducing (3.7) into (3.4) gives an alternative representation of Kepler's second law using rectangular coordinates:

$$x\dot{y} - y\dot{x} = c. \tag{3.8}$$

c is called the *constant of areas*. Equation (3.8) will also be derived using Newtonian mechanics [3.1.2].

3rd Law: *The cubes of semi-major axes of the planetary orbits are proportional to the squares of the planet's periods of revolution*

In mathematical formulation this means for different planets P_i with periods of revolution U_i, mean motions:

$$n_i = 2\pi/U_i, \tag{3.9}$$

and semi-major axes a_i:

$$\frac{a_i^3}{U_i^2} = \frac{C^2}{4\pi^2}. \tag{3.10}$$

C is a constant for the planetary system. Inserting (3.9) into (3.10) gives the commonly used expression

$$a_i^3 \cdot n_i^2 = C^2. \tag{3.11}$$

This law was found empirically by Kepler because it approximates very well the motion of the large planets. However, a completely different numerical value for C^2 is obtained for the system of Jovian moons. Therefore a more general relation is useful

$$\frac{a^3}{U^2} = \frac{k^2}{4\pi^2}(M + m), \tag{3.12}$$

where k is a universal constant and M, m are the respective masses. Using (3.12) it is possible to determine the masses of celestial bodies.

Kepler's laws describe the simplest form of motion of celestial bodies under the assumption that no external perturbing forces are present, and that the respective masses can be considered to be point masses or homogeneous bodies with spherical mass distribution. For the motion of an artificial Earth satellite these assumptions are only valid in a first approximation. Keplerian orbits, consequently, can only be used as a simple reference orbit and they give only qualitative information on the kind of motion. Kepler himself was convinced that his three, empirically found, laws followed a more general law. This more general law was formulated by *Isaac Newton* (1643–1727) in the form of the *Law of Gravitation*.

3.1.2 Newtonian Mechanics, Two-Body Problem

3.1.2.1 Equation of Motion

In the first book of "Principia" Newton introduced his three laws of motion:

1. *Every body continues in its state of rest or of uniform motion in a straight line unless it is compelled to change that state by an external impressed force.*

2. *The rate of change of momentum of the body is proportional to the force impressed and is in the same direction in which the force acts.*

3. *To every action there is an equal and opposite reaction.*

The second law expressed in mathematical form is

$$\boldsymbol{K} = m\ddot{\boldsymbol{r}} \tag{3.13}$$

where $\boldsymbol{K}$ is the vector sum of all forces acting on the mass m and $\ddot{\boldsymbol{r}}$ is the vectorial acceleration of the mass, measured in an inertial reference frame.

In addition we find Newton's famous *law of universal gravitation* (1687) in the third book, section I of the "Principia"

> *Every particle of matter in the Universe attracts every other particle of matter with a force directly proportional to the product of their masses and inversely proportional to the square of the distance between them*

$$K = -G\frac{Mm}{r^2}. \tag{3.14}$$

M and m are two particles of matter and G is the universal constant of gravitation with (Torge, 2001)

$$G = (6.67259 \pm 0.00085) \cdot 10^{-11} \mathrm{m}^3 \mathrm{kg}^{-1} \mathrm{s}^{-2}. \tag{3.15}$$

Notice that the law of gravitation in Newton's text is not written in the above closed form, but it can be taken from a somewhat broader explanation.

Within a Cartesian coordinate system with the axes x, y, z and with α, β, γ being the angles between the direction of force and the respective axes of the system we find from (3.13) for the motion of M with respect to m, expressed in components (Fig. 3.3)
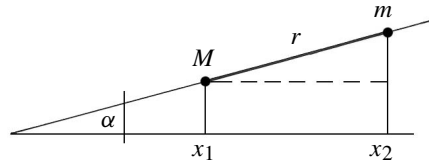
$$M\ddot{x}_1 = K_x = K \, \cos\alpha$$

Figure 3.3. Components within Newton's equation of motion

and after substituting (3.14)

$$M\ddot{x}_1 = -G\frac{Mm}{r^2}\cos\alpha = -G\frac{Mm}{r^3}(x_1 - x_2).$$

After rearrangement we obtain for all three components:

$$M\ddot{x}_1 = G\frac{Mm}{r^3}(x_2 - x_1), \quad M\ddot{y}_1 = G\frac{Mm}{r^3}(y_2 - y_1), \quad M\ddot{z}_1 = G\frac{Mm}{r^3}(z_2 - z_1). \tag{3.16}$$

For the motion of m with respect to M:

$$m\ddot{x}_2 = -G\frac{Mm}{r^3}(x_2 - x_1), \quad m\ddot{y}_2 = -G\frac{Mm}{r^3}(y_2 - y_1), \quad m\ddot{z}_2 = -G\frac{Mm}{r^3}(z_2 - z_1). \tag{3.17}$$

Transferring the origin of the coordinate system to the center of mass M; using the substitutions

$$x_2 - x_1 = x; \qquad y_2 - y_1 = y; \qquad z_2 - z_1 = z,$$

dividing (3.16) by M and (3.17) by m, and then subtracting (3.16) from (3.17):

$$\ddot{x} = -G(M+m)\frac{x}{r^3}, \quad \ddot{y} = -G(M+m)\frac{y}{r^3}, \quad \ddot{z} = -G(M+m)\frac{z}{r^3} \quad (3.18)$$

where $r^2 = x^2 + y^2 + z^2$. In vector form, (3.18) becomes

$$\ddot{\boldsymbol{r}} = \frac{d^2\boldsymbol{r}}{dt^2} = -G\frac{M+m}{r^3}\boldsymbol{r}. \quad (3.19)$$

For artificial Earth satellites the mass m can be neglected. The basic equation of satellite motion is then

$$\ddot{\boldsymbol{r}} = -\frac{GM}{r^3}\boldsymbol{r} \quad (3.20)$$

with $\boldsymbol{r}$ being the geocentric position vector of the artificial satellite.

Equation (3.20) is the vector form of a second order differential equation with six integration constants. In other words, the motion of a celestial body around its central body, governed by the mutual gravitation, has six independent parameters. Usually the six *Keplerian orbital parameters* (Fig. 3.4) are used. The equation of motion (3.20)

a semi-major axis
e numerical eccentricity
i orbit inclination
Ω right ascension of ascending node
ω argument of perigee
v true anomaly
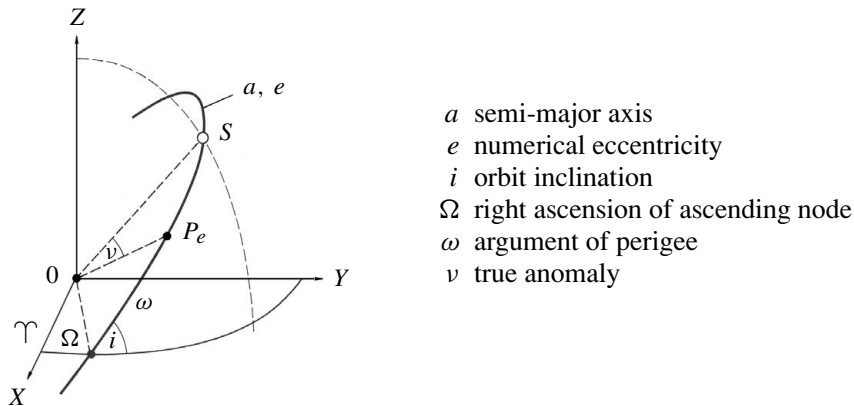
Figure 3.4. Keplerian orbital parameters

was derived under the assumption that only gravitational forces are present, that the mass of the satellite can be neglected, and that the central mass can be considered as a point mass. This is, in fact, not correct; in particular, the inhomogeneous structure of Earth is acting on the motion of the near-Earth satellite. As a consequence, the Keplerian orbit (3.20) can only be considered to be a first approximation to the true satellite orbit. The implications of the perturbing forces are treated in [3.2].

The integration of equation (3.20) leads to some important insights. Among others it is possible to derive the laws, which Kepler had found empirically, from Newton's

laws of gravitation and motion in a rigorous way. Some elements are presented in the sections that follow. A comprehensive treatment of the subject may be found in textbooks on celestial mechanics, such as Stumpff (1959/1965/1974), Brouwer, Clemence (1961), Schneider (1981, 1993), Taff (1985), Vinti (1998).

3.1.2.2 Elementary Integration

Multiplication of (3.18) by x, y, z respectively and forming pairs of differences yields

$$x\ddot{y} - y\ddot{x} = 0, \quad y\ddot{z} - z\ddot{y} = 0, \quad z\ddot{x} - x\ddot{z} = 0. \tag{3.21}$$

Integration of (3.21) gives:

$$x\dot{y} - y\dot{x} = C_1, \quad y\dot{z} - z\dot{y} = C_2, \quad z\dot{x} - x\dot{z} = C_3, \tag{3.22}$$

where C_1, C_2, C_3 are arbitrary constants. Multiplying the equations one after the other with z, x, y and forming the total sum cancels out the left hand terms to give

$$C_1 z + C_2 x + C_3 y = 0. \tag{3.23}$$

This is the equation of a plane containing the origin of the coordinate system. We can state that the satellite or planet is moving in a plane which contains the center of the central body. The orientation of the orbital plane in space can be specified using two parameters, for instance i and Ω as defined in Fig. 3.4. The relation between i, Ω and the constants C_1, C_2, C_3 is given by Roy (1978, p. 103 f), Montenbruck, Gill (2000, p. 28):

$$\frac{C_1}{N} = \cos i, \quad \frac{C_2}{N} = \sin \Omega \sin i \quad \text{and} \quad \frac{C_3}{N} = -\cos \Omega \sin i, \tag{3.24}$$

with $N = \sqrt{C_1^2 + C_2^2 + C_3^2}$ being the normal to the orbital plane.

Since the motion occurs in a plane, a rectangular plane coordinate system ξ, η can be introduced with the origin at the mass center of the central body (Fig. 3.5). The equations of motion, corresponding to (3.20) and described in components, hold

$$\ddot{\xi} = -GM\frac{\xi}{r^3}; \quad \ddot{\eta} = -GM\frac{\eta}{r^3} \tag{3.25}$$

with $r^2 = \xi^2 + \eta^2$. The equivalent of (3.21) follows

$$\xi\ddot{\eta} - \eta\ddot{\xi} = 0 \tag{3.26}$$
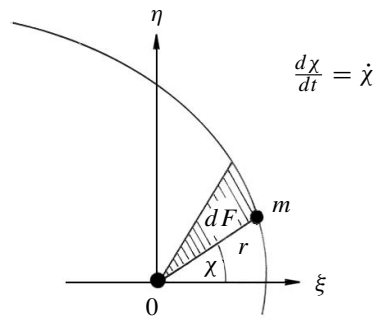
Figure 3.5. Illustration of the law of areas

which, after integration becomes

$$\xi\dot{\eta} - \eta\dot{\xi} = p_1. \tag{3.27}$$

Substituting (3.27) with polar coordinates

$$\xi = r\cos\chi \quad \text{and} \quad \eta = r\sin\chi \tag{3.28}$$

gives

$$r^2\dot{\chi} = p_1. \tag{3.29}$$

Considering an infinitesimal small area dF, being swept by r during the infinitesimal time interval dt (Fig. 3.5), we find for the area of the infinitesimal triangle:

$$dF = \frac{1}{2}r^2\dot{\chi}\,dt$$

$$\frac{dF}{dt} = \frac{1}{2}r^2\dot{\chi} = \frac{1}{2}p_1 \tag{3.30}$$

hence

$$F = \frac{1}{2}p_1 t + p_2. \tag{3.31}$$

The equations (3.30) and (3.31) contain in essence Kepler's second law. Hence it can be stated:

 − the motion takes place within a plane, and
 − the motion is governed by the law of areas.

What is missing is a statement about the shape of the orbit. Multiplying equations (3.25) by $2\dot{\xi}$ and $2\dot{\eta}$, respectively, gives

$$\ddot{\xi}\,2\dot{\xi} = -GM\frac{\xi}{r^3}2\dot{\xi}, \quad \ddot{\eta}\,2\dot{\eta} = -GM\frac{\eta}{r^3}2\dot{\eta} \tag{3.32}$$

and summing (3.32) gives

$$\frac{d}{dt}(\dot{\xi}^2 + \dot{\eta}^2) = -\frac{2\,GM}{r^3}(\xi\dot{\xi} + \eta\dot{\eta}). \tag{3.33}$$

With $r^2 = \xi^2 + \eta^2$

$$2\,r\dot{r} = 2\,\xi\dot{\xi} + 2\,\eta\dot{\eta}$$

it follows that

$$\frac{d}{dt}(\dot{\xi}^2 + \dot{\eta}^2) = -\frac{2\,GM}{r^2}\dot{r} = 2GM\frac{\dot{1}}{r} \tag{3.34}$$

which, after integration becomes

$$\dot{\xi}^2 + \dot{\eta}^2 = 2\frac{GM}{r} + p_3. \tag{3.35}$$

With polar coordinates (3.28) and after differentiation we get

$$\dot{r}^2 + r^2 \dot{\chi}^2 = \frac{2\,GM}{r} + p_3.$$

(3.36)

One solution of this differential equation is (Brouwer, Clemence, 1961)

$$r = \frac{p}{1 + e\cos(\chi - \varpi)}$$

(3.37)

where p, e, ϖ are constants. Equation (3.37) is the polar form of a conic section.

For $\chi = \varpi$ the satellite distance r becomes a minimum, i.e. the satellite passes through the perigee. Since the angular distance of a satellite from the perigee was introduced as the true anomaly v (3.4), the identity

$$\chi - \varpi = v$$

is valid. If $v = 90°$, then $r = p$. Further, we know from ellipsoidal geometry (3.2):

$$p = a(1 - e^2).$$

Now the integration constants p, e, ϖ can be expressed geometrically and (3.37) can be rewritten as

$$r = \frac{p}{1 + e\cos v}.$$

(3.38)

If the *argument of the latitude*

$$u = \omega + v$$

is measured from the ascending node of the satellite orbit, the origin of the angular measurement can be fixed. An alternative formulation of (3.38) is

$$r = \frac{p}{1 + e\cos(u - \omega)},$$

(3.39)

where ω is known as the *argument of the perigee* (3.4). With these substitutions we have dealt with five of the six integration constants, namely

$$\Omega, i, \omega, e, a.$$

The last free constant is the quantity p_2 in equation (3.31), Kepler's law of areas, which specifies the time-dependent position of the satellite in its orbit. Several equivalent parameters are in use, among others the time of transit through the perigee t_0 or the true anomaly v.

The following relations can be established between Kepler's orbital elements and the integration constants (e.g. Arnold, 1970; Brouwer, Clemence, 1961):

$$p = \frac{p_1^2}{GM}; \quad e^2 = 1 + \frac{p_1^2 p_3}{G^2 M^2}; \quad p_1 = \sqrt{GMp}; \quad p_3 = \frac{-GM}{a}.$$

(3.40)

With (3.31) and the use of (3.40) Kepler's third law can be deduced. The period of one satellite revolution is $T = t_2 - t_1$ and thus we obtain for the area swept after one revolution

$$F_2 - F_1 = \frac{1}{2} p_1 (t_2 - t_1) = \pi ab \tag{3.41}$$

i.e. the area of an ellipse. With

$$p_1 = \sqrt{GMp}; \quad b^2 = a^2 (1 - e^2); \quad p = a(1 - e^2)$$

we obtain after some rearrangement

$$T = \frac{2\pi}{\sqrt{GM}} a^{\frac{3}{2}}. \tag{3.42}$$

With the mean angular motion

$$n = \frac{2\pi}{T}$$

follows the mathematical expression for Kepler's third law:

$$n^2 \cdot a^3 = GM. \tag{3.43}$$

With this, we have derived Kepler's three laws using only Newton's basic equations (3.13) and (3.14). Using equation (3.35), another important relation can be deduced. Substituting (3.40) for p_3 gives the velocity of a satellite in its orbit:

$$v^2 = \dot{\xi}^2 + \dot{\eta}^2 = 2\frac{GM}{r} - \frac{GM}{a}$$

$$v^2 = GM \left(\frac{2}{r} - \frac{1}{a} \right). \tag{3.44}$$

Equation (3.36), with $\dot{\chi} = \dot{v}$, gives the following equation in polar coordinates:

$$v^2 = \dot{r}^2 + r^2 \dot{v}^2 = GM \left(\frac{2}{r} - \frac{1}{a} \right). \tag{3.45}$$

Equation (3.45) is well known as the *energy integral* and demonstrates that the velocity of a celestial body depends on the distance r and the semi-major axis a, but not on the eccentricity and thus not on the form of the orbit, cf. (3.61).

Furthermore equation (3.29), substituting (3.40) and (3.2) for p_1, leads to another form of Kepler's second law:

$$r^2 \dot{v} = \sqrt{GMa(1 - e^2)}. \tag{3.46}$$

By replacing $\dot{v}$ in (3.45) with (3.46) we obtain

$$\dot{r}^2 + r^2 \frac{GMa(1 - e^2)}{r^4} = GM \left(\frac{2}{r} - \frac{1}{a} \right). \tag{3.47}$$

Substituting (3.43) into (3.47) and rearranging gives (Brouwer, Clemence, 1961)

$$ndt = \frac{r}{a} \frac{dr}{\sqrt{a^2 e^2 - (r - a)^2}}. \tag{3.48}$$

The geometrical relation from Fig. 3.6

$$r = a(1 - e \cos E) \qquad (3.49)$$

is put into (3.48)

$$ndt = (1 - e \cos E)dE. \qquad (3.50)$$

Integration gives

$$n(t - t_0) = E - e \sin E. \qquad (3.51)$$

The variable E is called the *eccentric anomaly*. The integration constant t_0 can be considered to be the epoch of transit through the perigee. The lefthand side of (3.51) increases linearly with time. Instead of t a new variable $\overline{M}$, the *mean anomaly*, can be defined as

Figure 3.6. Relationship between true anomaly v and eccentric anomaly E

$$\overline{M} = n(t - t_0). \qquad (3.52)$$

The form

$$\overline{M} = E - e \sin E \qquad (3.53)$$

is called *Kepler's equation*. The relationship with the *true anomaly* v is given by

$$\tan v = \frac{\sqrt{1 - e^2} \sin E}{\cos E - e}. \qquad (3.54)$$

All three anomalies $E, \overline{M}, v$ are zero when the satellite passes through the perigee. They can be used to determine the satellite position within the orbit and hence each is suitable as a sixth orbital parameter.

In satellite geodesy the mean anomaly $\overline{M}$ is usually given preference because it can be interpolated linearly with time. In order to compute E from $\overline{M}$ equation (3.53) has to be transformed into an elliptical series expansion. Many solutions can be found in the literature. One example is (Brouwer, Clemence (1961, p. 76), Taff (1985, p. 61)):

$$\begin{aligned}
E = \overline{M} &+ \left(e - \frac{1}{8}e^3 + \frac{1}{192}e^5 - \frac{1}{9216}e^7 \right) \sin \overline{M} \\
&+ \left(\frac{1}{2}e^2 - \frac{1}{6}e^4 + \frac{1}{48}e^6 \right) \sin 2\overline{M} \\
&+ \left(\frac{3}{8}e^3 - \frac{27}{128}e^5 + \frac{243}{5120}e^7 \right) \sin 3\overline{M} \\
&+ \left(\frac{1}{3}e^4 - \frac{4}{15}e^6 \right) \sin 4\overline{M} + \left(\frac{125}{384}e^5 - \frac{3125}{9216}e^7 \right) \sin 5\overline{M} \\
&+ \frac{27}{80}e^6 \sin 6\overline{M} - \frac{16807}{46080}e^7 \sin 7\overline{M} + \cdots .
\end{aligned} \qquad (3.55)$$

For small eccentricities, the following iteration yields a very fast solution:

$$
\begin{aligned}
E_0 &= \overline{M} \\
E_i &= \overline{M} + e \sin E_{i-1}, \qquad i = 1 \dots .
\end{aligned}
\tag{3.56}
$$

3.1.2.3 Vectorial Integration

Starting with the equation of motion (3.20) it is possible to derive the basic theorem of the two-body motion in an explicit and elementary way. On the other hand the use of vector operations gives a rather transparent and closed representation (e.g. Bate et al., 1971; Taff, 1985). Basic equations for energy, angular momentum and orbital characteristics can be easily derived. Dot multiplication of equation (3.20)

$$
\ddot{\boldsymbol{r}} + \frac{GM}{r^3}\boldsymbol{r} = 0
$$

with the velocity vector $\dot{\boldsymbol{r}} = \boldsymbol{v}$ yields

$$
\dot{\boldsymbol{r}} \cdot \ddot{\boldsymbol{r}} + \dot{\boldsymbol{r}} \cdot \frac{GM}{r^3}\boldsymbol{r} = 0
\tag{3.57}
$$

or

$$
\boldsymbol{v} \cdot \dot{\boldsymbol{v}} + \frac{GM}{r^3}\boldsymbol{r} \cdot \dot{\boldsymbol{r}} = 0.
\tag{3.58}
$$

Since in general $\boldsymbol{a} \cdot \dot{\boldsymbol{a}} = a\dot{a}$, we find

$$
v\dot{v} + \frac{GM}{r^3}r\dot{r} = 0.
\tag{3.59}
$$

Substituting

$$
\frac{d}{dt}\left(\frac{v^2}{2}\right) = v\dot{v} \quad \text{and} \quad \frac{d}{dt}\left(-\frac{GM}{r}\right) = \frac{GM}{r^2}\dot{r}
$$

into (3.59), it follows that

$$
\frac{d}{dt}\left(\frac{v^2}{2}\right) + \frac{d}{dt}\left(-\frac{GM}{r}\right) = 0
$$

or

$$
\frac{d}{dt}\left(\frac{v^2}{2} - \frac{GM}{r}\right) = 0.
\tag{3.60}
$$

Integration of (3.60) results in a constant E_M, the *energy integral*,

$$
E_M = \frac{v^2}{2} - \frac{GM}{r}
\tag{3.61}
$$

which we know already from equation (3.45) in a different notation. The first term on the righthand side is the kinetic energy of the satellite with unit mass (m = 1),

and the second term its potential energy, $-\frac{GM}{r}$. Hence, the total mechanical energy of the satellite motion is constant, which is to be expected when no external forces are present. The negative sign for the potential energy comes from choosing its zero reference at infinity. Through this procedure a free integration constant, inherent in equation (3.60), is fixed.

Cross multiplication of equation (3.20) with r leads to

$$\boldsymbol{r} \times \ddot{\boldsymbol{r}} + \boldsymbol{r} \times \frac{GM}{r^3}\boldsymbol{r} = 0. \qquad (3.62)$$

The cross product of a vector with itself is zero, hence

$$\boldsymbol{r} \times \ddot{\boldsymbol{r}} = 0. \qquad (3.63)$$

With

$$\frac{d}{dt}(\boldsymbol{r} \times \dot{\boldsymbol{r}}) = \dot{\boldsymbol{r}} \times \dot{\boldsymbol{r}} + \boldsymbol{r} \times \ddot{\boldsymbol{r}}$$

it follows that

$$\frac{d}{dt}(\boldsymbol{r} \times \dot{\boldsymbol{r}}) = 0$$

and after integration

$$\boldsymbol{h} = \boldsymbol{r} \times \dot{\boldsymbol{r}} = \boldsymbol{r} \times \boldsymbol{v}. \qquad (3.64)$$

$\boldsymbol{h}$ is a constant vector, perpendicular to the plane containing $\boldsymbol{r}$ and $\boldsymbol{v}$. The physical interpretation of equation (3.64) is that the angular momentum of a satellite with unit mass (m $=$ 1) remains constant along its trajectory. With $\boldsymbol{h}$ constant the satellite motion must occur within a space-fixed plane. The orientation of this orbital plane can be described through the elements Ω and i, which have been introduced before. The relation is given by

$$\boldsymbol{h}_0 = \begin{pmatrix} \sin i & \sin \Omega \\ -\sin i & \cos \Omega \\ \cos i \end{pmatrix}; \quad \boldsymbol{h}_0 = \frac{\boldsymbol{h}}{|\boldsymbol{h}|}. \qquad (3.65)$$

At the same time $\boldsymbol{h}$ defines the direction of motion of the satellite. From the definition of the cross product we find for the magnitude of $\boldsymbol{h}$, according to Fig. 3.7

$$h = rv \cos \Phi. \qquad (3.66)$$

The *flight-path angle* Φ indicates the direction of the orbital motion.

A further vector multiplication of (3.20) with $\boldsymbol{h}$ leads to the form of the satellite orbit:

$$\ddot{\boldsymbol{r}} \times \boldsymbol{h} = \frac{GM}{r^3}(\boldsymbol{h} \times \boldsymbol{r}). \qquad (3.67)$$
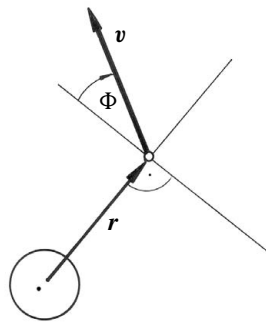
Figure 3.7. Flight-path angle

It can be easily proved that the lefthand side of (3.67) equals $\frac{d}{dt}(\dot{r} \times h)$. For the righthand side we find, with (3.65), and $r \cdot \ddot{r} = r \cdot \dot{r}$

$$
\begin{aligned}
\frac{GM}{r^3}(h \times r) &= \frac{GM}{r^3}(r \times v) \times r \\
&= \frac{GM}{r^3}[v(r \cdot r) - r(r \cdot v)] = \frac{GM}{r}v - \frac{GM\dot{r}}{r^2}r.
\end{aligned}
\tag{3.68}
$$

Furthermore,

$$
GM\frac{d}{dt}\left(\frac{r}{r}\right) = \frac{GM}{r}v - \frac{GM \cdot \dot{r}}{r^2}r.
$$

Thus, (3.67) can be re-formulated as

$$
\frac{d}{dt}(\dot{r} \times h) = GM\frac{d}{dt}\left(\frac{r}{r}\right).
\tag{3.69}
$$

The integration of both sides gives

$$
\dot{r} \times h = GM\frac{r}{r} + B.
\tag{3.70}
$$

The vector integration constant B is a linear combination of $\dot{r} \times h$ and r, and consequently it lies in the orbital plane. B is directed toward the perigee, as can easily be verified (e.g. Bate et al., 1971).

Dot multiplication of both sides of (3.70) creates a scalar-type equation (because $a \cdot b \times c = a \times b \cdot c$ and $a \cdot a = a^2$):

$$
\begin{aligned}
r \cdot (\dot{r} \times h) &= r \cdot GM\frac{r}{r} + r \cdot B \\
h^2 &= GMr + r|B| \cos v.
\end{aligned}
\tag{3.71}
$$

v is the angle between the constant vector B and the radius vector r. Solving for r gives

$$
r = \frac{\frac{h^2}{GM}}{1 + \frac{|B|}{GM} \cos v}.
\tag{3.72}
$$

Introducing

$$
p = \frac{h^2}{GM} \quad \text{and} \quad e = \frac{|B|}{GM}
\tag{3.73}
$$

it follows the already known equation of a conic section (3.38):

$$
r = \frac{p}{1 + e \cos v}.
$$

With the three basic integrals

> *Integral of energy*   (3.61)
>
> *Integral of momentum*   (3.64)
>
> *Integral of the orbit*   (3.72)

we can summarize our knowledge concerning orbital motion:

(1) The family of curves called *conic sections* (circle, ellipse, parabola, hyperbola) represent the only possible paths for an orbiting object in the two-body problem.

(2) The focus of the conic orbit must be located at the center of the central body.

(3) The sum of kinetic and potential energy does not change as the satellite moves along its conic orbit, which means that the satellite must slow down as it gains altitude and speed up as r decreases, in such a manner that the energy sum remains constant.

(4) The orbital motion takes place in a plane which is fixed in inertial space.

(5) The angular momentum of a satellite about the central body remains constant.

The reader should note that in the literature the basic integrals listed above are also called *constants of the motion*.

3.1.3   Orbit Geometry and Orbital Motion

Some fundamentals of orbit geometry have already been introduced. Table 3.1 summarizes the most important properties of the four conic sections which serve as possible orbits. One characteristic quantity is the parameter p, which is also called the *semi-latus rectum*.

Table 3.1. Characteristic properties of conic sections

		circle	ellipse	parabola	hyperbola
eccentricity	e	0	$0 < e < 1$	1	$e > 1$
parameter	p	a	$a(1 - e^2)$	p	$a(e^2 - 1)$
semi-major axis	a	a	a	∞	$a < 0$
semi-minor axis	b	a	$a\sqrt{1 - e^2}$	—	$a\sqrt{e^2 - 1}$
pericenter distance	r_p	a	$a(1 - e)$	$p/2$	$a(e - 1)$
apocenter distance	r_a	a	$a(1 + e)$	∞	∞

Comparing (3.72) with (3.38) demonstrates that the parameter p only depends on the angular momentum h

$$p = \frac{h^2}{GM}. \tag{3.74}$$

This is clarified in Fig. 3.8 (Bate et al., 1971), where it is supposed that a cannonball is fired horizontally from a cannon in an atmosphere-free environment. From equation (3.66) we obtain

$$h = rv$$

because the flight path angle equals zero. Therefore, progressively increasing the muzzle velocity v is equivalent to increasing the angular momentum h, generating a

family of curves as is shown in the figure.

Each of the curves represents a conic section with the focal point located at the Earth's center of mass. Furthermore, it is illustrated that the velocity vector at perigee v_p, is horizontal, i.e. perpendicular to the radius vector r_p, as follows from (3.66)

$$h = r_p \, v_p. \qquad (3.75)$$

The same is valid at apogee.

Formulating the energy integral (3.61) at perigee we get, with (3.75),
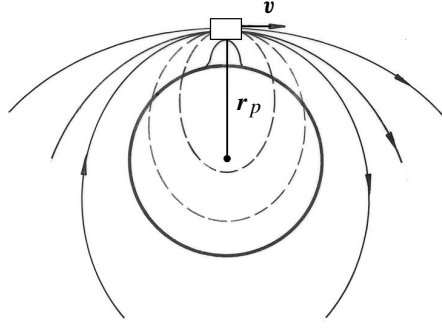
Figure 3.8. Orbital velocity and orbital curve

$$E_M = \frac{v^2}{2} - \frac{GM}{r} = \frac{h^2}{2\,r_p^2} - \frac{GM}{r_p}. \qquad (3.76)$$

With $r_p = a\,(1-e)$ and $p = a\,(1-e^2)$ (from Table 3.1), and with (3.74), it follows that

$$h^2 = GMa(1-e^2)$$

and

$$E_M = \frac{GMa\,(1-e^2)}{2a^2\,(1-e)^2} - \frac{GM}{a\,(1-e)}. \qquad (3.77)$$

This can be reduced to

$$E_M = -\frac{GM}{2\,a}. \qquad (3.78)$$

This simple equation (3.78) is valid for all conic sections and demonstrates that the semi-major axis of a satellite orbit only depends on the total energy. In other words, the energy of the orbital motion is characterized by the size of the semi-major axis. Since, on the other hand, the orbital angular momentum h only depends on p, following (3.74), it can easily be shown that the shape of the orbit, which is characterized through e,

$$e = \sqrt{1 + \frac{2E_M h^2}{(GM)^2}}, \qquad (3.79)$$

is determined both by the orbital energy and the angular momentum. With substitutions from Table 3.1 equation (3.78) develops to

$$
\begin{aligned}
E_M &= -\frac{GM}{2a} &&\text{for the ellipse,} \\
E_M &= 0 &&\text{for the parabola, and} \\
E_M &= +\frac{GM}{2a} &&\text{for the hyperbola.}
\end{aligned}
\qquad (3.80)
$$

For a closed curve (circle and ellipse, $a > 0$), the total energy is negative; it has to be considered that the zero reference is at infinity by convention. All along the orbit, the magnitude of the potential energy exceeds the kinetic energy, thus m remains tied to M. With a positive energy balance, there is kinetic energy left even at infinity, thus the orbiting body will leave the gravity field of the central body. For escape to just take place, the velocity must be changed until the total energy is zero.

Inserting the expressions of (3.80) into the energy equation (3.61), and solving for the velocity v, we obtain

$$v_E^2 = GM \left(\frac{2}{r} - \frac{1}{a} \right) \qquad \text{for the ellipse,}$$

$$v_P^2 = \frac{2GM}{r} \qquad \text{for the parabola, and} \tag{3.81}$$

$$v_H^2 = GM \left(\frac{2}{r} + \frac{1}{a} \right) \qquad \text{for the hyperbola.}$$

These equations, known as *vis-viva-equations*, allow the computation of velocities for every point of the orbit, when a is known. For circular orbits we obtain the simple equation

$$v_C = \sqrt{\frac{GM}{r}}. \tag{3.82}$$

For an elliptic orbit (cf. Table 3.1) the satellite's velocity varies between a maximum at perigee:

$$v_{Pe} = \sqrt{\frac{GM}{a}} \sqrt{\frac{1+e}{1-e}} \tag{3.83}$$

and a minimum at apogee:

$$v_{Ap} = \sqrt{\frac{GM}{a}} \sqrt{\frac{1-e}{1+e}}. \tag{3.84}$$

According to (3.80), Earth's gravity field can be left on a parabolic orbit. The required minimum velocity at the surface (*escape velocity*) is, using (3.81)

$$v_e = \sqrt{\frac{2GM}{r_0}}. \tag{3.85}$$

With $r_0 = 6370$ km and $GM = 398\,600$ km^{3}s^{-2} it follows that

$$v_e = 11.2 \text{ km/s.} \tag{3.86}$$

The period of a satellite revolution in an elliptic orbit was found to be (3.42)

$$T = \frac{2\pi}{\sqrt{GM}} a^{\frac{3}{2}}.$$

Evidently, the period only depends on a, but not on the eccentricity e.

The description of a satellite orbit with the six Keplerian elements $a, e, i, \omega, \Omega, \nu$ is demonstrated in Fig. 3.4. This list is not exhaustive; many other parameters are in use. Instead of the true anomaly ν often the *mean anomaly* $\overline{M}$ or the time of perigee passage is used. The parameter p can be utilized instead of the semi-major axis a. Instead of the argument of perigee ω the following arguments are used (Fig. 3.9):

$$\begin{aligned}
\pi &= \Omega + \omega & &\text{longitude of perigee,} \\
u &= \omega + \nu & &\text{argument of latitude,} \\
l &= \Omega + \omega + \nu = \pi + \nu = \Omega + u. & &\text{true longitude.}
\end{aligned} \tag{3.87}$$

The right ascension Ω is an angle in the equatorial plane; ω and ν are angles in the orbital plane; ω and derived quantities containing ω are not defined for circular orbits.
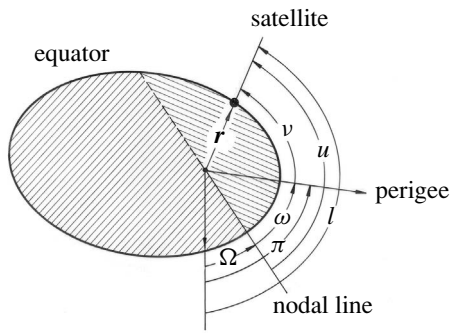
Figure 3.9. Orbital elements for the representation of longitudes
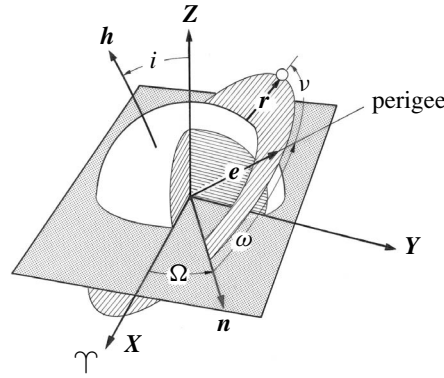
Figure 3.10. The fundamental vectors $\boldsymbol{h}, \boldsymbol{n}, \boldsymbol{e}$

It may be helpful to introduce three fundamental vectors $\boldsymbol{h}, \boldsymbol{n}, \boldsymbol{e}$ (Fig. 3.10) to facilitate transformations. $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ are unit vectors in the equatorial system and h_x, h_y, h_z or n_x, n_y, n_z are the respective vector components of the fundamental vectors. The *angular momentum vector*

$$\boldsymbol{h} = \boldsymbol{r} \times \boldsymbol{v},$$

being perpendicular to the plane of the orbit, is already known from (3.64). The *node vector* is defined as

$$\boldsymbol{n} = \boldsymbol{Z} \times \boldsymbol{h}. \tag{3.88}$$

The *perigee vector*

$$\boldsymbol{e} = \frac{\boldsymbol{B}}{GM} \tag{3.89}$$

points from the center of mass to the perigee and can be derived from (3.70) as (Bate et al., 1971, p. 62)

$$\boldsymbol{e} = \frac{1}{GM} \left(\left(v^2 - \frac{GM}{r} \right) \boldsymbol{r} - (\boldsymbol{r} \cdot \boldsymbol{v}) \boldsymbol{v} \right). \tag{3.90}$$

The relationship to Kepler's elements of orientation is given via simple vector operations (Fig. 3.10):

$$
\begin{aligned}
\cos i &= \frac{\boldsymbol{h} \cdot \boldsymbol{Z}}{h} = \frac{h_z}{h}, \\
\cos \Omega &= \frac{\boldsymbol{n} \cdot \boldsymbol{X}}{n} = \frac{n_x}{n}, \\
\cos \omega &= \frac{\boldsymbol{n} \cdot \boldsymbol{e}}{n\, e}, \\
\cos \nu &= \frac{\boldsymbol{e} \cdot \boldsymbol{r}}{e\, r}.
\end{aligned}
\tag{3.91}
$$

Also,

$$
p = \frac{h^2}{GM} \quad \text{and} \quad e = |\boldsymbol{e}|.
$$

We can compute a from p using (3.2), and $\overline{M}$ from ν using (3.53) and (3.54), because the inversion of (3.54) gives

$$
\cos E = \frac{e + \cos \nu}{1 + e \cos \nu}.
\tag{3.92}
$$

For inclinations between $0°$ and $90°$ the satellite motion is towards the east and is called *direct motion*. For inclinations between $90°$ and $180°$ the motion is oriented westward and is called *retrograde motion*.

Formulas for the conversion between Keplerian elements and equatorial satellite coordinates, as well as between Keplerian elements and the position and velocity vector $\boldsymbol{r}, \boldsymbol{v}$, are developed in section [3.3.1].

Kepler's elements are less suitable for circular orbits or orbits with small eccentricities, because singularities may occur in the computations. In such cases alternative sets of orbital parameters are advisable, such as *Hill's orbital variables*. They can be interpreted as the satellite's position and velocity vector in spherical coordinates (r, u, Ω). The variables are (Fig. 3.11):
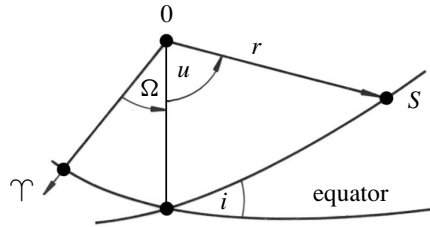
Figure 3.11. Hill's orbital parameters

$$
\begin{aligned}
&r, \dot{r}, u, \Omega, \\
&G = r^2 \cdot \dot{u}, \\
&H = G \cdot \cos i.
\end{aligned}
\tag{3.93}
$$

G is here a variable and not the gravitational constant. The relationships between Keplerian elements and Hill's variables are given by the equations (3.94). In order to

avoid misunderstanding, we introduce here the symbol μ for the product of Earth's mass M and the gravitational constant G, so that

$$
\begin{aligned}
&G = (\mu a(1 - e^2))^{\frac{1}{2}} & &u = v + \omega \\
&r = \frac{G^2}{\mu}\frac{1}{1 + e\cos v} & &H = G\cos i \\
&\dot{r} = \frac{\mu}{G}e\sin v & &\Omega = \Omega.
\end{aligned}
\tag{3.94}
$$

More details, including the first derivatives with respect to time, can be found for example in Lelgemann (1979); Cui, Mareyen (1992). A similar set of variables, avoiding singularities, and also named *non-singular elements* (Montenbruck, Gill, 2000, p. 29) is given by *Delaunay* (cf. Brouwer, Clemence, 1961, p. 540). Hill's or Delaunay's variables are considered as *canonical orbital elements*. These elements are referred to units, for instance to the

unit of mass	$M = 1$, more exactly $M + m = 1$,
unit of length	mean distance a of the particle m from the gravitational center of the central mass (e.g. the Astronomical Unit AU for the Earth-Sun system),
unit of time	selected such that $GM = 1$, i.e. according to Kepler's third law $T = \frac{1}{n} = 2\pi\sqrt{\dfrac{a^3}{GM}}$.

The satellite's orbital velocity can then be represented in terms of [unit of length/unit of time].

The advantage of the canonical representation can be seen in the fact that the equations can be used without accurate knowledge of the numerical values of the constants. This was of great importance in the history of celestial mechanics, because the distance and mass of the Sun were not very well known. On the other hand, it is clear that a scale is brought into the solution of satellite geodesy by fixing numerical values for GM.

In mathematics a system of differential equations is called *canonic* when

$$
\frac{d\xi_j}{dt} = \frac{\partial F}{\partial \eta_j}; \quad \frac{d\eta_j}{dt} = -\frac{\partial F}{\partial \xi_j}, \quad j = 1, \dots, n.
\tag{3.95}
$$

F is the *Hamilton-function*. This function plays an important role in the dynamics of mechanical systems and thus in celestial mechanics (Brouwer, Clemence, 1961, p. 530).

3.2   Perturbed Satellite Motion

Up to now we have considered the motion of a satellite with negligible mass under the central gravitative force of a single point mass M. This *Keplerian motion* was

described by the basic equation (3.20):

$$\ddot{r} = -\frac{GM}{r^3} r.$$

The integration of this equation formally gives the solution

$$r(t) = r(t; a_1, \ldots, a_6), \quad \dot{r}(t) = \dot{r}(t; a_1, \ldots, a_6) \tag{3.96}$$

with $a_1, \ldots, a_6$ being free selectable integration constants. Preferably, the Kepler elements $a, e, i, \omega, \Omega, \overline{M}$ are used.

In reality a certain number of additional forces act on the near-Earth satellite. To distinguish them from the central force (*central body acceleration*) these are called *perturbing forces*. The satellite experiences additional accelerations because of these forces, which can be combined into a resulting *perturbing vector* k_s. The extended equation of motion reads as

$$\ddot{r} = -\frac{GM}{r^3} r + k_s. \tag{3.97}$$

Perturbing forces are in particular responsible for:

1. Accelerations due to the non-spherically and inhomogeneous mass distribution within Earth (central body), $\ddot{r}_E$.
2. Accelerations due to other celestial bodies (Sun, Moon and planets), mainly $\ddot{r}_S$, $\ddot{r}_M$.
3. Accelerations due to Earth and oceanic tides, $\ddot{r}_e$, $\ddot{r}_o$.
4. Accelerations due to atmospheric drag, $\ddot{r}_D$.
5. Accelerations due to direct and Earth-reflected solar radiation pressure, $\ddot{r}_{SP}$, $\ddot{r}_A$.

The perturbing forces causing 1 to 3 are gravitational in nature; the remaining forces are non-gravitational. The total is

$$k_s = \ddot{r}_E + \ddot{r}_S + \ddot{r}_M + \ddot{r}_e + \ddot{r}_o + \ddot{r}_D + \ddot{r}_{SP} + \ddot{r}_A. \tag{3.98}$$

Fig. 3.12 gives a graphical description of the perturbing forces and accelerations. The resulting total acceleration depends on the location r of the satellite, i.e. a quantity which first has to be determined from the solution of the differential equation (3.97) as a function of time. Consequently, the integration of (3.97) is a complex problem.

One usual way to solve such complex problems in celestial mechanics is to start with reasonable simplifications and to correct the resulting "error" in a separate second
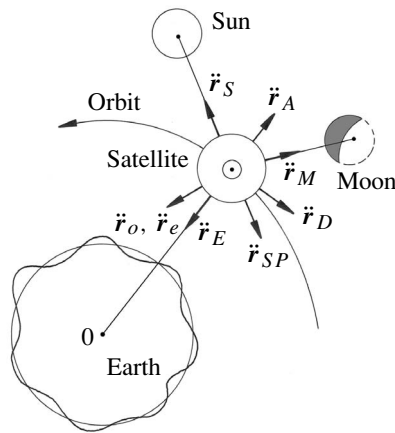
Figure 3.12. Perturbing forces acting on a satellite

step. Such simplifications, considered to be reasonable, are called *intermediate*. In this sense, the satellite motion described by equations (3.20) and (3.96) is an *intermediate motion* or *basic motion*.

Two procedures are used to solve the perturbed motion problem. In the first case the satellite's coordinates are considered to be disturbed directly. The effect of the perturbing forces on the coordinates is determined numerically, and the differences between disturbed and undisturbed coordinates are computed. The analytical form of the trajectory is not analyzed. This *numerical integration* of the orbit is treated in more detail in [3.3.2.2]; this type of solution is also called *variation of coordinates*.

For the second type of solution the integration constants a_i of the undisturbed case (3.96) are considered to be time dependent functions. Formally, the solution of the perturbed problem is

$$r(t) = r(t; a_1(t), \ldots, a_6(t)), \quad \dot{r}(t) = \dot{r}(t; a_1(t), \ldots, a_6(t)), \qquad (3.99)$$

where analytical expressions have to be found for the time-dependent orbital elements. This procedure, the *analytical integration*, is also called *variation of constants*; it is treated in more detail in [3.3.2.1]. The basic concept is that perturbations are regarded as deviations between the (intermediate) Keplerian elements at a given initial epoch and at some further epochs. In order to apply the method of variation of constants it is necessary to start with an approximate solution for the equation of motion. For the planetary system and for artificial near-Earth satellites this precondition is fulfilled with the Keplerian elements.

3.2.1 Representation of the Perturbed Orbital Motion

3.2.1.1 Osculating and Mean Orbital Elements

The requirement for orbital elements to be time-dependent leads to the concept of *osculating orbital elements*. Let a satellite, whose motion is described through equation (3.97), have the position vector $r(t_k)$ at time $t = t_k$ and the velocity vector $\dot{r}(t_k)$. Suppose that all perturbations k_s could be removed at this particular epoch t_k, then the further motion of the satellite would be in an undisturbed Keplerian orbit, governed by the initial conditions $r(t_k)$ and $\dot{r}(t_k)$. This orbit is called an *osculating* (lat. osculare = to kiss, to embrace) or *instantaneous orbit*, because it coincides with the true, disturbed, orbital path at epoch t_k, when the initial parameters are equal.

In reality the perturbing forces do not disappear; this is why the satellite is located on a different osculating orbit for each particular epoch. The true satellite orbit can be regarded to be the envelope of all successive osculating orbits with the osculating elements $a(t_k)$, $e(t_k)$, $\ldots \overline{M}(t_k)$. With t_k as a time parameter, continuously increasing, the perturbed satellite motion can be interpreted to be a Keplerian motion with time-variable elements

$$a(t), \; e(t), \; i(t), \; \omega(t), \; \Omega(t), \; \overline{M}(t).$$

For applications in satellite geodesy, with Earth's gravitation as the primary force, the osculating elements serve very well for orbit approximation because they change

slowly. This is because the acceleration of the central body exceeds the remaining perturbing accelerations at least by a factor 10^3 (cf. [3.2.4], [12.2]). It is therefore possible to approximate, for the use of orbit predictions, the orbital elements by a power series in time differences, $(t - t_0)$, with t_0 being a mean epoch:

$$a_i(t) = a_i(t_0) + \dot{a}_i(t - t_0) + \ddot{a}_i(t - t_0)^2 + \cdots \quad (i = 1, \ldots, 6). \qquad (3.100)$$

The "history" of an osculating element $a_i(t)$ is represented as the sum of long- and short-periodic terms:

$$a_i(t) = \bar{a}_i(t) + \Delta a_i(t). \qquad (3.101)$$

$\bar{a}_i(t)$ contains the sum of low frequency, secular and constant parts, $\Delta a_i(t)$ represents the high-frequency oscillations. The terms $\bar{a}_i(t)$ are also called *mean elements*. Thus, mean elements can be considered as osculating elements with vanishing periodic terms.

3.2.1.2 Lagrange's Perturbation Equations

We now have to establish a relation between the acting perturbing forces and the time dependent variations of the orbital elements. The appropriate basic equations were formulated by *Lagrange* (1736–1813). The explicit derivations can be found in textbooks on celestial mechanics (e.g. Brouwer, Clemence, 1961; Taff, 1985), also Kaula (1966).

Following equations (3.61) and (3.78) the total energy of the satellite motion is determined by

$$E_M = \frac{v^2}{2} - \frac{GM}{r} = -\frac{GM}{2a}.$$

The negative term of the total energy, $GM/2a$, is also named the *force function F*. With the potential V as the negative value of the potential energy and the symbol T for the kinetic energy we find the following form of the force function (e.g. Kaula, 1966):

$$F = V - T. \qquad (3.102)$$

In a non-central force field

$$V = \frac{GM}{r} + R,$$

$$F = \frac{GM}{r} + R - T = \frac{GM}{2a} + R. \qquad (3.103)$$

The function R contains all components of V excluding the central term $\frac{GM}{r}$, and is called the disturbing function or *disturbing potential*.

For the sake of completeness an alternative form of the equation of motion (3.97) in a non-central force field is given

$$\ddot{\boldsymbol{r}} = \operatorname{grad} V = \nabla V. \qquad (3.104)$$

This form will be used later in [3.2.2.3].

With Lagrange's perturbation equations a relationship between the disturbing potential R and the variation of the orbital elements is established, e.g. Brouwer, Clemence (1961, p. 284), Kaula (1966, p. 29):

$$
\begin{aligned}
\frac{da}{dt} &= \frac{2}{na}\frac{\partial R}{\partial \overline{M}}, \\
\frac{de}{dt} &= \frac{1-e^2}{na^2 e}\frac{\partial R}{\partial \overline{M}} - \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial R}{\partial \omega}, \\
\frac{d\omega}{dt} &= -\frac{\cos i}{na^2\sqrt{1-e^2}\sin i}\frac{\partial R}{\partial i} + \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial R}{\partial e}, \\
\frac{di}{dt} &= \frac{\cos i}{na^2\sqrt{1-e^2}\sin i}\frac{\partial R}{\partial \omega} - \frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial R}{\partial \Omega}, \\
\frac{d\Omega}{dt} &= \frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial R}{\partial i}, \\
\frac{d\overline{M}}{dt} &= n - \frac{1-e^2}{na^2 e}\frac{\partial R}{\partial e} - \frac{2}{na}\frac{\partial R}{\partial a}.
\end{aligned}
\tag{3.105}
$$

In order to avoid singularities, alternative forms of perturbation equations are available, e.g. Brouwer, Clemence (1961, p. 287), Taff (1985, p. 308ff). In particular for orbits with small eccentricity (ω indeterminate) or with small inclination (Ω indeterminate) other forms are preferable.

For the canonical set of Hill's elements (3.94) the following relations hold with the geocentric gravitational constant $GM = \mu$:

$$
\begin{aligned}
\frac{d\dot{r}}{dt} &= \left(-\frac{\mu}{r^2} + \frac{G^2}{r^3}\right) + \frac{\partial R}{\partial r}, \\
\frac{dr}{dt} &= \dot{r} - \frac{\partial R}{\partial \dot{r}}, \\
\frac{dG}{dt} &= \frac{\partial R}{\partial u}, \\
\frac{du}{dt} &= \frac{G}{r^2} - \frac{\partial r}{\partial G} - \frac{\cos i}{G\sin i}\frac{\partial R}{\partial i}, \\
\frac{di}{dt} &= \frac{\cos i}{G\sin i}\frac{\partial R}{\partial u} - \frac{1}{G\sin i}\frac{\partial R}{\partial \Omega}, \\
\frac{d\Omega}{dt} &= \frac{1}{G\sin i}\frac{\partial R}{\partial i}.
\end{aligned}
\tag{3.106}
$$

The analytical integration of the perturbation equations (3.96) or (3.106) requires that the disturbing potential R is written as a function of the orbital elements. With the derivatives at hand, the integration can then be executed, cf. [3.3.2.2]. So long as R only depends on Earth's anomalous gravity field, the relation between the coefficients of the potential expansion and the orbit perturbations can be formulated. This aspect is

used when Earth's gravity field is derived from an analysis of perturbed satellite orbits [12.2].

It becomes evident that the implication of particular perturbations on satellite orbits can at best be investigated via the analytical solution of the perturbation equations.

3.2.1.3   Gaussian Form of Perturbation Equation

In some cases it is useful to formulate the disturbing accelerations directly at the satellite in componental form, instead of using partial derivatives of the disturbing potential in the elements. This is, for example, true for orbits with large eccentricities, where series expansions would require many terms in e. The formulas of type (3.105) are also less suitable for numerical treatment. An appropriate alternative form was developed by *Gauss*. According to Fig. 3.13 the perturbing forces at the satellite are resolved into three mutually perpendicular components

$$\text{grad } R = \nabla R = \begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix}, \quad (3.107)$$
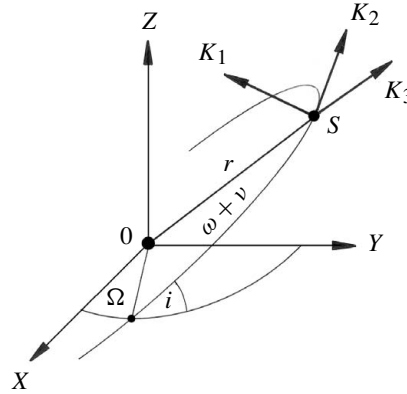
Figure 3.13. Gaussian form of perturbing forces

with

  K_1 perpendicular to the orbital plane, positive toward the north pole,

  K_2 perpendicular to the radius vector in the orbital plane, positive in the direction of increasing longitude, and

  K_3 in the direction of the radius vector, positive in the direction of increasing radial distance.

The corresponding perturbation equations are, Brouwer, Clemence (1961, p. 301), Arnold (1970, p. 28), Taff (1985, p. 314), Beutler et al. (1998, p. 61f):

$$\frac{da}{dt} = \frac{2}{n\sqrt{1-e^2}} \left(e \sin \nu \, K_3 + \frac{p}{r} K_2 \right),$$

$$\frac{de}{dt} = \frac{\sqrt{1-e^2}}{na} (\sin \nu \, K_3 + (\cos E + \cos \nu) K_2),$$

$$\frac{d\omega}{dt} = \frac{\sqrt{1-e^2}}{nae} \left(-\cos\nu\,K_3 + \left(\frac{r}{p} + 1 \right) \sin\nu\,K_2 \right) - \cos i \frac{d\Omega}{dt},$$

$$\frac{di}{dt} = \frac{1}{na\sqrt{1-e^2}} \frac{r}{a} \cos(\omega + \nu)K_1,$$

$$\frac{d\Omega}{dt} = \frac{1}{na\sqrt{1-e^2}} \frac{r}{a} \frac{\sin(\omega + \nu)}{\sin i}K_1, \tag{3.108}$$

$$\frac{d\overline{M}}{dt} = n - \frac{1}{na} \left(\frac{2r}{a} - \frac{1-e^2}{e} \cos\nu \right) K_3 - \frac{1-e^2}{nae} \left(1 + \frac{r}{p} \right) \sin\nu\,K_2.$$

Equations (3.108) are convenient in that they allow the influences of the components K_1, K_2, K_3 to be separately discussed. We see immediately that only K_1 is capable of changing the orientation of the orbital plane (elements Ω and i). A change of the semi-major axis a can be achieved by K_2 for $e \ll 1$, i.e. by the component in the direction of satellite motion. This is of importance for satellite maneuvers. Note that in the literature the symbols W, S, R are also used instead of K_1, K_2, K_3.

3.2.2   Disturbed Motion due to Earth's Anomalous Gravity Field

The dominant, by far, perturbing force on orbits of near-Earth artificial satellites is due to Earth's oblateness, illustrated in Fig. 3.14. The equatorial bulge produces a slight torque on the satellite and tries to rotate the satellite orbit into the equatorial plane. It results in an effect similar to the precession of Earth's rotational axis [2.1.2]; the vector of the orbit's angular momentum moves around Earth's axis of maximum momentum, i.e. the orbital plane and the nodes move westward for direct orbits and eastward for retrograde orbits.
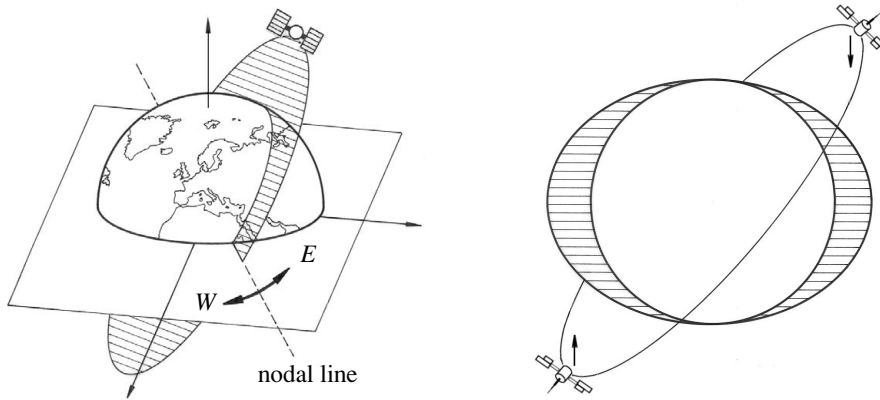
Figure 3.14. Nodal precession caused by Earth's oblateness

The nodal regression rate is visualized in Fig. 3.15 (left). The effect can reach $9°$/day for low orbits (< 200 km) and for low inclination. Polar orbits have no nodal precession when we neglect higher order terms in the mass distribution.
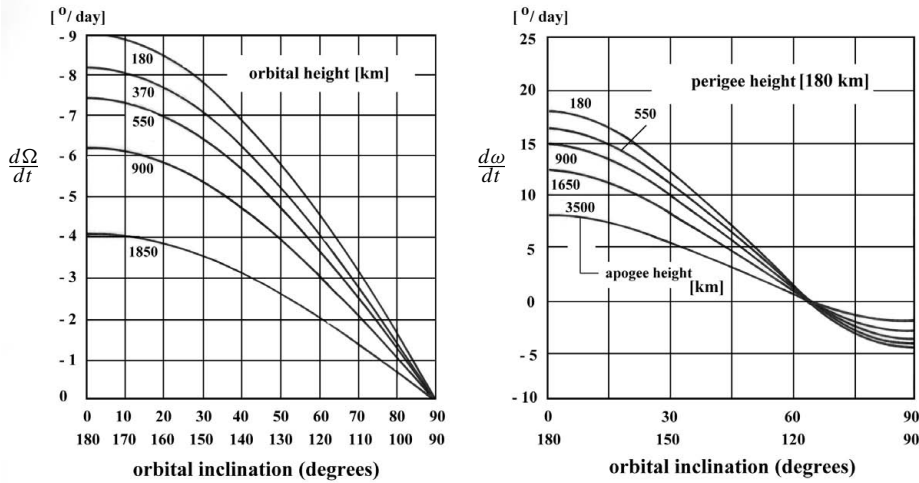
Figure 3.15. Relation between orbit inclination, height and nodal regression (left); relation between orbit inclination, height and perigee rotation (right) (cf. Bate et al., 1971)

The effect of nodal precession can be used to advantage, e.g. to generate specific ground track repetition rates for Earth observation satellites, or to generate sun-synchronous orbits, cf. [3.4.3]. Undesired nodal regression can lead to uneven orbital coverage within a satellite configuration. This happened with the TRANSIT Navy Navigation Satellite System, cf. [6.2].

The oblateness of Earth has another effect, which acts most on eccentric orbits; this is the rotation $d\omega/dt$ of the line of apsides. Fig. 3.15 (right) gives an impression of this effect for orbits with perigee heights of 180 km and different apogee heights. The rotation of apsides happens in the direction of the satellite motion for orbit inclinations $i < 63.4°$ and $i > 116.6°$. For inclinations $63.4° < i < 116.6°$ the rotation of apsides is against the satellite motion. Again, this feature can be used for the planning of satellite missions, when the orbit inclination is sufficiently different to 63.4°. Predictions of the behavior of orbits are difficult for the *critical inclination $i = 63.4°$*, cf. [3.2.2.2], (Taff, 1985, p. 340).

3.2.2.1   Perturbation Equation and Geopotential

For more detailed investigations of the relationship between Earth's anomalous gravitational potential and perturbations of the satellite orbit it is helpful to use a series expansion with spherical harmonics for the potential, well known from physical geodesy

(e.g. Torge, 2001):

$$V = \frac{GM}{r}\left(1 + \sum_{n=1}^{\infty}\sum_{m=0}^{n}\left(\frac{a_e}{r}\right)^n (C_{nm}\cos m\lambda + S_{nm}\sin m\lambda)P_{nm}(\cos\vartheta)\right). \quad (3.109)$$

The harmonic coefficients C_{nm}, S_{nm} are integrals of the mass and describe the mass distribution within the central body; a_e is the equatorial radius and P_{nm} are the so-called *associated Legendre functions* or *Legendre polynomials* (cf. Moritz, 1990). A further discussion of (3.109) follows in chapter [12.2].

The first term $\frac{GM}{r}$ describes the potential of a homogeneous sphere and thus refers to Keplerian motion; consequently it is also named the *Keplerian term*. The remaining expressions within the double-summation are due to the disturbing potential R, cf. (3.103). With the origin of the coordinate system transferred to the center of mass of the primary, the terms with $n = 1$ and $m = 0, 1$ become zero, hence

$$R = \frac{GM}{r}\left(\sum_{n=2}^{\infty}\sum_{m=0}^{n}\left(\frac{a_e}{r}\right)^n (C_{nm}\cos m\lambda + S_{nm}\sin m\lambda)P_{nm}(\cos\vartheta)\right). \quad (3.110)$$

The coefficients are named

$$\begin{aligned}
\text{zonal} \quad &\text{when} \quad m = 0, \\
\text{tesseral} \quad &\text{when} \quad m \neq 0, \text{ and} \\
\text{sectorial} \quad &\text{when} \quad m = n.
\end{aligned}$$

For further developments the following steps are essential:
- reformulate the expansion (3.110) as a function of the orbital elements,
- form the derivatives with respect to the elements, and
- substitute the derivatives into the disturbing equations (3.105).

The related transformations and representations are quite cumbersome. In the following only the principal relations are summarized. Detailed derivations and discussions can be found in the fundamental book of Kaula (1966) and also with Mueller (1964), Groten (1979, 1980). A short and clear summary is given by Khan (1983).

In the *first step* the disturbing potential

$$R = \sum_{n=2}^{\infty}\sum_{m=0}^{n} R_{nm} \quad (3.111)$$

is re-formulated as a function of the orbital elements:

$$R_{nm} = \frac{GMa_e^n}{a^{n+1}}\sum_{p=0}^{n} F_{nmp}(i)\sum_{q=-\infty}^{+\infty} G_{npq}(e)S_{nmpq}(\omega, \overline{M}, \Omega, \Theta). \quad (3.112)$$

Θ is the Greenwich sidereal time. The following definitions hold for the particular functions. The *inclination function* is defined as follows:

$$F_{nmp}(i) = \sum_{t} \frac{(2n - 2t)!}{2^{2n-2t}(n - m - 2t)!n!}\sin^{n-m-2t} i$$

$$\sum_{s=0}^{m} \binom{m}{s} \cos^s i \sum_c \binom{n-m-2t+s}{c} \binom{m-s}{p-t-c} (-1)^{c-k} \qquad (3.113)$$

with the summation rules (Kaula, 1966):

t is summed from 0 to the lesser of p or k, where k is the integer part of $(n-m)/2$, thus

$$0 \le t \le \left\{ \begin{array}{c} p \\ k = (n-m)/2 \\ k = (n-m-1)/2 \end{array} \right\} \quad \text{for} \quad \left\{ \begin{array}{c} p < (n-m)/2 \\ p \ge (n-m)/2 \\ p \ge (n-m-1)/2 \end{array} \right\}.$$

c is summed over all values making the binomial coefficients non-zero, thus (Khan, 1983):

$$\left[\begin{array}{c} p-t \le m-s \\ p-t \ge m-s \end{array} \right] \quad \text{then} \quad \left\{ \begin{array}{c} 0 \\ p-t-m+s \end{array} \right\} \le c$$

$$\le \left\{ \begin{array}{c} n-m-2t+s \\ p-t \end{array} \right\} \quad \text{when} \quad \left[\begin{array}{c} p-t \ge n-m-2t+s \\ p-t \le n-m-2t+s \end{array} \right].$$

Equation (3.113) is suitable for computer use. A table with expressions for the inclination function $F_{nmp}(i)$ up to $n = m = p = 4$, already in a form suitable for hand calculation, is given in Kaula (1966, p. 34)

The *eccentricity function* $G_{npq}(e)$ is defined separately for long period and short period terms [3.2.2.2]. For *long period terms* not containing the mean anomaly $\overline{M}$:

$$n - 2p + q = 0 \quad \text{or} \quad q = 2p - n$$

and

$$G_{np(2p-n)}(e) = \frac{1}{(1-e^2)^{n-1/2}} \sum_{d=0}^{p'-1} \binom{n-1}{n+2d-2p'} \binom{n+2d-2p'}{d} \left(\frac{e}{2}\right)^{n+2d-2p'}$$
$$(3.114)$$

with

$$p' = \left\{ \begin{array}{c} p \\ n-p \end{array} \right\} \quad \text{for} \quad \begin{array}{c} p \le \frac{n}{2} \\ p \ge \frac{n}{2} \end{array}.$$

For short period terms, $n - 2p + q \ne 0$, the development of the eccentricity function is more complicated. Kaula (1966) gives one solution which refers to *Tisserand* (1889):

$$G_{npq}(e) = (-1)^{|q|} \left(1 + \frac{e^2}{(1+\sqrt{1-e^2})^2}\right)^n \left(\frac{e}{1+\sqrt{1-e^2}}\right)^{|q|}$$
$$\sum_{k=0}^{\infty} P_{npqk} Q_{npqk} \left(\frac{e}{1+\sqrt{1-e^2}}\right)^{2k} \qquad (3.115)$$

with

$$P_{npqk} = \sum_{r=0}^{h} \binom{2p' - 2n}{h - r} \left(\frac{(n - 2p' + q')(1 + \sqrt{1 - e^2})}{2} \right)^r \frac{(-1)^r}{r!}.$$

The summation rules are

$$h = \begin{Bmatrix} k + q' \\ k \end{Bmatrix} \quad \text{for} \quad \begin{matrix} q' > 0 \\ q' < 0 \end{matrix}$$

as well as

$$Q_{npqk} = \sum_{r=0}^{h} \binom{-2p'}{h - r} \left(\frac{(n - 2p' + q')(1 + \sqrt{1 - e^2})}{2} \right)^r \frac{1}{r!}$$

and

$$h = \begin{Bmatrix} k \\ k - q' \end{Bmatrix} \quad \text{for} \quad \begin{matrix} q' > 0 \\ q' < 0 \end{matrix}.$$

For both summations:

$$h = \begin{Bmatrix} k \\ n - p \end{Bmatrix} \quad \text{and} \quad q' = \begin{Bmatrix} q \\ -q \end{Bmatrix} \quad \text{for} \quad \begin{matrix} p \leq \frac{n}{2} \\ p \geq \frac{n}{2} \end{matrix}.$$

Tabulated values of the eccentricity function $G_{npq}(e)$ (equations (3.114) and (3.115)) can be found in Kaula (1966, p. 38).

The remaining elements, $\omega, \overline{M}, \Omega, \Theta$, are related to the harmonic coefficients C_{nm}, S_{nm}, by the definition

$$\begin{aligned}
S_{nmpq}(\omega, \overline{M}, \Omega, \Theta) = & \begin{pmatrix} C_{nm} \\ -S_{nm} \end{pmatrix} \begin{matrix} n - m \text{ even} \\ n - m \text{ odd} \end{matrix} \times \cos((n - 2p)\omega \\
& + (n - 2p + q)\overline{M} + m(\Omega - \Theta)) \\
& + \begin{pmatrix} S_{nm} \\ C_{nm} \end{pmatrix} \begin{matrix} n - m \text{ even} \\ n - m \text{ odd} \end{matrix} \times \sin((n - 2p)\omega \\
& + (n - 2p + q)\overline{M} + m(\Omega - \Theta)).
\end{aligned} \tag{3.116}$$

In the *second step* the expressions (3.112) have to be derived in terms of the orbital elements, and substituted into the perturbation equations (3.105). This yields (Kaula,

1966; Mueller, 1964; Khan, 1983):

$$\frac{d\Omega_{nmpq}}{dt} = \frac{GMa_e^n F'_{nmp} G_{npq} S_{nmpq}}{\bar{n}a^{n+3}\sqrt{1-e^2}\sin i},$$

$$\frac{di_{nmpq}}{dt} = \frac{GMa_e^n F_{nmpq} G_{npq} S'_{nmpq}}{\bar{n}a^{n+3}\sqrt{1-e^2}\sin i}((n-2p)\cos i - m),$$

$$\frac{d\omega_{nmpq}}{dt} = GMa_e^a \left(\frac{\sqrt{1-e^2}}{e}F_{nmp}G'_{npq} - \frac{\cot i}{\sqrt{1-e^2}}F'_{nmp}G_{npq}\right)\frac{S_{nmpq}}{\bar{n}a^{n+3}},$$

$$\frac{da_{nmpq}}{dt} = \frac{2GMa_e^n F_{nmp} G_{npq} S'_{nmpq}}{\bar{n}a^{n+2}}(n-2p+q), \qquad (3.117)$$

$$\frac{de_{nmpq}}{dt} = \frac{GMa_e^n F_{nmp} G_{npq} S'_{nmpq}}{\bar{n}a^{n+3}e}((1-e^2)(n-2p+q) - \sqrt{1-e^2}(n-2p)),$$

$$\frac{d\overline{M}_{nmpq}}{dt} = \frac{GMa_e^n F_{nmp} S_{nmpq}}{\bar{n}a^{n+3}}\left(2(n+1)G_{npq} - \frac{1-e^2}{e}G'_{npq}\right) + \bar{n}.$$

With equations (3.117) relations have been established between the spherical harmonic coefficients S_{nm}, C_{nm} and the pertinent variations of the orbital elements. $\bar{n}$ is the mean motion from (3.52). The functions S', G', F' are derivatives with respect to the related arguments.

Integration of the perturbation equations (3.117) over a given time interval leads, for a particular $nmpq$-term of the disturbing function R, to equation (3.119). This holds under the assumption that the effect of a particular perturbation has no significant influence on the formulation of another perturbation, i.e. that a first order approximation is sufficient. This is true for the potential of Earth. With

$$\dot{\psi}_{nmpq} = (n-2p)\dot{\omega} + (n-2p+q)\dot{\overline{M}} + m(\dot{\Omega} - \dot{\Theta}), \qquad (3.118)$$

(i.e. when only monotonous time-dependent variations are considered for ω, Ω and $\overline{M}$ [3.2.2.2]) and with $\overline{C}_{nmpq}$ as the integral of C_{nmpq}, it follows (Mueller, 1964; Khan, 1983) that

$$\Delta\Omega_{nmpq} = GMa_e^n \frac{F'_{nmp} G_{npq} \overline{S}_{nmpq}}{\bar{n}a^{n+3}\sqrt{1-e^2}\sin i\,\dot{\psi}_{nmpq}},$$

$$\Delta i_{nmpq} = GMa_e^n \frac{F_{nmp} G_{npq}((n-2p)\cos i - m)C_{nmpq}}{\bar{n}a^{n+3}\sqrt{1-e^2}\sin i\,\dot{\psi}_{nmpq}},$$

$$(3.119)$$

$$\Delta\omega_{nmpq} = GMa_e^n \frac{(e^{-1}\sqrt{1-e^2}F_{nmp}G'_{npq} - \dfrac{\cot i}{\sqrt{1-e^2}}F'_{nmp}G_{mpq})\overline{C}_{nmpq}}{\bar{n}a^{n+3}\dot{\psi}_{nmpq}},$$

$$\Delta a_{nmpq} = GMa_e^n \frac{2F_{nmp} G_{npq} C_{nmpq}(n-2p+q)}{\bar{n}a^{n+2}\dot{\psi}_{nmpq}},$$

$$\Delta e_{nmpq} = GMa_e^n \frac{F_{nmp}G_{npq}C_{nmpq}((1-e^2)(n-2p+q) - \sqrt{1-e^2}(n-2p))}{\bar{n}a^{n+3}e\dot{\psi}_{nmpq}},$$

$$\Delta \overline{M}_{nmpq} = GMa_e^n \frac{F_{nmp}\overline{C}_{nmpq}(2(n+1)G_{npq} - (1-e^2)e^{-1}G'_{npq})}{\bar{n}a^{n+3}\dot{\psi}_{nmpq}}.$$

The total perturbation caused by the disturbing potential R follows as a first order approximation from linear superposition of particular perturbations:

$$\Delta\Omega(t) = \sum_{nmpq} \Delta\Omega_{nmpq}(t), \quad \Delta i(t) = \sum_{nmpq} \Delta i_{nmpq}(t), \ldots. \tag{3.120}$$

Formulas for equivalent perturbation equations, written with Hill's orbital elements, can be found in Lelgemann (1979).

3.2.2.2 Perturbations of the Elements

The explicit algebraic formulation of equations (3.117) and (3.119) is rather cumbersome, but it gives a qualitative and quantitative insight into the relation between the spherical harmonic coefficients and the nature of the perturbations. It is convenient to divide the perturbations into three groups, corresponding to their periods. These are *secular* (linear), *long period* and *short period* perturbations. They can be superimposed (Fig. 3.16).
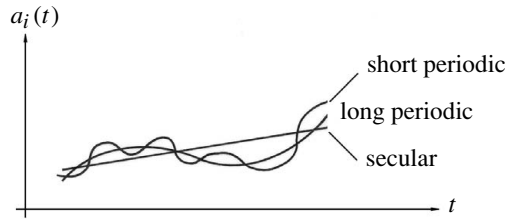
Figure 3.16. Perturbations of the elements $a_i(t)$

The time-dependent variations of the perturbations are described with the functions $S_{nmpq}(\omega, \overline{M}, \Omega, \Theta)$, following (3.112). The nature of the particular perturbation is therefore governed by the temporal behavior of the argument for S_{nmpq} in (3.116), namely by (3.118)

$$\dot{\psi}_{nmpq} = (n-2p)\dot{\omega} + (n-2p+q)\dot{\overline{M}} + m(\dot{\Omega} - \dot{\Theta}).$$

In (3.118) the whole spectrum of frequencies, which can be generated from the nmpq-terms, is present. It can be demonstrated that the perturbations in the elements are governed by the same frequencies that are present in the functions S_{nmpq} (Arnold, 1970). With

$$n - 2p = n - 2p + q = m = 0$$

the total angular argument in equation (3.112)

$$\psi = (n - 2p)\,\omega + (n - 2p + q)\,\overline{M} + m\,(\Omega - \Theta) \qquad (3.121)$$

equals zero. Hence, only secular perturbations are generated by such *nmpq*-combinations. Again, from the condition $m = 0$, it is evident that only the zonal coefficients $C_{n,0}$ can cause secular perturbations. Consequently, the influence of the tesseral and sectorial spherical harmonics must be smaller than the influence of the zonal harmonics.

The two most important *secular perturbations* of low satellite orbits have been already mentioned at the beginning of this section, and demonstrated in Fig. 3.14 and 3.15. These are the *oblateness perturbations*, namely the rotation of the nodal and apsidal lines caused by the second order zonal harmonic C_{20}. Furthermore, the mean anomaly $\overline{M}$ experiences secular perturbations. This can be derived from (3.121) in connection with the perturbation equation (3.112).

With $n = 2, m = 0, p = 1, q = 0$ it follows (Arnold, 1970; Goad, 1977) that

$$
\begin{aligned}
\frac{d\Omega}{dt} &= C_{20} \frac{3\bar{n}a_e^2}{2a^2(1 - e^2)^2} \cos i, \\[2mm]
\frac{d\omega}{dt} &= C_{20} \frac{3\bar{n}a_e^2}{4a^2(1 - e^2)^2} (1 - 5\cos^2 i), \\[2mm]
\frac{d\overline{M}}{dt} &= \bar{n} - C_{20} \frac{3\bar{n}a_e^2}{4a^2\sqrt{(1 - e^2)^3}} (3\cos^2 i - 1), \\[2mm]
\frac{di}{dt} &= \frac{da}{dt} = \frac{de}{dt} = 0.
\end{aligned}
\qquad (3.122)
$$

C_{20} does not produce secular perturbations in the elements i, a, e. However, C_{20} does give rise to secular variations of the elements $\Omega, \omega, \overline{M}$, because the numerical value of C_{20} exceeds all other potential coefficients by a factor of 10^3. These variations can be used as *reference elements*; they represent a secularly preceding Kepler-ellipse with the elements $a, e, i, \dot{\Omega}, \dot{\omega}, \dot{\overline{M}}$. The remaining perturbations are then described as deviations from the reference values. This type of procedure is particularly suitable when very small perturbations have to be analyzed, like those produced by geodynamical effects such as solid Earth tides.

From equation (3.122) it can be deduced that the variation of ω vanishes for $\cos i \approx \pm\sqrt{5}/5$ or $i \approx 63.4°$ or $116.6°$. This situation is called the *critical inclination*. The behavior of the orbit is no longer predictable with elementary methods. Furthermore it becomes clear that satellites in polar orbits ($i = 90°$) do not experience nodal variations caused by C_{20}. For ($i > 90°$) the nodes proceed (*prograde motion*), and for ($i < 90°$) the motion is *retrograde*. *Long-period perturbations* often show periods longer than 100 days. They are caused by $\dot{\omega}$, a variation of the line of apsides. They are present when the relations

$$n - 2p \neq 0, \quad n - 2p + q = 0, \quad \text{and} \quad m = 0 \qquad (3.123)$$

are valid in equation (3.112) under consideration of (3.121). Because $m = 0$, the long-period perturbations are related to the zonal harmonics. It can be shown that the even zonal harmonics induce secular variations in Ω, ω and $\overline{M}$ as well as long-period variations in e and i. The odd zonal harmonics produce long-period variations in Ω, ω, i, e and $\overline{M}$ (Khan, 1983). Because of their particular behavior, the even zonal harmonics are generally determined from secular perturbations in Ω and ω, whereas the odd zonals are derived from variations in e and i.

A more detailed analysis cannot be restricted to first order perturbations because the perturbations in the elements produced by C_{20} cannot be neglected in the integration process. Since C_{20} is 10^3 times bigger than the remaining terms the non-linear terms C_{20}^2 or $C_{20}C_{30}, C_{20}C_{40}\ldots$ also have to be considered. For more details see Kaula (1966, p. 41ff).

Short-period perturbations are present when (3.121)

$$n - 2p + q \neq 0 \qquad \text{and/or} \qquad m \neq 0. \qquad (3.124)$$

The short-period perturbations, caused by Earth's gravity field, thus have periods which are related to the satellite's revolution (through $\overline{M}$) and Earth rotation (through $\Omega - \Theta$). A tesseral coefficient of order m will cause perturbations with m cycles/day because of the term $m(\dot{\Omega} - \dot{\Theta})$ in (3.118). The corresponding $nmpq$-combinations are called *m-daily terms* (Goad, 1977).

The analysis of short-period perturbations is useful for the determination of the tesseral and sectorial harmonics, because $m \neq 0$ (3.124). The longest period of a tesseral harmonic of order $m = 1$ is one day. This is why the selected satellite orbits and the computational methods have to be sufficiently sensitive with respect to short-period perturbations.

Possible periods for particular index combinations can be derived from $\dot{\psi}_{nmpq}$ (3.118) by

$$P_{nmpq} = \frac{2\pi}{|\dot{\psi}_{nmpq}|}. \qquad (3.125)$$

Finally, applying equations (3.119) possible types of perturbations can be identified for the particular elements. The total sum of perturbations experienced by a satellite is then composed from secular and/or periodic perturbation, with periods related to Earth rotation, satellite revolution, nodal and apsidal rotation as well as combinations thereof. Table 3.2 gives a simplified overview. In Table 3.5 [3.2.4] the quantitative effects of perturbations on orbital elements is demonstrated in examples.

3.2.2.3 Perturbations Caused by the Zonal Coefficients J_n

In order to estimate the effect of Earth's anomalous gravity field on particular satellite orbits, it is often sufficient to determine the accelerations caused by the first zonal harmonics. Here the expansion (3.109) is simplified. In satellite geodesy it is usual to replace C_{n0} as

$$C_{n0} = -J_n,$$

Table 3.2. Characteristics of perturbations in the elements

Parameter	secular perturbations	long-period perturbations	short-period perturbations
a	-	-	×
e	-	×	×
i	-	×	×
Ω	×	×	×
ω	×	×	×
$\overline{M}$	×	×	×

and it then follows that

$$
V = \frac{GM}{r}\left(1 - \sum_{n=2}^{\infty} J_n \left(\frac{a_e}{r}\right)^n P_n(\cos\vartheta)\right).
\tag{3.126}
$$

Restricting the expansion to the order $n = 6$, and writing L for the geocentric latitude, we obtain (Bate et al., 1971, p. 421)

$$
\begin{aligned}
V = \frac{GM}{r}\Bigg(&1 - \frac{J_2}{2}\left(\frac{a_e}{r}\right)^2 (3\sin^2 L - 1) + \frac{J_3}{2}\left(\frac{a_e}{r}\right)^3 (5\sin^3 L - 3\sin L)\\
&- \frac{J_4}{8}\left(\frac{a_e}{r}\right)^4 (35\sin^4 L - 30\sin^2 L + 3)\\
&+ \frac{J_5}{8}\left(\frac{a_e}{r}\right)^5 (63\sin^5 L - 70\sin^3 L + 15\sin L)\\
&- \frac{J_6}{16}\left(\frac{a_e}{r}\right)^6 (231\sin^6 L - 315\sin^4 L + 105\sin^2 L - 5)\Bigg).
\end{aligned}
\tag{3.127}
$$

Using the unit vectors X, Y, Z, introduced for the equatorial system in Fig. 3.10, we obtain for the acceleration $\ddot{r}$ (3.104):

$$
\ddot{r} = \nabla V = \frac{\partial V}{\partial x}X + \frac{\partial V}{\partial y}Y + \frac{\partial V}{\partial z}Z,
\tag{3.128}
$$

and hence for the acceleration components $\ddot{x}$, $\ddot{y}$, $\ddot{z}$ at the satellite location x, y, z:

$$
\begin{aligned}
\ddot{x} = \frac{\partial V}{\partial x} = \frac{GMx}{r^3}\Bigg(&1 - J_2\frac{3}{2}\left(\frac{a_e}{r}\right)^2\left(5\frac{z^2}{r^2} - 1\right) + J_3\frac{5}{2}\left(\frac{a_e}{r}\right)^3\left(3\frac{z}{r} - 7\frac{z^3}{r^3}\right)\\
&- J_4\frac{5}{8}\left(\frac{a_e}{r}\right)^4\left(3 - 42\frac{z^2}{r^2} + 63\frac{z^4}{r^4}\right)
\end{aligned}
\tag{3.129}
$$

$$- J_5 \frac{3}{8} \left(\frac{a_e}{r}\right)^5 \left(35\frac{z}{r} - 210\frac{z^3}{r^3} + 231\frac{z^5}{r^5}\right)$$

$$+ J_6 \frac{1}{16} \left(\frac{a_e}{r}\right)^6 \left(35 - 945\frac{z^2}{r^2} + 3465\frac{z^4}{r^4} - 3003\frac{z^6}{r^6}\right) + \ldots \Bigg),$$

$$\ddot{y} = \frac{\partial V}{\partial y} = \frac{y}{x}\ddot{x}, \tag{3.130}$$

$$\ddot{z} = \frac{\partial V}{\partial z} = -\frac{GMz}{r^3} \left(1 + J_2 \frac{3}{2} \left(\frac{a_e}{r}\right)^2 \left(3 - 5\frac{z^2}{r^2}\right)\right.$$

$$+ J_3 \frac{3}{2} \left(\frac{a_e}{r}\right)^3 \left(10\frac{z}{r} - \frac{35}{3}\frac{z^3}{r^3} - \frac{r}{z}\right)$$

$$- J_4 \frac{5}{8} \left(\frac{a_e}{r}\right)^4 \left(15 - 70\frac{z^2}{r^2} + 63\frac{z^4}{r^4}\right) \tag{3.131}$$

$$- J_5 \frac{1}{8} \left(\frac{a_e}{r}\right)^5 \left(315\frac{z}{r} - 945\frac{z^3}{r^3} + 693\frac{z^5}{r^5} - 15\frac{r}{z}\right)$$

$$\left. - J_6 \frac{1}{16} \left(\frac{a_e}{r}\right)^6 \left(315 - 2205\frac{z^2}{r^2} + 4851\frac{z^4}{r^4} - 3003\frac{z^6}{r^6}\right) + \ldots \right).$$

Numerical values for the low zonal harmonics $J_n = -C_{n0}$ are, based on the geopotential model EGM96 [12.2], and using (3.188) for the denormation:

$$J_2 = 1082.63 \times 10^{-6}, \quad J_3 = -2.54 \times 10^{-6}, \quad J_4 = -1.62 \times 10^{-6},$$
$$J_5 = -0.23 \times 10^{-6}, \quad J_6 = +0.54 \times 10^{-6}.$$

3.2.3   Other Perturbations

The dominant perturbing influence on the orbits of low artificial satellites is Earth's non-central gravitational field. In order to obtain high accuracy in orbit computations, which is necessary for most geodetic applications of satellite observations, it is necessary to estimate also the accelerations caused by other perturbing forces. This is in particular true for the gravitational influences of the Sun and Moon. For low orbiting satellites atmospheric drag also plays an important role.

3.2.3.1   Perturbing Forces Caused by the Sun and Moon

Assuming that the Sun and the Moon can be considered to be point-masses, as with the satellite, we can take advantage of the basic equation of motion (3.20)

$$\ddot{\boldsymbol{r}} = -\frac{GM}{r^3}\boldsymbol{r}.$$

The following equations are valid for the accelerations acting on a satellite with neg-ligible mass. X, Y, Z are the axes of an arbitrary, but inertial, reference frame.

Using the notation of Fig. 3.17 we find for the acceleration of the satellite caused by the mass m_M of the Moon and the mass m_E of Earth

$$\ddot{\boldsymbol{r}}_0 = G\left(-\frac{m_E}{|\boldsymbol{r}|^3}\boldsymbol{r} + \frac{m_M}{|\boldsymbol{\rho}_M|^3}\boldsymbol{\rho}_M\right). \tag{3.132}$$

Furthermore, the acceleration of m_E caused by m_M

$$\ddot{\boldsymbol{r}}_{0E} = G\frac{m_M}{|\boldsymbol{r}_M|^3}\boldsymbol{r}_M. \tag{3.133}$$

The relative acceleration of the satellite with respect to Earth is

$$\ddot{\boldsymbol{r}}_0 - \ddot{\boldsymbol{r}}_{0E} = \ddot{\boldsymbol{r}} = G\left\{\frac{m_M}{|\boldsymbol{\rho}_M|^3}\boldsymbol{\rho}_M - \frac{m_M}{|\boldsymbol{r}_M|^3}\boldsymbol{r}_M - \frac{m_E}{|\boldsymbol{r}|^3}\boldsymbol{r}\right\}. \tag{3.134}$$
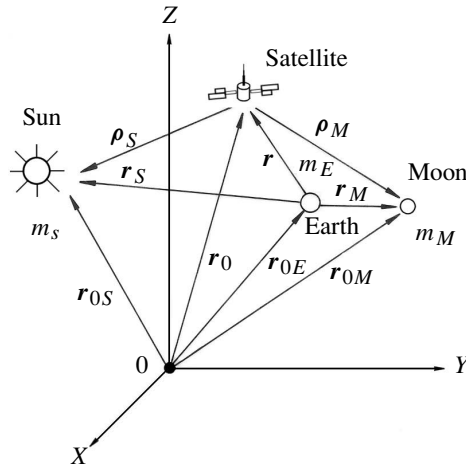
Figure 3.17. Gravitational attraction of the Sun and the Moon on a satellite

With $|\boldsymbol{\rho}| = \rho$, $|\boldsymbol{r}| = r$ and $m_E = M$, and with the origin of the coordinate system at Earth's center of mass, we find

$$\ddot{\boldsymbol{r}} = -\frac{GM}{r^3}\boldsymbol{r} + Gm_M\left(\frac{\boldsymbol{\rho}_M}{\rho_M^3} - \frac{\boldsymbol{r}_M}{r_M^3}\right). \tag{3.135}$$

The first term is due to the acceleration caused by Earth (central term). The additional perturbing acceleration, caused by the gravitational attraction of the Moon acting on

the satellite, is then

$$\ddot{\boldsymbol{r}}_M = Gm_M \left(\frac{\boldsymbol{r}_M - \boldsymbol{r}}{|\boldsymbol{r}_M - \boldsymbol{r}|^3} - \frac{\boldsymbol{r}_M}{r_M^3} \right). \tag{3.136}$$

The corresponding influence $\ddot{\boldsymbol{r}}_S$, due to the Sun, is

$$\ddot{\boldsymbol{r}}_S = Gm_S \left(\frac{\boldsymbol{r}_S - \boldsymbol{r}}{|\boldsymbol{r}_S - \boldsymbol{r}|^3} - \frac{\boldsymbol{r}_S}{r_S^3} \right). \tag{3.137}$$

The masses of the disturbing bodies and their locations within a geocentric reference frame have to be known for numerical computations. Useful constants are

for the mass of the Sun: $Gm_S \approx 1325 \cdot 10^8$ km^3 s^{-2}, and
for the mass of the Moon: $Gm_M \approx 49 \cdot 10^2$ km^3 s^{-2}.

The rather simple formulas derived above can also be used to calculate the perturbations on artificial satellites caused by the planets. Those perturbations are, however, negligible in most cases.

Cartesian coordinates (ephemerides) of the Sun and the Moon (and of the major planets) are available upon request from the Jet Propulsion Laboratory (JPL) in the form of Chebyshev approximations (Seidelmann (ed.), 1992, chapt. 5). The so-called "Development Ephemerides" (DE) are the fundamental ephemerides of the *Astronomical Almanac*. Currently the DE200 (includes nutations but not librations) and DE405 (includes both nutations and librations) ephemerides are widely used (Montenbruck, Gill, 2000). In the most recent series all data are referred to the International Celestial Reference Frame (ICRF) [2.1.2.1].

For many purposes it is sufficient to make use of simple equations for the solar and lunar coordinates, since the forces generated by the Sun and the Moon are much smaller than the central attraction of Earth. For appropriate formulas, based on mean orbital elements and the assumption of an unperturbed motion of Earth around the Sun, see Montenbruck, Gill (2000, p. 70ff).

In order to study the effect of the perturbations on orbital elements we use the procedure adopted for the derivation of (3.112), i.e. the perturbing potential of a celestial body has to be expanded in terms of the orbital elements and substituted into the perturbation equations (3.105) after differentiation. Corresponding expressions for the Moon are given e.g. by Kozai (1959, 1966). Ω and ω are subject to secular perturbations. Following Kozai (1959) we get

$$\begin{aligned}
\frac{d\omega}{dt} &= \frac{3}{4}\frac{n_M^2}{n}m_M\frac{1}{\sqrt{1-e^2}}\left(2 - \frac{5}{2}\sin^2 i + \frac{1}{2}e^2\right)\left(1 - \frac{3}{2}\sin^2 i_M\right), \\
\frac{d\Omega}{dt} &= -\frac{3}{4}\frac{n_M^2}{n}m_M\frac{\cos i}{\sqrt{1-e^2}}\left(1 + \frac{3}{2}e^2\right)\left(1 - \frac{3}{2}\sin^2 i_M\right),
\end{aligned} \tag{3.138}$$

with n_M, i_M as the lunar mean motion and the inclination of the lunar orbit. m_M is given as the mass ratio of the Moon to Earth.

Furthermore there are periodic perturbations in the elements Ω, ω, a, e, i. They are related to the annual and monthly orbital motions of the Sun and the Moon. Explicit derivations of the formulas are given by Giacaglia (1973) and Taff (1985, p. 348ff).

Orbital perturbations caused by the Sun and the Moon may be significant (see Table 3.4). They have to be taken into account for orbit computation. The acceleration acting on a GPS satellite at a height of 20 000 km is about $5 \cdot 10^{-6}$ m/s^2 for the Moon, and $2 \cdot 10^{-6}$ m/s^2 for the Sun. The influence of the planets is only about $3 \cdot 10^{-10}$ m/s^2.

It should be noted that (3.135) gives the basic equation of the *N-body problem* in celestial mechanics.

3.2.3.2 Solid Earth Tides and Ocean Tides

Solid Earth tides and ocean tides change Earth's gravitational potential and thus cause additional accelerations acting on the satellite, which can be considered as an *indirect gravitational effect* of the Sun and the Moon.

The acceleration of the satellite caused by solid Earth tides is (Rizos, Stolz, 1985)

$$\ddot{\boldsymbol{r}}_e = \frac{k_2}{2} \frac{Gm_d}{r_d^3} \frac{a_e^5}{r^4} (3 - 15 \cos^2 \theta) \frac{\boldsymbol{r}}{r} + 6 \cos \theta \frac{\boldsymbol{r}_d}{r_d}, \tag{3.139}$$

where
- m_d mass of the disturbing body (Sun, Moon),
- $\boldsymbol{r}_d$ geocentric position vector of the disturbing body,
- θ angle between the geocentric position vector $\boldsymbol{r}$ of the satellite and $\boldsymbol{r}_d$, and
- k_2 Love number, describing the elasticity of Earth's body.

The acceleration of a GPS satellite is rather small, being $2 \cdot 10^{-9}$ m/s^2. For low-orbiting satellites, such as STARLETTE [8.2] the influences are much greater. This is why STARLETTE is used for the determination and modeling of solid body tides. The tidally induced variations in Earth's external potential can be expressed as variations in the spherical harmonic geopotential coefficients. For explicit formulas see e.g. Balmino (1973), Eanes et al. (1983), Dow (1988), Kang (1998), Rim, Schutz (1999).

The effects of ocean tides on satellites are rather difficult to model because of the irregular coast lines. Using a global tide model, e.g. from Schwiderski (1984), Eanes, Bettadpur (1996), LeProvost et al. (1998) it is possible to compute for each point P on the ocean surface the tidal heights and the resulting tidal-induced mass variations:

$$dm_p = \rho_0 h(P, t) d\sigma. \tag{3.140}$$

ρ_0 is the average density of water, t the time and $d\sigma$ a surface element. The potential variation, caused by this mass variation, is (Rizos, Stolz, 1985)

$$\Delta U = \frac{Gdm_p}{a_e} \sum_n (1 + k_n') P_{n0} \cos \psi, \tag{3.141}$$

which can be related to the orbit perturbations via the perturbation equation (3.110). In (3.141) k_n' denotes the deformation coefficients, P_{n0} the Legendre polynomials, and ψ the geocentric angle between the initial point A and the surface point P.

The effect of ocean tides on satellite orbits is very small. The largest influences are sensed in the inclination angle i and in the nodal longitude Ω. The effects have periods between $\sim$10 days and $\sim$100 days, and they are mostly below $0.''1$ (Goad, 1977). For GPS satellites the acceleration is of the order of $5 \cdot 10^{-10}$ m/s^2 (corresponding to less than 1 m after 2 days). For the orbital analysis of near-Earth satellites a detailed modeling of tidal influences is essential. Explicit derivations and discussions can be found in Lambeck et al. (1975), Goad (1977), Dow (1988), Kang (1998), Rim, Schutz (1999). Detailed formulas for the practical computation of solid Earth and ocean tides are described in McCarthy (2000).

3.2.3.3 Atmospheric Drag

For low-orbiting satellites the most important non-gravitational perturbation is caused by a drag, due to the interaction between the satellite and particles of the atmosphere. The aerodynamic forces, acting on the surface of the spacecraft, depend on:

- the geometry of the satellite,
- the velocity of the satellite,
- the orientation of the satellite with respect to the flow, and
- the density, temperature and composition of the atmospheric gas.

Hence, the appropriate mathematical modeling of the resulting forces turns out to be a rather complicated problem. Based on many years of empirical investigation, the following formula proves to give useful results. Here the acceleration $\ddot{\boldsymbol{r}}_D$ is in a direction opposite to the force of the atmospheric resistance:

$$\ddot{\boldsymbol{r}}_D = -\frac{1}{2}C_D\rho(\boldsymbol{r},t)\frac{A}{m_s}(\dot{\boldsymbol{r}} - \dot{\boldsymbol{r}}_a)|\dot{\boldsymbol{r}} - \dot{\boldsymbol{r}}_a|, \qquad (3.142)$$

where

m_s	mass of the satellite,
A	effective cross-sectional area of the satellite,
C_D	drag coefficient (satellite specific),
$\rho(\boldsymbol{r},t)$	density of the atmosphere near the satellite,
$\boldsymbol{r},\dot{\boldsymbol{r}}$	position and velocity vector of the satellite, and
$\dot{\boldsymbol{r}}_a$	velocity of the atmosphere near the satellite.

Assuming that the atmosphere rotates rigidly with Earth, we obtain in a geocentric equatorial coordinate system the relative velocity of the satellite with respect to the atmosphere:

$$\dot{\boldsymbol{r}} - \dot{\boldsymbol{r}}_a = \begin{pmatrix} \dot{x} + \dot{\theta}y \\ \dot{y} - \dot{\theta}x \\ \dot{z} \end{pmatrix}, \quad \dot{\theta} = \text{Earth rotation rate.} \qquad (3.143)$$

Typical values of the coefficient C_D range from 1.5 to 3.0, and are usually estimated as parameters along with the orbit determination process (Montenbruck, Gill, 2000). For a spherical satellite, C_D is approximated as 2. For more complicated surfaces, like a cylinder, a cone or a plane, C_D becomes larger. For the area-to-mass ratio the spacecraft attitude must be known.

The density of the atmosphere depends not only on the height but also on other parameters like geographic location, season, time of the day, Sun activity and geomagnetism, and it can be computed with suitable models. *Jacchia's* (1971) models were derived from satellite orbit perturbations, and they form part of the *CIRA 72 atmosphere* (Conventional International Reference Atmosphere). For recent model developments see e.g. Montenbruck, Gill (2000).

The influence of atmospheric drag decreases rather rapidly with increasing height. Table 3.3 gives an idea of the mean atmospheric density for different heights (Cappelari et al., 1976; Montenbruck, Gill, 2000).

Table 3.3. Density of the upper atmosphere

Height km	Density g/km^3	Height km	Density g/km^3
100	497 400	600	0.081 – 0.639
200	255 – 316	700	0.020 – 0.218
300	17 – 35	800	0.007 – 0.081
400	2.2 – 7.5	900	0.003 – 0.036
500	0.4 – 2.0	1000	0.001 – 0.018

The perturbing forces on the satellite may cause accelerations varying between 10^{-3} and $10^{-9} \mathrm{m/s}^2$ (see Fig. 3.20). The atmospheric drag for higher regions strongly varies with the solar and geomagnetic activity. This is why predictions are required for spacecraft operations, e.g. for remote sensing satellites.

The mean anomaly $\overline{M}$ will be the most disturbed by the drag effect, because the influence of the atmospheric friction is directed toward the orbital motion of the satellite. Because of the energy balance the total energy of the satellite motion will decrease together with the kinetic energy, and thus the semi-major axis a will reduce. According to Kepler's third law, the angular velocity of the satellite will increase through the friction force of the atmospheric drag.

For satellites with an orbital height of about 1000 km or less, for example Earth observation satellites or satellites of the former TRANSIT system (cf. [4.3]) the drag effects can be rather large. For TOPEX/POSEIDON, with an orbital height of 1340 km, the atmospheric acceleration is of the order $4 \cdot 10^{-10} \mathrm{m/s}^2$. For satellites in higher orbits, like GPS, the atmospheric drag has no effect.

3.2.3.4 Direct and Indirect Solar Radiation Pressure

The particle radiation, continuously emitted by the Sun, has two effects on a satellite. These are the direct radiation pressure, resulting from the interaction of the solar radiation with the spacecraft, and the indirect, Earth-reflected portion (*albedo*).

The force acting directly on the satellite is proportional to the effective satellite surface area, to the reflectivity of the surface and to the solar flux; it is inversely proportional to the velocity of light and to the square of the distance between the satellite and the Sun. The resulting perturbing acceleration is thus (Cappelari et al., 1976; Montenbruck, Gill, 2000):

$$\ddot{\boldsymbol{r}}_{SP} = \nu P_S \frac{C_r O}{m} (AU)^2 \frac{(\boldsymbol{r} - \boldsymbol{r}_S)}{|\boldsymbol{r} - \boldsymbol{r}_S|^3}, \tag{3.144}$$

with

P_S Sun-constant (quotient of solar flux and velocity of light in the Astronomical Unit),

(AU) Astronomical Unit ($1.5 \cdot 10^8$ km),

O/m cross-section area of the satellite as seen from the Sun divided by its mass,

$\boldsymbol{r}, \boldsymbol{r}_S$ geocentric position vector of the satellite and of the Sun in the space-fixed equatorial system,

C_r factor of reflectivity for the satellite surface ($C_r = 1.95$ for aluminium), and

ν shadow function:

       $\nu = 0$, satellite in Earth's shadow,

       $\nu = 1$, satellite in sunlight, and

       $0 < \nu < 1$, satellite in half-shadow.

A simple cylindric shadow-model is sufficient for the decision whether the satellite is in Earth's shadow or not (Cappelari et al., 1976). According to Fig. 3.18 the satellite is in sunlight, when

$$D = \boldsymbol{r}'\boldsymbol{r}'_S > 0 \tag{3.145}$$

and in shadow, when

$$D < 0$$

$$|\boldsymbol{S}_c| = |\boldsymbol{r}' - D\boldsymbol{r}'_S| < a_e.$$

Figure 3.18. Cylindric shadow model

a_e is the semi-axis of the shadow generating body (Earth), and $\boldsymbol{r}'_S$ is the unit vector to the Sun. This simple model is not sufficient for higher accuracy demands. In order to avoid discontinuities in the numerical orbit computation near the shadow transits it is possible to use a suitable smoothing shadow-transfer-function for ν. Also

the "Sun-constant" P_S can be introduced as a variable in order to model variable Sun activity. The solar radiation pressure shows an annual variation of about 3% due to the elliptic Earth orbit. Finally, detailed models for complicated satellite surfaces can be used, to account for the effect of the satellite's orientation with respect to the Sun. Commonly C_r is estimated as a free parameter in orbit determination programs, and thus absorbs the effect of unknown details of the satellite orientation and reflectivity.

The radiation pressure on GPS satellites is rather complicated to model accurately due to the complex structure of the spacecraft. Usually a body-fixed spacecraft coordinate system is introduced. The Y-axis is along the solar panel beam (Fig. 3.19). In the tailored *ROCK4* and *ROCK42* models the GPS satellite is subdivided into several surfaces which are modeled either as a plane or a cylinder with particular reflectivity characteristics (Landau, 1988; Fliegel, Gallini, 1989; Fliegel et al., 1992; Rothacher, 1992). The influence of the direct radiation pressure is mainly effective in the direction of the orbital motion (along track), and can reach 10 m and more after a few hours. The acceleration is in the order of $1 \cdot 10^{-7} \text{m/s}^2$. An excellent introduction into the problem is given by Ziebart et al. (2002).
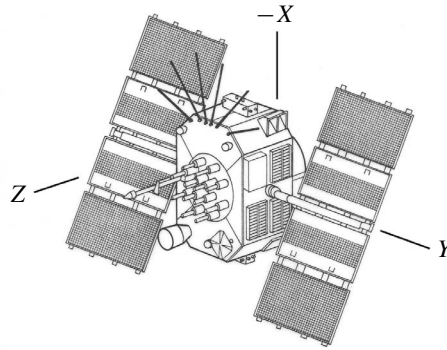
Figure 3.19. The satellite-fixed coordinate system

Part of the solar radiation is reflected by Earth. The ratio between the reflected radiation and the incoming flux is called *albedo*. The albedo part of the radiation pressure is very difficult to model, because of the varying distribution of land, sea and clouds, but in most cases it is less than 10% of the direct radiation. The estimate for GPS satellites lies between 1% and 2% (King et al., 1987). This is why the albedo effect is usually neglected for GPS orbit computations, except for very long orbital arcs. The magnitude of the perturbing acceleration is approximately $4 \cdot 10^{-10} \text{m/s}^2$.

A very detailed discussion of the direct and indirect solar radiation effect, as well as other non-gravitational perturbations, can be found in Milani et al. (1987). For recent models in GPS orbit computation see McCarthy (2000).

3.2.3.5 Further Perturbations

For very precise orbital analysis other perturbations may be considered, whose individual contributions to the acceleration of the satellite are usually far below 10^{-9}m/s^2. These are, for example:
  − friction caused by charged particles in the upper atmosphere,
  − thermal radiation of the satellite,
  − heating effects at shadow boundaries,
  − electromagnetic interaction in the geomagnetic field, and

    − influences of the inter-planetary dust.

A discussion and evaluation of such perturbing forces can be found in special literature, e.g. Ries (1997); Ries, Tapley (1999). For practical purposes in satellite geodesy these perturbations are mostly not considered.

    It should be mentioned that so-called *apparent forces* arise if the equation of motion is formulated with respect to a moving reference system instead of a space-fixed one (Schneider, 1981, p. 30, Reigber, 1981). These are, if the moving system is non-uniformly rotating and/or accelerated as against the space-fixed system, the

<div>

centrifugal force $\qquad \boldsymbol{Z} = -m\boldsymbol{d}' \times (\boldsymbol{d}' \times \boldsymbol{r}')$,

gyro force $\qquad \boldsymbol{T} = -m\dfrac{\partial \boldsymbol{d}'}{\partial t} \times \boldsymbol{r}'$,

Coriolis force $\qquad \boldsymbol{C} = -2m\boldsymbol{d}' \times \dfrac{\partial \boldsymbol{r}'}{\partial t}$,

</div>

(3.146)

which an observer in a space-fixed system would not notice. From the equation of motion in the stationary system

$$m\frac{d^2 \boldsymbol{r}}{dt^2} = \boldsymbol{K}(t, \boldsymbol{r}, \dot{\boldsymbol{r}}) \qquad (3.147)$$

follows the equation of motion in the moving system (Schneider, 1981):

$$m\frac{\partial^2 \boldsymbol{r}'}{\partial t^2} = \boldsymbol{K}' - m\boldsymbol{d}' \times (\boldsymbol{d}' \times \boldsymbol{r}') - m\frac{\partial \boldsymbol{d}'}{\partial t} \times \boldsymbol{r}' - 2m\boldsymbol{d}' \times \frac{\partial \boldsymbol{r}'}{\partial t}, \qquad (3.148)$$

with $\boldsymbol{K}'$ the force in the rotating system, $\boldsymbol{r}'$ the position vector in the rotating system, and $\boldsymbol{d}'$ the rotation vector.

    For an explicit computation of the apparent accelerating forces, using the expression for the rotational vector $\boldsymbol{d}'$ and the derivatives thereof in the moving system, see for example Reigber (1981).

    Relativistic effects are, for most applications in satellite geodesy, smaller than the observation accuracy. In many cases they are cancelled by the observation technique, or they are modeled through other parameters. Insofar as relativistic effects are of importance, they will be discussed together with the particular satellite methods (e.g. [7.4.1]). With respect to orbital dynamics it follows from general relativity that the orbital elements are subject to additional secular perturbations. These influences are much greater for the orbits of near-Earth satellites than for planets (cf. the relativistic perihelion rotation of Mercury). Cugusi, Proverbio (1978) give the appropriate formulas, and they find as mean values for satellites of geodetic interest:

$$10''/\text{year for } \omega, \quad 0\overset{.}{.}2/\text{year for } \Omega, \quad \text{and} \quad 0\overset{.}{.}2/\text{year for } \overline{M}.$$

The correction to the acceleration of an artificial satellite, based on general relativity, is (McCarthy, 2000)

$$\ddot{\boldsymbol{r}}_{\text{rel}} = \frac{GM}{c^2 r^3}\left(\left(4\frac{GM}{r} - \dot{r}^2\right)\boldsymbol{r} + 4\,(\boldsymbol{r} \cdot \dot{\boldsymbol{r}})\,\dot{\boldsymbol{r}}\right), \qquad (3.149)$$

with

c speed of light,
r satellite position vector,
$\dot{r}$ satellite velocity vector, and
GM geocentric constant of gravitation.

The relativistic correction of the accelerations is in the order of $3 \cdot 10^{-10}$m/s^2 for GPS satellites and $1 \cdot 10^{-8}$m/s^2 for TOPEX/POSEIDON.

For some satellite systems particular, non-gravitational accelerations are generated from *thrust* or *attitude control* maneuvers. They have to be considered in orbital analyses. Thrust forces appear in connection with maneuvers for orbit corrections. Attitude control systems change the satellite's orientation in space. Cappelari et al. (1976) or Montenbruck, Gill (2000, p. 104f) give formulas for the consideration of such effects.

In dynamical orbit determination it is not possible to model all perturbations perfectly. This holds in particular for the non-conservative force models which are limited by uncertainties in the knowledge of platform orientation, material properties, and surface temperatures (Montenbruck, Gill, 2000). This is why *empirical accelerations* are employed to take account of this effect. In general the empirical forces are described by an equation of the following type:

$$\ddot{r}_{\text{em}} = E(a_0 + a_1 \sin \nu + a_2 \cos \nu), \tag{3.150}$$

were a_0 is a constant acceleration bias, a_1 and a_2 are coefficients related to the frequency (e.g. one cycle per orbital revolution), and E is a matrix to transform the acceleration biases from the local orbital frame (radial, cross-track, and along-track) into the inertial system. For details see e.g. Montenbruck, Gill (2000, p. 112).

3.2.3.6 Resonances

Resonances occur when the period of a satellite revolution is an integer multiple of Earth's rotation period. This leads to an amplification of certain non-zonal harmonics S_{nm}, C_{nm}, resulting in much higher amplitudes in the element perturbation than normal. In geometric terms, resonances appear when consecutive revolutions of a satellite are separated exactly by an interval which corresponds to the wave-length of the particular harmonic coefficient. After a given number of revolutions, the sub-satellite orbit repeats, i.e. the satellite crosses the same regions and is subject to the same perturbations. This causes an amplification of the initial perturbation and generates the *resonance effect*. Consequently, a satellite with $\approx m$ revolutions/day will be sensitive to resonant influences from the tesseral coefficients C_{nm}, S_{nm}.

From a mathematical point of view, resonances develop when the denominator (3.118) in the perturbation equation (3.119) becomes very small:

$$\dot{\psi}_{nmpq} = (n - 2p)\dot{\omega} + (n - 2p + q)\dot{\overline{M}} + m(\dot{\Omega} - \dot{\Theta}) \approx 0. \tag{3.151}$$

Satellite orbits can be explicitly selected to determine particular tesseral harmonics with high accuracy, using the resonance effect and equation (3.118). The corollary is that in orbit computation it is essential to know whether specific high order potential coefficients can give rise to large perturbations, caused by resonances. Low orbiting satellites, because of their frequent revolutions, are particularly affected by short wave structures of the geopotential. Resonances may be present also for Earth observation and remote sensing satellites, because of their dedicated orbital design with particular repetition rates. GPS satellites experience resonance effects caused by Earth's ellipticity (Delikaraoglou, 1989).

Insofar as different coefficients generate resonances of identical phase and amplitude, they cannot be separated, and only derived jointly from orbital analyses. The determination of such so-called *lumped coefficients* is treated for example by Klokočník (1982). For a detailed discussion of resonances in high satellite orbits (GPS, geosynchronous) see e.g. Hugentobler (1998).

3.2.4 Implications of Perturbations on Selected Satellite Orbits

The effect of perturbations on the motion of satellites which are used in geodesy is demonstrated for some typical orbits in Fig. 3.20.
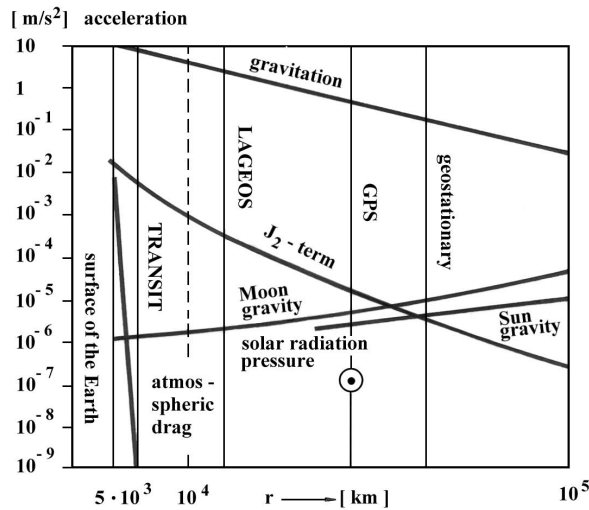
Figure 3.20. Relation between the orbital height and the magnitude of different perturbing forces

At a height of about *1000 km* above Earth's surface we find the TRANSIT navigation satellites, satellites for remote sensing (e.g. SPOT, LANDSAT), the altimeter satellites (GEOS-3, SEASAT-1, ERS-1/2), and many other satellites with corner cube reflectors, which are used in the determination of Earth models. At a height of about *6000 km*

are the orbits of the "Laser Geodynamic Satellites" LAGEOS-1 and LAGEOS-2. At a height of about *20 000 km* are the GPS and GLONASS satellites.

Fig. 3.20 vizualizes the relation between the orbital height of satellites and the most important perturbing forces (Landau, Hagmeier, 1986). The effects of particular perturbations on the orbit of a GPS satellite are listed in Table 3.4 (King et al., 1987).

Table 3.4. Implications of perturbations on the orbit of a GPS satellite

Perturbation	Acceleration m/s^2	Effect on the orbit 2^h-orbit	3-days orbit
Central force (for comparison)	0.56		
C_{20}	$5 \cdot 10^{-5}$	2 km	14 km
Further harmonics	$3 \cdot 10^{-7}$	50–80 m	100–1500 m
Solar & Lunar gravitation	$5 \cdot 10^{-6}$	5–150 m	1000–3000 m
Body tides	$1 \cdot 10^{-9}$	–	0.5–1.0 m
Ocean Tides	$1 \cdot 10^{-9}$	–	0.0-2.0 m
Solar radiation pressure	$1 \cdot 10^{-7}$	5–10 m	100–800 m
Albedo	$1 \cdot 10^{-9}$	–	1.0–1.5 m

Finally, the influence of perturbations on particular orbital elements of a GPS satellite is shown. Table 3.5 gives the deviations of the true orbit from a Keplerian reference orbit after 4 hours (Nakiboglu et al., 1985).

Table 3.5. Effects of perturbations on GPS satellites after 4 hours

Element	C_{20}	Higher order geopotential	Sun Moon	Solar radiation pressure
a	2600 m	20 m	220 m	5 m
e	1600 m	5 m	140 m	5 m
i	800 m	5 m	80 m	2 m
Ω	4800 m	3 m	80 m	5 m
$\omega + \overline{M}$	1200 m	4 m	500 m	10 m

3.3   Orbit Determination

The basic task in orbit determination is to derive elements for the description of orbits from observations or from a priori known information. The fundamental ideas, equations and procedures for the solution of this problem were developed in [3.1] and [3.2]. In this section some additional and more advanced aspects are considered.

In classical celestial mechanics, for the sake of simplified computations, the task of orbit determination was, in general, divided into

- initial orbit determination without over-determination, and then
- orbit improvement using all available observations.

With the *initial orbit determination* a preliminary orbit is usually derived from three sets of directional observations at different epochs, neglecting all perturbations. Today, with all the electronic computation facilities at hand, this distinction is no longer necessary. For some purposes, however, a simplified orbit determined without taking perturbations into account, or only considering mean osculating elements [3.3.1], may be completely sufficient. This is especially true when long, but less accurate, orbital arcs are required for orbit predictions, and for the establishment of observation programs, or for the identification of unknown objects (e.g. Hugentobler, 1998).

In parameter estimation mostly shorter arcs are required with very high accuracy. Today this is an appropriate field for *numerical orbit integration methods* using digital computers [3.3.2.2]. On the other hand, *analytical* considerations are required [3.3.2.1] to study the time dependent behavior of orbits, and to select appropriate orbital characteristics. When satellite observations are analyzed, or when predicted orbits are used, it is often necessary to represent short orbital arcs with suitable smoothing algorithms [3.3.3]. A new tool for *precise orbit determination* (POD) stems from the fact that satellite trajectories can now be determined directly [3.3.2.3] with onboard packages like GPS [7], PRARE [4.3.3.3], or DORIS [6.7].

3.3.1   Integration of the Undisturbed Orbit

We distinguish between the *initial value problem*, connected with the name of *Laplace*, and the *boundary value problem*, named after *Gauss*. In the *initial value problem* the orbital elements of a space vehicle are determined from the position and velocity vector r and v of a space vehicle at a given epoch. These may be known from the burn-out position and velocity of the launching rocket, hence

$$r, v \rightarrow (a, e, i, \Omega, \omega, \overline{M}). \tag{3.152}$$

One particularly clear derivation of these relations was given with equations (3.88) to (3.91), using the normal vector h, the nodal vector n and the perigee vector e, see also Montenbruck, Gill (2000, p. 28 f).

Based on the known Keplerian elements the position and the velocity vector of the satellite can be determined for arbitrary epochs

$$(a, e, i, \Omega, \omega, \overline{M}) \rightarrow r, v. \tag{3.153}$$

The solution of (3.153) is called *ephemeris computation*. In the first step r and v are expressed in the orbital plane coordinate system (Fig. 3.21), also called the *perifocal system*, and they are then transformed into geocentric-equatorial coordinates.

Here the unit vectors P (in direction of the perigee) and Q (perpendicular to P) are introduced. With known orbital elements, the position vector r can easily be described as

$$r = r \cos v \, P + r \sin v \, Q, \quad (3.154)$$

where (cf. (3.38))

$$r = \frac{p}{1 + e \cos v}.$$
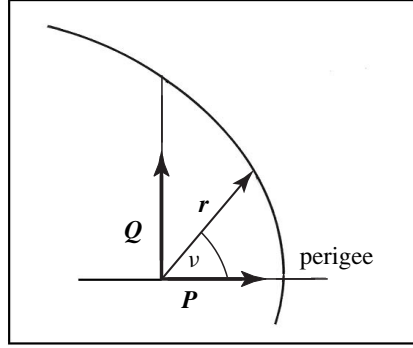
Figure 3.21. Unit vectors P and Q in the perifocal system

In order to obtain v, we have to form the first derivative of (3.154), assuming the orbital system to be time-invariant for the undisturbed motion, hence

$$\dot{P} = \dot{Q} = 0,$$

$$v = \dot{r} = (\dot{r} \cos v - r \dot{v} \sin v) \, P + (\dot{r} \sin v + r \dot{v} \cos v) \, Q. \quad (3.155)$$

If we refer to Fig. 3.7, p. 75, and 3.21, we will see that the horizontal component of velocity of a satellite is $v \cos \Phi$ which can also be expressed as $r \dot{v}$ in the perifocal system. We hence obtain from (3.66) and (3.73)

$$h = rv \cos \Phi = r^2 \dot{v}$$

and

$$p = \frac{h^2}{GM},$$

providing for the derivative of (3.38)

$$\dot{r} = \sqrt{\frac{GM}{p}} e \sin v \quad (3.156)$$

and

$$r \dot{v} = \sqrt{\frac{GM}{p}} (1 + e \cos v). \quad (3.157)$$

Substituting into (3.155), and simplifying, yields (e.g. Bate et al., 1971, p. 72)

$$v = \sqrt{\frac{GM}{p}} (-\sin v \, P + (e + \cos v) \, Q). \quad (3.158)$$

The conversion of a into p is given by (3.2), and the conversion of $\overline{M}$ into v follows from (3.54). The transformation from the orbital system into the equatorial system is

realized through rotations; elements of the rotation matrices are i, ω, Ω (cf. (3.163) and Fig. 3.22).

What follows is a well-known elementary description of the ephemeris computation, where equatorial Cartesian satellite coordinates X_S, Y_S, Z_S are derived from orbital elements

$$(a, e, i, \Omega, \omega, \overline{M}) \rightarrow X_s, Y_s, Z_s:$$

1. Calculate the eccentric anomaly E from the mean anomaly $\overline{M}$ with (3.55)

$$E = \overline{M} + \left(e - \frac{1}{8}e^3 + \frac{1}{192}e^5 - \frac{1}{9216}e^7 \right) \sin \overline{M} \dots .$$

2. Calculate the true anomaly ν from the eccentric anomaly E with (3.54)

$$\tan \nu = \frac{\sqrt{1 - e^2} \sin E}{\cos E - e}.$$

3. Calculate the distance r between the satellite and the center of gravity with (3.49)

$$r = a(1 - e \cos E).$$

4. Calculate Cartesian coordinates x_s, y_s of the satellite in the orbital system (Fig. 3.1, p. 64) with (3.5)

$$y_s = r \sin \nu, \quad \text{and} \quad x_s = r \cos \nu.$$

5. Calculate geocentric coordinates X_s, Y_s, Z_s of the satellite in the equatorial system (Fig. 3.22); the transformation is explicitly realized with the cosine relations between all axes:

$$
\begin{aligned}
\cos(xX) &= -\cos i \sin \omega \sin \Omega + \cos \omega \cos \Omega, \\
\cos(yX) &= -\cos i \cos \omega \sin \Omega - \sin \omega \cos \Omega, \\
\cos(xY) &= \cos i \sin \omega \cos \Omega + \cos \omega \sin \Omega, \\
\cos(yY) &= \cos i \cos \omega \cos \Omega - \sin \omega \sin \Omega, \\
\cos(xZ) &= \sin i \sin \omega, \\
\cos(yZ) &= \sin i \cos \omega.
\end{aligned}
\tag{3.159}
$$

Hence

$$
\begin{aligned}
X_s &= x_s \cos (xX) + y_s \cos (yX) = r \cos \delta \cos \alpha, \\
Y_s &= x_s \cos (xY) + y_s \cos (yY) = r \cos \delta \sin \alpha, \\
Z_s &= x_s \cos (xZ) + y_s \cos (yZ) = r \sin \delta.
\end{aligned}
\tag{3.160}
$$

The transformation (3.160) can also be written with matrices as

$$X_s = R x_s. \tag{3.161}$$

X_s and x_s are the position vectors of the satellite in the geocentric system and in the orbital system respectively. The rotation matrix is (cf. [2.1.1])

$$R = R_3(-\Omega) R_1(-i) R_3(-\omega), \tag{3.162}$$

with

$$R_3(-\omega) = \begin{pmatrix} \cos\omega & -\sin\omega & 0 \\ \sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

Figure 3.22. Orbital system and equatorial system

$$R_1(-i) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos i & -\sin i \\ 0 & \sin i & \cos i \end{pmatrix}, \quad \text{and} \quad R_3(-\Omega) = \begin{pmatrix} \cos\Omega & -\sin\Omega & 0 \\ \sin\Omega & \cos\Omega & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In the *boundary value problem* observations of different epochs are given. These can be, for example:

     distances (Laser, GPS),

     distance differences (Doppler),

     directions (satellite photography with cameras or CCD), or

     3D-positions (GPS, PRARE, DORIS).

For a solution of the problem we identify in Fig. 3.22 the following relations:

$$\begin{pmatrix} X_s \\ Y_s \\ Z_s \end{pmatrix} = r \begin{pmatrix} \cos(\nu + \omega)\cos\Omega - \sin(\nu + \omega)\sin\Omega\cos i \\ \cos(\nu + \omega)\sin\Omega + \sin(\nu + \omega)\cos\Omega\cos i \\ \sin(\nu + \omega)\sin i \end{pmatrix}, \tag{3.163}$$

$$r = \frac{a(1 - e^2)}{1 + e\cos\nu} = a\,(1 - e\cos E).$$

Further we know from ellipsoidal geometry (3.2):

$$p = a(1 - e^2).$$

For the velocity vector components we have (e.g. Rothacher, 1992, p. 224):

$$\dot{x}_s = -\sqrt{\frac{GM}{p}}\sin\nu, \quad \dot{y}_s = \sqrt{\frac{GM}{p}}(e + \cos\nu), \quad \dot{z}_s = 0. \tag{3.164}$$

The velocity vector in the equatorial coordinate system is then

$$\dot{\boldsymbol{r}}_s = \begin{pmatrix} \dot{X}_s \\ \dot{Y}_s \\ \dot{Z}_s \end{pmatrix} = \boldsymbol{R}_3(-\Omega)\boldsymbol{R}_1(-i)\boldsymbol{R}_3(-\omega) \begin{pmatrix} \dot{x}_s \\ \dot{y}_s \\ \dot{z}_s \end{pmatrix}. \tag{3.165}$$

The coordinates X_s, Y_s, Z_s are written as functions of the observations. These relations are then used for the formulation of observation equations, which describe a functional relationship between the observations on the one side and the elements as unknown parameters on the other side. Approximations for the unknowns are usually taken from launching conditions or available ephemerides. For partials of position and velocity with respect to osculating elements see (e.g. Rothacher, 1992, p. 227).

With spaceborne GPS receivers onboard a satellite the coordinates X_s, Y_s, Z_s can be determined directly [3.3.2.3]. Examples of orbit determination from two positions or three sets of angles are given in Montenbruck, Gill (2000, p. 40ff).

3.3.2   Integration of the Perturbed Orbit

The integration of the equation (3.97) for the perturbed satellite motion

$$\ddot{\boldsymbol{r}} = -\frac{GM}{r^3}\boldsymbol{r} + \boldsymbol{k}_s$$

can be worked out analytically or numerically.

In the *analytical orbit integration* an attempt is made to find algebraic expressions for all acting forces of interest, and to integrate them in a closed form. Usually a representation in form of parameters, in preference the Keplerian elements, is applied.

In the *numerical orbit integration* all forces acting at a particular satellite position are explicitly calculated, and they are then used as starting conditions for a step-wise integration. Hence, the accelerations are integrated directly.

3.3.2.1   Analytical Methods of Orbit Integration

The analytical orbit integration can be solved without difficulties for the undisturbed Keplerian motion, using the formulas from section [3.3.1]. In the case where only perturbations caused by the gravitational potential of the central body are present, a first order perturbation theory has already been given with the solution of Lagrange's planetary equations (3.105), namely with the development of the disturbing potential in form of the equation (3.112). With a linear superposition of the particular n-, m-, p-, q-terms in the perturbation equations for the elements, following (3.119) and (3.120), the time-dependent variation of each particular element can be calculated. This yields

$$\begin{aligned} \Omega(t) &= \Omega_0(t) + \Delta\Omega(t) \\ i(t) &= i_0(t) + \Delta i(t) \\ &\ \vdots \end{aligned} \tag{3.166}$$

$\Omega_0(t), i_0(t), \ldots$ are the mean Keplerian elements at epoch t. Using the transformation equations (3.153) to (3.158)

$$(a, e, i, \omega, \Omega, \overline{M})_t \to (\boldsymbol{r}, \boldsymbol{v})_t,$$

we can describe the satellite motion within Earth's gravity field. In this representation no coupling between the different perturbation terms has been considered. A more general formulation of (3.166) is

$$
\begin{aligned}
\Omega(t) &= \Omega_0(t_0) + (t - t_0)\, F_1(a_0, e_0, \Omega_0, i_0, \omega_0, \overline{M}_0, \text{perturbation parameters}) \\
&\quad + F_2(a_0, e_0, \Omega_0, i_0, \omega_0, \overline{M}_0, \text{perturbation parameters}, t - t_0), \\
i(t) &= \ldots, \\
&\ \vdots
\end{aligned}
$$

Perturbation parameters can be, for example, the coefficients of the terrestrial gravitational potential or characteristic parameters of the other perturbations. The term $F_1(t - t_0)$ describes secular perturbations, which grow out of limits with increasing $(t - t_0)$. The term F_2 summarizes periodic perturbations.

The explicit formulation of (3.166) is extremely laborious and requires much algebra. Computer-assisted algorithms are used for the computation of the rather complex expressions (computer-algebra). Particularly cumbersome is the analytical treatment of non-gravitational perturbations like air drag and radiation pressure, because discontinuities may occur. For such "non-conservative" force fields the Gaussian form of Lagrange's perturbation equations (3.108) is preferable.

In order to have a chance of solving the analytical problem, we must proceed in two steps; the true orbit is divided into two portions:

  (a) an approximate orbit is selected as a reference orbit, e.g. the undisturbed Kepler motion, and
  (b) the deviations between the reference orbit and the true orbit are formulated as perturbations.

The analytical orbit integration remains, nevertheless, an approximate solution, because it results from the truncation of series expansions. Today the analytical methods have almost completely disappeared from the field of routine orbit determination, because of the increasing requirements for the accuracy of the orbits. More than 500 $nmpq$-terms have to be determined in order to model the ephemeris of a typical geodetic satellite with an accuracy better than 1 m (Goad, 1977). Modern applications demand at least 1 cm orbit consistency for very long arcs (Beutler et al., 1998).

For several aspects of celestial mechanics and satellite geodesy the analytical orbit theory still plays an important role, in particular for understanding phenomena like resonances. It should be emphasized that early Earth gravity field models resulting from satellite observations (e.g. SAO Standard Earth I, II, III [12.2]) were all based on analytical developments.

The advantages and disadvantages of analytical orbit integration methods are summarized as follows:

Main advantages

- The relations and dependencies between disturbing forces and variations of elements can be explicitly formulated and studied. The characteristic behavior of the orbital motion can be identified, and predictions of the long-term stability and the development of orbits can be made.

- The numerical value of an individual orbital element at a particular epoch can be determined with a single evaluation of equation (3.166).

Main disadvantages

- Near-Earth satellites are very sensitive to perturbations, so the algebraic expressions rapidly become rather complex and bulky.

- Perturbations, caused by non-conservative forces like the solar radiation-pressure, are discontinuous functions and thus difficult to model with analytical expressions.

- Analytical solutions are approximations because they depend on truncated series expansions.

- Singularities occur for elliptical elements, when $e = i = 0$.

- The efficiency of computation is rather low, because of the many trigonometric functions in the algebraic expressions.

3.3.2.2   Numerical Methods of Orbit Integration

The numerical methods are distinguished by their simplicity and universal applicability when compared with analytical methods. With the use of the modern computer techniques the numerical effort only plays a minor role. This is why numerical methods are now used almost exclusively for orbit computations in satellite geodesy.

One basic requirement for the numerical integration is a suitable *orbit determination method*, for example a procedure named after Cowell or Encke (Noton, 1998). The *method of Cowell* (1910) was developed at the beginning of the last century and has been applied to the orbit determination of Halley's comet and the moons of Jupiter. With the availability of fast and efficient computers the method is now particularly suitable. The basic idea is that the equation of motion (3.97)

$$\ddot{\boldsymbol{r}} = -\frac{GM}{r^3}\boldsymbol{r} + \boldsymbol{k}_s,$$

including all perturbations, is integrated stepwise. Equation (3.97) can be re-written in the form of two first order differential equations:

$$\dot{\boldsymbol{r}} = \boldsymbol{v}, \quad \dot{\boldsymbol{v}} = \boldsymbol{k}_s - \frac{GM}{r^3}\boldsymbol{r}, \tag{3.167}$$

and further broken down into vector components:

$$\dot{x} = v_x, \quad \dot{v}_x = k_{sx} - \frac{GM}{r^3}x,$$

$$\dot{y} = v_y, \quad \dot{v}_y = k_{sy} - \frac{GM}{r^3}y, \qquad (3.168)$$

$$\dot{z} = v_z, \quad \dot{v}_z = k_{sz} - \frac{GM}{r^3}z.$$

The perturbing acceleration k_s, following the equations of section [3.2.3], can as well be written in the form $\dot{r} = v$ and $\dot{v}$ for each particular perturbation effect, e.g. the gravitational attraction of the Moon or of the Sun. The satellite state (r, v) for the desired epoch is then calculated using suitable methods of numerical integration (see later).

The main advantage of Cowell's method lies in its conceptual simplicity. The influences of all individual perturbations can be considered at the same time. One disadvantage is that smaller integration steps are required near a large attracting body, which leads to an increase of computation time and to an accumulation of round-off errors. In such cases some improvements can be made by formulating the equation (3.167) in spherical polar coordinates (r, θ, Φ). The integration steps can be larger because of much smaller coordinate variations. The equations of motion are then (cf. Bate et al., 1971, p. 389):

$$\ddot{r} - r(\dot{\theta}\cos^2\Phi + \dot{\Phi}^2) = -\frac{GM}{r^2}$$

$$r\ddot{\theta}\cos\Phi + 2\dot{r}\dot{\theta}\cos\Phi - 2r\dot{\theta}\dot{\Phi}\sin\Phi = 0 \qquad (3.169)$$

$$r\ddot{\Phi} + 2\dot{r}\dot{\Phi} + r\dot{\theta}^2\sin\Phi\cos\Phi = 0.$$

In Cowell's method the total force acting on the satellite is integrated, whereas in the *method of Encke* (1857) a reference orbit is introduced, and only the difference between the primary acceleration and all perturbing accelerations is subject to integration. Usually an osculating Kepler ellipse is chosen as a reference trajectory for the initial epoch t_0 (Fig. 3.23). Hence, one portion of the total acceleration is separated from the numerical integration, and is given a simplified analytical solution. The osculating orbit serves until the deviations from the true orbit become too large.
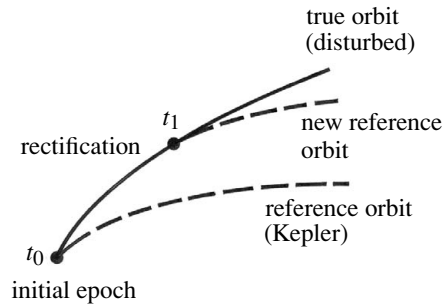
Figure 3.23. Numerical integration of the orbit, after Encke

Then a new ellipse is introduced as a reference orbit for the new initial epoch t_1, with the true position and velocity vector valid for t_0. This process is called *rectification*.

Let r be the position vector of the true perturbed orbit and ρ the position vector of the osculating (reference) orbit for a particular epoch $\tau = t - t_0$. Then we find for the true orbit

$$\ddot{r} + \frac{GM}{r^3} r = k_s, \qquad (3.170)$$

and for the osculating orbit

$$\ddot{\rho} + \frac{GM}{\rho^3} \rho = 0.$$

For the initial epoch t_0 we have

$$r(t_0) = \rho(t_0) \quad \text{and} \quad v(t_0) = \dot{\rho}(t_0).$$

Figure 3.24. Explanation of Encke's method of orbit integration

Referring to Fig. 3.24, we introduce the deviation vector δr between the reference orbit and the true orbit (we skip the parameter t for the sake of simplicity):

$$\delta r = r - \rho, \quad \delta\ddot{r} = \ddot{r} - \ddot{\rho}. \qquad (3.171)$$

Substituting (3.170) into (3.171) yields

$$\delta\ddot{r} = k_s + \left(\frac{GM}{\rho^3} \rho - \frac{GM}{r^3} r \right) = k_s + \left(\frac{GM}{\rho^3} (r - \delta r) - \frac{GM}{r^3} r \right)$$

$$\delta\ddot{r} = k_s + \frac{GM}{\rho^3} \left(\left(1 - \frac{\rho^3}{r^3} \right) r - \delta r \right). \qquad (3.172)$$

The deviation vector δr can be calculated for arbitrary epochs $\delta r(t_0 + \Delta t)$ with numerical integration techniques. Equation (3.172), however, may lead to numerical difficulties, because the expression $(1 - \rho^3/r^3)$ nearly equals zero. These difficulties can be solved by substituting

$$2q = 1 - \frac{r^2}{\rho^2}, \qquad (3.173)$$

and series expansions (Bate et al., 1971, p. 393). With δr changing much more slowly than r, Encke's method usually requires fewer integration steps, and consequently less computer time than Cowell's method (Taff, 1985, p. 393).

Generally speaking, Cowell's method is adequate when the perturbing acceleration equals or exceeds the primary acceleration, and Encke's method is more suitable when perturbing accelerations are small compared with the primary acceleration (Battin et al., 1978).

Many alternative orbit determination methods have been developed in celestial mechanics (see e.g Stumpff, 1959/1965/1974; Roy, 1978; Battin et al., 1978); none have particular advantages when compared with the two above-mentioned methods.

The *numerical integration* itself is realized with methods taken from approximation theory. Basically a polynomial has to be fitted to a limited series of consecutive points, in order to generate an additional point through extrapolation of the polynomial. This process is repeated at will. The polynomial coefficients are derived from the given points and their derivatives, based on the equation of motion. Different methods are used, depending on the number of points required, on the extrapolated values, on the smoothing process, and so on. They are generally subdivided into single-step and multi-step methods.

A well-known member of the family of *single-step methods* is the *Runge–Kutta method*. Here, a Taylor series of a certain order is used as an extrapolation function. A special feature of the single-step method is that only the last integration step is used, so the knowledge of the "history" of the function to be integrated is neglected. To avoid this *multi-step solutions* are usually applied in satellite geodesy. They are also called *predictor-corrector methods*.

The basic idea is to first predict a satellite position with a certain algorithm and then to correct that position. As a first step a predicted value X_{n+1} is calculated from X_n. It is then substituted in the differential equation of the process in order to obtain with $\dot{X}_{n+1}$ a corrected value of X_{n+1}. The process can be iterated, until the result does not change. Many predictor-corrector formulas may be found in the literature. The formulas of *Adam–Moulton* or *Gauss–Jackson* are frequently used. In principle these are filter techniques; this is also why the *Kalman filter* is now used more and more (Battin et al., 1978). Here r and v are state vectors with a relevant variance-covariance matrix, see e.g. Egge (1985).

Two main error sources have to be considered when numerical integration methods are applied. These are round-off errors and truncation errors. The *round-off errors* depend on the numerical representation accuracy in the computer being used. In order to limit these influences, rather large step-sizes are an advantage in the integration. *Truncation errors* develop when the last terms of a series expansion, which is used for the integration, are cut off. The errors can be minimized with small step-sizes. These two conflicting requirements have to be fulfilled with appropriate compromises.

Except for these deficiencies, the numerical integration methods can be considered as rigorous orbit integration methods, without approximations. The only disadvantage is that many, unwanted, intermediate satellite positions have to be calculated before the final solution is obtained.

For a detailed treatment of current numerical integration solutions, including exercises, see Montenbruck, Gill (2000). A short review is given by Beutler et al. (1998).

3.3.2.3 Precise Orbit Determination with Spaceborne GPS

A new technique of orbit determination evolved with the placement of GPS receivers into space vehicles, to determine directly the position and velocity vector of the space-

craft. A first experiment was flown on TOPEX/POSEIDON (Melbourne et al., 1994b), cf. [9.2]. Since then, GPS receivers have been positioned on a number of satellites, in particular on *Low Earth Orbiters* (LEO) like CHAMP or GRACE [10.2]. Three basic strategies are distinguished for *Precision Orbit Determination* (POD). They are:

- − the dynamic strategy,
- − the kinematic strategy, and
- − reduced-dynamic strategy.

The *dynamic* strategy corresponds to the classical approach of dynamic orbit modeling as treated in the previous chapters. A mathematical model of the forces acting on the satellite is used to estimate the accelerations over time and to integrate the equation of motion. A final trajectory is then estimated by a best (e.g. least squares) fit to the GPS measurements. In this procedure the effect of noisy measurements on the solution is reduced whereas the dynamical model remains unchanged. The dynamic strategy is essential in missions where forces have to be modeled, for example in gravity field missions with LEO observations.

In the *kinematic* or non-dynamic strategy the high accuracy potential of the GPS position estimates is exploited. This is in particular true for low (e.g. 400 km) orbits where an accurate modeling of perturbing accelerations is difficult to make. The rationale for this is that the actual path of the LEO may be closer to the GPS position estimates than to the trajectory determined from a dynamic model. The kinematic strategy is based on an underlying dynamic model. However, the dynamic modeling errors are avoided by high-weighting of the precise GPS observations.

The *reduced-dynamic* strategy combines the advantages and reduces the disadvantages of the two previous strategies. The kinematic technique neglects any existing dynamic information on the platform behavior in order to eliminate dynamic mis-modeling but maintains the GPS measurement noise. The dynamic technique neglects the high inherent precision of GPS measurements but maintains dynamic mis-modeling. A combination of both techniques in a Kalman filter process with appropriate weighting counterbalances the disadvantages.

The basis of the strategy is again correction of the dynamic solution with continuous GPS data. The key factor is a proper weighting of the Kalman filter process noise. The process noise model is characterized by two parameters, the process variance V, and the time constant (correlation-length) T. When the time constant is large and the variance approaches 0, the geometric information is suppressed and the model depends exclusively on the dynamic strategy. When the time constant $T = 0$ (white noise) and the process variance is high, the solution depends mainly on the geometric data. For more information see Melbourne et al. (1994b); Schwintzer et al. (1995); Yunck (1996).

3.3.3 Orbit Representation

For many practical applications in satellite geodesy only a short part of the orbit is used, for instance that portion which can be directly observed from the participating

stations. In such cases simplified orbit representation techniques, which do not require orbital dynamics and the correct modeling of the acting forces, may be sufficient. Instead, aspects of computer speed and memory requirements are well to the fore. The following procedures are frequently used:

— modeling of the deviations from a Keplerian orbit,

— polynomial representation, and

— "short-arc" representation with a simplified force-model.

In many cases a combination of the different representations is used.

3.3.3.1 Ephemeris Representation for Navigation Satellites

Here low memory requirements and efficient algorithms are of particular importance. Two operational navigation systems have been in use: the TRANSIT system [6] until 1996, and currently the Global Positioning System (GPS) [7]. For both, the ephemeris representation is based on a mean Kepler ellipse with secular terms, e.g. $d\Omega/dt$, di/dt, and $d\omega/dt$. The time-dependent deviations of the predicted orbit from this reference ellipse are transmitted to the user. In the so-called TRANSIT *broadcast ephemerides* [6.2.2] the differences, in the three components of the orbital Cartesian system, were transmitted for every even minute (UTC) (cf. [3.2.1.3], Fig. 3.13). These are:

$$\begin{aligned} \Delta E(t) \quad & \text{correction in the direction of the motion,} \\ \Delta a(t) \quad & \text{correction in the radial component, and} \\ \eta(t) \quad & \text{component perpendicular to the orbital plane.} \end{aligned} \qquad (3.174)$$

If the satellite coordinates are required for epochs in between the even minutes, an appropriate interpolation algorithm, such as a polynomial interpolation [3.3.3.2] or a short-arc-solution [3.3.3.3], has to be applied.

For GPS ephemerides [7.1.5.2] the differences are given in the form of harmonic coefficients for modeling time-dependent sine and cosine correction terms, namely (cf. Fig. 7.12, p. 225):

$$\begin{aligned} C_{us}, C_{uc} \quad & \text{amplitudes of the harmonic correction terms} \\ & \text{to the argument of latitude u} = v + \omega, \\ C_{is}, C_{ic} \quad & \text{amplitudes of the harmonic correction terms} \\ & \text{to the inclination angle } i, \\ C_{rs}, C_{rc} \quad & \text{amplitudes of the harmonic correction terms} \\ & \text{to the orbit radius.} \end{aligned} \qquad (3.175)$$

This representation is continuous in time and thus suitable for real-time applications; no interpolation is required. Each representation is, however, only valid for a limited time span, for example one or two hours. A smoothing algorithm may be necessary to remove the "jump" between adjacent portions of the orbit representation.

For GLONASS satellites the structure of the navigation message is different (Stewart, Tsakari, 1998). Every 30 minutes the geocentric vector components for position and velocity as well as for the lunisolar acceleration are transmitted. The user has to apply adequate interpolation algorithms for intervening observation epochs (cf. [7.7.1]).

3.3.3.2   Polynomial Approximation

The main advantages of this procedure are the simplicity of computation and the rather modest requirements for computer-time and memory. The main disadvantage is that polynomials are not suitable for the representation of trajectories longer than one or more revolutions. For this reason polynomials are not used for orbit predictions. On the other hand, polynomials can be successfully used for the representation of short arcs. The computing time required for the representation of GPS orbits with polynomials is considerably less than for conventional orbit representation. This is why polynomials are preferable for real-time navigation, when small field computers are used. Linear approximation functions of the type

$$F(x) = a_0\varphi(x) + a_1\varphi_1(x) + \cdots + a_m\varphi_m(x) \tag{3.176}$$

are common in practice. Different *base functions* $\varphi_i(x)$ can be selected. Well known are power series $\varphi_i(x) = x^i$, i.e. the approximation with polynomials

$$F(x) = a_0 + a_1x + \cdots + a_mx^m. \tag{3.177}$$

Other common selections in satellite geodesy are trigonometric polynomials and Chebyshev polynomials.

For the approximation of orbits from the TRANSIT-type (broadcast ephemerides), the following *trigonometric base functions* have been used (Wells, 1974):

$$\varphi = \{1, t, \sin 2\bar{n}t, \cos 2\bar{n}t\} \tag{3.178}$$

where $\bar{n}$ is the mean motion of the satellite. $F(x)$ can be approximated by

$$P = \sum_{i=0}^{3} a_i\varphi_i \tag{3.179}$$

and $\dot{F}(x)$ by

$$\dot{P} = \sum_{i=0}^{3} a_i\dot{\varphi}_i, \tag{3.180}$$

with

$$\dot{\varphi} = \{0, 1, 2\bar{n}\cos\bar{n}t, -2\bar{n}\sin\bar{n}t\}. \tag{3.181}$$

The first derivatives of the base functions are helpful when both the satellite coordinates $X(t)$ and the velocities $\dot{X}(t)$ are given. This was, for instance, the case with the

TRANSIT "precise ephemerides" for every minute, and it is the case for the GLO-NASS "broadcast ephemerides" [7.7.1].

Higher order Chebyshev polynomials are also frequently used for the approximation of satellite orbits. Here we find for the satellite coordinates, velocities and accelerations (Kouba, 1983c):

$$X(t) = \sum_{i=0}^{n} C_{X_i} T_i(\tau), \quad \dot{X}(t) = \sum_{i=1}^{n} C_{X_i} T_i'(\tau), \quad \ddot{X}(t) = \sum_{i=2}^{n} C_{X_i} T_i''(\tau), \quad (3.182)$$

with

$$\tau = \frac{2}{\Delta t}(t - t_0) \quad \text{and} \quad t \in \langle t_0; (t_0 + \Delta t) \rangle.$$

t_0 and Δt are the starting epoch and the length of the fitting interval, n is the polynomial order, C_{X_i} are the adjusted Chebyshev coefficients for the satellite coordinates x, y, z. The Chebyshev polynomials T_i and their derivatives are determined recursively:

$$
\begin{aligned}
T_0(\tau) &= 1, \\
T_1(\tau) &= \tau, \\
T_n(\tau) &= 2\tau T_{n-1}(\tau) - T_{n-2}(\tau); |\tau| \le 1, \quad n \ge 2; \\
T_1'(\tau) &= \frac{d\tau}{dt} = \dot{\tau}, \\
T_2'(\tau) &= 4\tau\dot{\tau}, \\
T_n'(\tau) &= \frac{2n}{n-1}\tau T_{n-1}'(\tau) - \frac{n}{n-2}T_{n-2}'(\tau); \quad n \ge 3, \\
T_2''(\tau) &= 4(\dot{\tau})^2, \\
T_3''(\tau) &= 24\tau(\dot{\tau})^2, \\
T_n''(\tau) &= \frac{2n}{n-1}(\dot{\tau}T_{n-1}'(\tau) + \tau T_{n-1}''(\tau)) - \frac{n}{n-2}T_{n-2}''(\tau), \quad n \ge 4.
\end{aligned}
\quad (3.183)
$$

For TRANSIT orbits a polynomial order of $n = 8$ to 10 proved to be successful. For GPS orbits an order of 7 to 8 is sufficient. The advantage of the Chebyshev polynomials compared to other polynomial representations is that they give a much better approximation, even at the limits of the interval.

In addition to the previously mentioned representations of TRANSIT orbits (broadcast and precise ephemerides), polynomial representations are suitable for fitting offsets between the adjacent sections of the GPS broadcast orbits. These offsets can reach several decimeters and thus may complicate continuous carrier phase solutions [7.1.5.2]. After a polynomial approximation the remaining offsets are below 2 cm.

In satellite laser ranging a smoothing of the orbit with polynomials aids the detection of blunders and a first estimation of the observation accuracy [8.4.2]. Finally, large numbers of single observations can be condensed into *normal points* via smoothing functions. Such normal points play an important role in laser ranging, and they are also applied in the evaluation process of GPS observations.

A *spatial smoothing of orbits* with polynomials was proposed early on by Wolf (1967, 1970) in connection with the geometrical evaluation of satellite triangulations. Through this procedure all observations could be referred to the same "smooth" orbital arc.

3.3.3.3 Simplified Short Arc Representation

In the *short arc method* forces are considered in the process of orbit approximation. Here a short arc is defined as a portion of an orbit of less than one revolution. In this case it is usually sufficient to use a potential series expansion up to degree and order (10,10) for satellites in TRANSIT orbits ($h \approx 1000$ km) and only (4,4) or (6,6) for GPS or GLONASS satellites ($h \approx 20\,000$ km).

Starting with initial conditions, for instance from a broadcast ephemeris, the orbital arc is determined using a method of numerical integration [3.3.2.2]. Depending on the quality of the approximate start values, several iteration steps may be necessary.

Compared with alternative approximation techniques, the short arc method requires a lot of computer time. It is, however, possible (Kouba, 1983b), to arrive at equally good results with a simplified and much faster short arc algorithm, when accelerations, caused by Earth's gravity field, are introduced into a polynomial approximation. The accelerations acting on the satellite can be derived from the well known potential expansion up to degree and order N, and be calculated, for example every minute. Equation (3.109) yields

$$V = \frac{GM}{r}\left(1 + \sum_{n=1}^{N}\sum_{m=0}^{n}\left(\frac{a_e}{r}\right)^n P_{nm}(\sin\Phi)\{C_{nm}\cos m\lambda + S_{nm}\sin m\lambda\}\right),$$

where Φ, λ are the geocentric latitude and longitude of the satellite. The components of acceleration at the satellite location, described in a Cartesian coordinate system, are obtained using the chain rule:

$$\begin{aligned}
\ddot{x} &= \frac{\partial V}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial V}{\partial \Phi}\frac{\partial \Phi}{\partial x} + \frac{\partial V}{\partial \lambda}\frac{\partial \lambda}{\partial x}, \\
\ddot{y} &= \frac{\partial V}{\partial r}\frac{\partial r}{\partial y} + \frac{\partial V}{\partial \Phi}\frac{\partial \Phi}{\partial y} + \frac{\partial V}{\partial \lambda}\frac{\partial \lambda}{\partial y}, \\
\ddot{z} &= \frac{\partial V}{\partial r}\frac{\partial r}{\partial z} + \frac{\partial V}{\partial \Phi}\frac{\partial \Phi}{\partial z} + \frac{\partial V}{\partial \lambda}\frac{\partial \lambda}{\partial z}.
\end{aligned}\qquad(3.184)$$

The partial derivatives in (3.184) can be found in Anderle (1974) or Egge (1985). Rewriting (3.184) gives (Kouba, 1983c,b):

$$\ddot{x} = \frac{\partial V}{\partial x} + 2\omega\dot{y} + \omega^2 x, \quad \ddot{y} = \frac{\partial V}{\partial y} - 2\omega\dot{x} + \omega^2 y, \quad \ddot{z} = \frac{\partial V}{\partial z}.\qquad(3.185)$$

ω is Earth's rotation velocity; x, y, z are the satellite coordinates. The following

expressions also hold:

$$
\begin{aligned}
\frac{\partial V}{\partial x} &= \frac{\partial V}{\partial r}\frac{x}{r} - \frac{\partial V}{\partial(\sin\Phi)}\frac{zx}{r^3} - \frac{\partial V}{\partial\lambda}\frac{y}{r^2\sin^2\Phi}, \\
\frac{\partial V}{\partial y} &= \frac{\partial V}{\partial r}\frac{y}{r} - \frac{\partial V}{\partial(\sin\Phi)}\frac{zy}{r^3} - \frac{\partial V}{\partial\lambda}\frac{y}{r^2\sin^2\Phi}, \\
\frac{\partial V}{\partial z} &= \frac{\partial V}{\partial r}\frac{z}{r} - \frac{\partial V}{\partial(\sin\Phi)}\left(\frac{z^2}{r^3} - \frac{1}{r}\right).
\end{aligned}
\tag{3.186}
$$

Finally, the partial derivatives required in equation (3.186) are given by:

$$
\begin{aligned}
\frac{\partial V}{\partial r} &= -\frac{GM}{r^2}\Big(1 - \sum_{n=2}^{N}(n+1)\left(\frac{a_e}{r}\right)^n\sum_{m=0}^{n}P_{nm}(\sin\Phi) \\
&\qquad\times (C_{nm}\cos m\lambda + S_{nm}\sin m\lambda)\Big), \\
\frac{\partial V}{\partial(\sin\Phi)} &= \frac{GM}{r}\Big(\sum_{n=2}^{N}\left(\frac{a_e}{r}\right)^n\sum_{m=0}^{n}P'_{nm}(\sin\Phi) \\
&\qquad\times (C_{nm}\cos m\lambda + S_{nm}\sin m\lambda)\Big), \\
\frac{\partial V}{\partial\lambda} &= \frac{GM}{r}\Big(\sum_{n=2}^{N}\left(\frac{a_e}{r}\right)^n\sum_{m=0}^{n}P_{nm}(\sin\Phi) \\
&\qquad\times (-C_{nm}\sin m\lambda + S_{nm}\cos m\lambda)\Big).
\end{aligned}
\tag{3.187}
$$

The Legendre polynomials $P_{nm}(\sin\Phi)$ and their derivatives $P'_{nm}(\sin\Phi)$ with respect to Φ can be determined recursively (e.g. Tscherning et al., 1983; Egge, 1985). For potential coefficients $\overline{C}_{nm}$ and $\overline{S}_{nm}$, which are given in fully normalized form, a *denormation* is required:

$$
\begin{Bmatrix} C_{nm} \\ S_{nm} \end{Bmatrix} = \sqrt{k(2n+1)\frac{(n-m)!}{(n+m)!}}\begin{Bmatrix} \overline{C}_{nm} \\ \overline{S}_{nm} \end{Bmatrix},
\tag{3.188}
$$

with $k = 1$ for $m = 0$, and $k = 2$ for $m \neq 0$, (cf. Torge, 2001, p. 72).

Now, the satellite positions, velocities, and the accelerations (3.185) can be smoothed with a Chebyshev polynomial algorithm. For TRANSIT satellites the agreement with precise ephemerides was found to be at the 5 cm level (Kouba, 1983c; Schenke, 1984) with a series expansion of the potential up to degree and order (10, 10).

3.4   Satellite Orbits and Constellations

3.4.1   Basic Aspects

Based on the formulas which have been derived earlier in this chapter we can present some basic relations for a better understanding of satellite orbital motion. For the velocity of a satellite in the undisturbed orbit we found from the equation of energy (3.61) the equations (3.81):

$$
\begin{aligned}
\text{for an elliptical orbit (a) :} \quad & v^2 = GM\left(\frac{2}{r} - \frac{1}{a}\right), \\
\text{for a parabolic orbit (b) :} \quad & v^2 = \frac{2GM}{r}, \\
\text{for a hyperbolic orbit (c) :} \quad & v^2 = GM\left(\frac{2}{r} + \frac{1}{a}\right).
\end{aligned}
\tag{3.189}
$$

It should be stated again that the velocity does not depend on the particular shape of the orbit, but only on GM, r and a. The satellites used in satellite geodesy are mostly in orbits with very small eccentricities. Hence, equation (3.189a) can be rewritten for a circular orbit in order to estimate velocities

$$
v_c = \sqrt{\frac{GM}{r}}.
\tag{3.190}
$$

With the mean radius of Earth
$$
r_0 = 6370 \text{ km}
$$

and with
$$
GM = 3.986 \cdot 10^{20} \text{ cm}^3/\text{s}^2 = 3.986 \cdot 10^5 \text{ km}^3/\text{s}^2,
$$

a table for different altitudes of satellite orbits can be readily computed. The mean period of one revolution follows from Kepler's third law (3.43),

$$
n^2 a^3 = GM,
$$

and with the above numerical values as

$$
U = 84.491\left(\frac{r}{r_0}\right)^{\frac{3}{2}}.
\tag{3.191}
$$

Within Table 3.6 we only find mean velocities to a first approximation. For elliptical orbits the velocity decreases with increasing distance from Earth, and is at its maximum at the perigee (cf. (3.189a)). In the extreme case (parabola), the velocity has to be multiplied by $\sqrt{2}$. The minimal velocity of a body required to leave Earth's gravity field was determined in (3.86) to be 11.2 km/s.

The intersection of the satellite orbital plane with a non-rotating Earth would produce a great circle on Earth's surface. Fig. 3.25 shows this so-called *subsatellite track*.

Table 3.6. Mean orbital altitudes, velocities and periods of selected satellites

r [km]	h [km]	v_c [km/s]	U [min]	Examples
6 378	7	7.91	84.49	near Earth's surface
6 770	400	7.67	92.57	space station, gravity field missions
7 400	1 000	7.34	105.6	Earth observation satellites
7 730	1 360	7.18	112.9	TOPEX/POSEIDON
10 000	3 600	6.31	165.6	PAGEOS
12 300	5 900	5.69	226.2	LAGEOS
26 600	20 200	3.87	12h	GPS
42 160	35 790	3.07	23h 56m	geostationary satellite
384 400		1.02	27d 08h	Moon
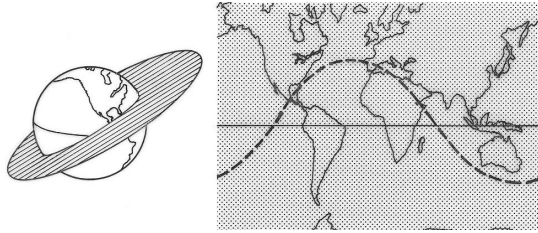
Figure 3.25. Subsatellite track
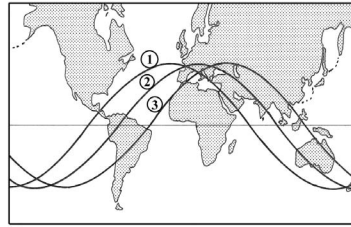
Figure 3.26. Western displacement of tracks

It is evident that the maximum northern or southern latitude which can be attained with the subsatellite track exactly equals the orbital inclination i, or $(180° - i)$ for retrograde satellites [3.3.3]. Due to Earth's rotation, subsequent satellite orbits show a western displacement with respect to each other (Fig. 3.26). The amount of this shift, $\Delta\lambda$, follows from the relationship between the satellite period and Earth's rotation rate; the latter being $23^h56^m04^s$ for a sidereal day. At the equator

$$\Delta\lambda = 0°.2507 \cdot U[\text{min}]. \tag{3.192}$$

The number of revolutions in a sidereal day is

$$R \approx \frac{1436}{U[\text{min}].} \tag{3.193}$$

In general R is not an integer. The ground track repeat period is the number of days needed to complete an integer number of revolutions. The ERS satellites, for example, with an orbital height of 770 km perform a total of 43 orbits in a period of 3 days.

A rough estimate of the geographical coordinates of subsatellite points can be determined rather easily. According to Fig. 3.22, equatorial coordinates α, δ of the

satellite can be derived from Cartesian coordinates with (2.15) or directly from orbital elements (e.g. Taff, 1985, p. 95). Some straightforward trigonometry gives, with $u = \omega + v$:

$$
\begin{aligned}
\tan(\alpha - \Omega) &= \tan(\omega + v)\cos i = \tan u \cos i, \quad \alpha = \Omega + \tan^{-1}(\cos i \tan u), \\
\sin \delta &= \sin(\omega + v)\sin i = \sin u \sin i, \quad \delta = \sin^{-1}(\sin i \sin u).
\end{aligned}
\tag{3.194}
$$

Hence, we find for the geographical coordinates of the subsatellite points (with Θ_0 the Greenwich siderial time):

$$
\Lambda = \alpha - \Theta_0, \quad \Phi = \delta.
\tag{3.195}
$$

For more accurate computations we have to consider Earth's ellipticity (cf. [2.1.4]) and the effect of nodal regression, caused by the J_2 term (cf. [3.2.2], Fig. 3.15).

The azimuth A and elevation angle E or zenith angle z of a satellite S_j as seen from a ground station B_i can be computed from the coordinates of the ground station r_i and of the satellite r_j (cf. Fig. 4.1 and (4.1)):

$$
r_j - r_i = \Delta r_{ij} = \Delta X = \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix}.
\tag{3.196}
$$

ΔX is given in the global geocentric system.

In order to express a satellite's position in the local tangent coordinate system (cf. Fig. 2.7) we refer to [2.1.3] (see also Torge (2001, chap. 6.2)). By substituting (2.28) into (2.31) we obtain the observation equation for the azimuth A, the zenith angle z, and the distance ρ:

$$
\begin{aligned}
A &= \arctan \frac{-\sin \Lambda \Delta X + \cos \Lambda \Delta Y}{-\sin \Phi \cos \Lambda \Delta X - \sin \Phi \sin \Lambda \Delta Y + \cos \Phi \Delta Z}, \\
z &= \arccos \frac{\cos \Phi \cos \Lambda \Delta X + \cos \Phi \sin \Lambda \Delta Y + \sin \Phi \Delta Z}{(\Delta X^2 + \Delta Y^2 + \Delta Z^2)^{1/2}}, \\
\rho &= (\Delta X^2 + \Delta Y^2 + \Delta Z^2)^{1/2}.
\end{aligned}
\tag{3.197}
$$

Note that Φ and Λ are the astronomical latitude and longitude. In approximate calculations Φ and Λ can be substituted by the ellipsoidal coordinates φ and λ. For an alternate set of formulas see e.g. (Montenbruck, Gill, 2000, p. 37).

3.4.2 Orbits and Constellations

Satellite orbits are commonly characterized by their orbital height. We distinguish

$$
\begin{aligned}
\text{LEO} &= \text{Low Earth Orbit; up to 2000 km,} \\
\text{MEO} &= \text{Medium Earth Orbit; 5000–20\,000 km,} \\
\text{GEO} &= \text{Geostationary Orbit; 36\,000 km.}
\end{aligned}
$$

In addition we have particular orbits such as

$$IGSO = \text{Inclined Geo-synchronous Orbit,}$$
$$HEO = \text{Highly Elliptical Orbit,}$$

and orbits without commonly used abbreviations such as the sun-synchronous orbit, geo-synchronous orbit, and polar orbit.

Low Earth Orbits

LEOs in satellite geodesy are mostly circular. Typically they may accommodate gravity field missions [10] (such as CHAMP, GRACE, or GOCE) at orbital heights of about 400 km, remote sensing satellites (such as SPOT, LANDSAT, ERS) at orbital heights of about 800–1000 km, and altimeter satellites [9] (such as TOPEX/POSEIDON, ENVISAT, JASON) at orbital heights of 1000–1500 km. LEOs are also used for communication satellite constellations like *Globalstar* and *Iridium*. The orbital period at these altitudes varies between 90 minutes and two hours. The radius of the satellite footprint (i.e. the area on the surface from where the satellite is visible above the horizon) is rather small and varies between 2000 and 4000 km.

Advantages: Low launch costs; low-power radio transmitters provide sufficient signal strength for simple receivers on the ground; the rapidly varying Doppler shift can be used for high precision navigation purposes (e.g. TRANSIT, DORIS [6]).

Disadvantages: Satellites are only in sight for 15 to 20 minutes, because of the short orbital period. Continuous data transfer to ground stations requires relay satellites in high orbits. For communication purposes a rather large number of LEOs is required because of the small footprint (e.g. more than 60 satellites for Iridium, $h = 780$ km).

Medium Earth Orbits

MEOs are used for constellations of navigation satellites such as GPS, GLONASS (about 20 000 km), or the European GALILEO (about 24 000 km). The laser satellites LAGEOS-1,2 (about 6000 km) also belong to this group. Circular MEOs are also called *Intermediate Circular Orbits* (ICO).

Advantages: Satellites are in view for several hours; communication satellite systems require less "satellite swapping". The orbits are not affected by atmospheric drag and hence are quite stable.

Disadvantages: Low Doppler shift; more expensive launch costs.

Geostationary Earth Orbits

GEOs are mainly used for communication satellites. A satellite placed into a circular orbit of inclination $i = 0°$ at an altitude of 35 800 km has a 24^h period and appears fixed to an Earth-bound observer. The footprint of a GEO satellite covers almost 1/3 of Earth's surface (from about 75 degrees N to about 75 degrees S), so that near-global coverage can be provided with a minimum of three satellites. These favorable characteristics have led to international regulations and the assignment of individual longitude slots ($\pm 0.°1$) to interested countries and agencies (Montenbruck, Gill, 2000).

Advantages: The orbits are very stable, and few satellites are required for global coverage.

Disadvantages: Rather high launch costs, limited slots in the geostationary belt, and no coverage around the poles.

Inclined Geosynchronous Orbits
IGSOs are circular orbits with a 24 hours period. They differ from GEOs in that they are inclined with respect to the equatorial plane. For an observer on Earth the satellite will move. The ground track is like a large figure eight.
Advantages: Excellent coverage of the areas close to the poles.

Highly Elliptical Orbits
HEOs typically have a perigee at about 500 km and an apogee as high as about 50 000 km. The orbits are near the "critical inclination" at 63.4 degrees (cf. (3.2.2)) in order to avoid rotation of the line of apsides. The sub-satellite point beneath the apogee is at latitude 63.4 North or South; HEOs hence provide communication services to locations in high northern or southern latitudes.

Polar Orbits
Polar orbits have an inclination of $i = 90°$. The orbits are fixed in space, and Earth rotates underneath. A single satellite in a polar orbit hence provides coverage of the entire globe. One example of a configuration with polar orbits is the "Navy Navigation Satellite System" (cf. [6.2]).

    A good overview of particular orbits for slow moving satellites is given by Hugentobler (1998).

Constellations
Constellations consist of multiple satellites, mainly of the same type, with similar orbits, but placed into suitably shifted orbital planes or rotated trajectories. One well known example is the Global Positioning System (GPS) [7].
LEO-constellation. These consist of about 48 to 65 LEO satellites at any time augmented by a number of payloads on GEO communication satellites in order to provide global coverage. *Advantages*: small, cheap, easily replaceable satellites, low-power radio transmitter, strong Doppler shift, high signal strength for simple user equipment. *Disadvantages*: short orbital period, only 15 to 20 minutes in view, requires rapid inter-satellite switching and a large number of ground control stations.
MEO-constellation. These are ideal for navigation purposes. It is currently used for GPS and GLONASS, and will be used for the European GALILEO. It provides global coverage except for polar areas. The performance can be augmented by a number of GEO satellites (e.g. EGNOS, WAAS [7.7.2]). MEOs are mostly distributed in several orbital planes, hence each plane can contain spare satellites. *Advantages*: excellent ground coverage except for polar areas; satellites remain visible for several hours. *Disadvantages*: slow moving satellites provide rather slow Doppler shift; high launch costs.
IGSO/GEO-constellation. These are regional systems consisting of several GEO satellites and several IGSO satellites. A system can at first be deployed to cover a given region and later expanded to other regions. *Advantages*: good coverage over polar

areas; excellent geometry and satellite visibility. *Disadvantages*: large power requirements for sufficient signal strength; high launch costs.

3.4.3   Sun-synchronous, Geostationary, and Transfer Orbits

The nodal regression of a satellite orbit can amount to several degrees per day, and it can be used to select specific satellite orbits. A given geographical area may be covered on a daily basis or with a pre-defined repetition rate, leading to a total coverage of Earth's surface between the extreme latitudes after a certain time of operation. This is of special importance for Earth-observing satellites. Two specific orbital selections are the sun-synchronous orbit and the geostationary orbit.

Sun-synchronous orbits are important for satellites which have to stay continuously within the sunlight, or which have to see the surface under specific illumination conditions. Such a situation is demonstrated in Fig. 3.27. The requirement is to have the ecliptical component of the orbital plane, with respect to inertial space, remain perpendicular to a line connecting Earth and Sun. This can only be achieved when the nodal motion, caused by C_{20}, exactly compensates the annual motion of Earth around the Sun, i.e. when
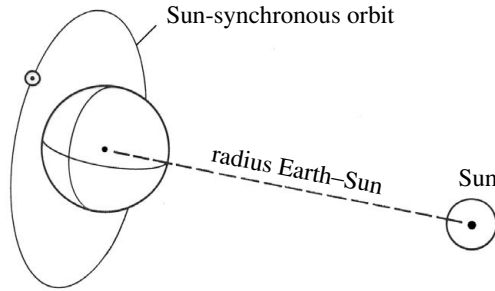
Figure 3.27. Sun-synchronous orbit

$$\frac{d\Omega}{dt} = 0°.9863/\text{day}. \tag{3.198}$$

Using (3.122), yields

$$\frac{d\Omega}{dt} = C_{20} \frac{3na_e^2}{2a^2(1-e^2)^2} \cos i.$$

With

$$n = \sqrt{\frac{GM}{a^3}} \quad \text{(following (3.43))}$$

and

$$GM = 3.986 \cdot 10^5 \text{km}^3 \text{ s}^{-2}, \quad a_e = 6380 \text{ km}, \quad C_{20} = -1.08 \cdot 10^{-3},$$

we obtain

$$\frac{d\Omega}{dt} = -\frac{2.0597 \cdot 10^{14} \cos i}{\sqrt{a^7}(1-e^2)^2} \, [°/\text{day}]. \tag{3.199}$$

The elements a, e, i have to be selected in such a way that $\dot{\Omega}$ equals $0°.9863/\text{day}$. In many cases the elements a and e are predetermined, in order to fulfill certain

requirements, and only the inclination i remains freely selectable. For a satellite at an altitude of 1000 km with a near circular orbit, for example, we get the following requirement for a sun-synchronous orbit

$$\cos i = -\frac{(0.9863)(7380)^{3.5}}{2.0597 \cdot 10^{14}}, \quad i = 99.°47.$$

If a satellite appears to hover more or less motionless over the same geographical location, it is called a *geostationary* satellite. From Fig. 3.25 it becomes clear that only equatorial orbits ($i = 0$) can be geostationary. The duration of one revolution must correspond exactly to Earth's rotation rate in inertial space, i.e. to $23^{\mathrm{h}}56^{\mathrm{m}}$. Thus we obtain with equation (3.43) and $n = 2\pi/86160$

$$a = \left(\frac{GM}{n^2}\right)^{\frac{1}{3}} \tag{3.200}$$

and find

$$a = 42\,165 \text{ km.}$$

A further condition is $e \approx 0$, in order to achieve a near-uniform orbital velocity. For orbital inclinations $i \neq 0$ the satellite ground track will be a "figure-eight" curve, oscillating from the equator to the north and to the south. A rigorous geostationary orbit cannot be realized because of the acting perturbing forces. However, an approximate geostationary orbit can be achieved with appropriate orbital corrections. For more details see Soop (1983); Maral, Bousquet (1986); Ploner (1996); Hugentobler (1998). Satellites with a period of revolution equal to that of Earth's rotation, but with any non-zero inclination or eccentricity are called *geosynchronous satellites*.

Placing a high altitude satellite in orbit is often achieved in several steps. First the satellite is put into a near-Earth, almost circular, *parking orbit* and then via an *ascent ellipse* into the final orbit. This is also the procedure when satellites are launched using the *Space Shuttle*.

A commonly used *transfer orbit* between the coplanar, nearly circular, orbits is the so-called *Hohmann transfer orbit* (named after *Hohmann*, who described the principle in 1925). The transfer itself results from two velocity changes at the perigee and at the apogee. From the geometry shown in Fig. 3.28 we find for the semi-major axis of the transfer orbit:

$$2a_t = r_1 + r_2. \tag{3.201}$$

The total energy of the transfer orbit is, from (3.78),

$$E_t = -\frac{GM}{2a} = -\frac{GM}{r_1 + r_2}. \tag{3.202}$$
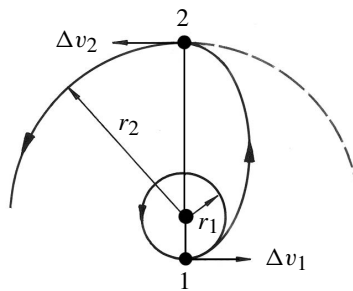
Figure 3.28. Hohmann transfer orbit

With the energy equation (3.76) we find for the velocity at point 1 (perigee) in the elliptical orbit:

$$v_1 = \sqrt{2\left(\frac{GM}{r_1} + E_t\right)}. \tag{3.203}$$

The speed which the satellite has already at point 1 in the lower orbit is, from (3.190),

$$v_{c_1} = \sqrt{\frac{GM}{r_1}}. \tag{3.204}$$

To move the satellite from the low circular orbit to the transfer ellipse, its speed has to be increased from v_{c_1} to v_1, thus

$$\Delta v_1 = v_1 - v_{c_1}. \tag{3.205}$$

A corresponding increase of the velocity is required at point 2 (apogee), to transfer from the elliptical orbit to the higher circular orbit, so

$$\Delta v_2 = v_{c_2} - v_2. \tag{3.206}$$

The following complete and closed formulas can be used (Kaplan, 1976):

$$\Delta v_1 = \sqrt{\frac{GM}{r_1}}\left(\sqrt{\frac{2(r_2/r_1)}{1 + (r_2/r_1)}} - 1\right), \quad \Delta v_2 = \sqrt{\frac{GM}{r_2}}\left(1 - \sqrt{\frac{2}{1 + (r_2/r_1)}}\right). \tag{3.207}$$

The same principles apply for the transfer from a higher orbit into a lower orbit.

The Hohmann transfer is the most economical solution with respect to the required speed change Δv. However, it takes longer than any other possible transfer orbit (Bate et al., 1971, p. 165). This is why in practice other transfer orbits are also applied.

To change the inclination of the orbit, velocity changes Δv perpendicular to the orbital plane are necessary. If an equatorial orbit is required, the Δv changes must be applied at one of the nodal points. This is essential, for example, when satellites from non-equatorial launching sites need to be placed in geostationary orbits.

Some points in orbit of interest to space transportation, or for the location of particular satellites (e.g. GAIA [5.3.3]) are the so-called *Lagrange* or *Libration points*, $L1$ to $L5$ (Fig. 3.29). These points are positions of equilibrium for a body in a two-body system, (e.g. the Earth-Moon or the Earth-Sun system). The points $L1$, $L2$, and $L3$ lie on a straight line through the main bodies and are points of unstable equilibrium; that is, a satellite positioned at one of these points will drift away caused by small perturbations and hence needs propellant for station keeping. In the Earth-Moon system $L1$ and $L2$ are lunar-stationary or lunar-synchronous points and, for an observer on the Earth, move with the same angular velocity as the Moon. The $L4$ and $L5$ points form an
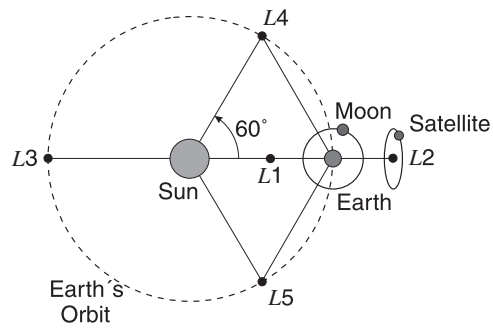
Figure 3.29. The five Lagrange points (libration points) in the Sun-Earth system

equilateral triangle with the other two bodies; they are in stable equilibrium. For more details see e.g. Roy (1978, chap. 5).

4 Basic Observation Concepts and Satellites Used in Geodesy

In this chapter, a brief overview of the principal observation techniques, and of the satellites that can be used, is given. A more detailed treatise follows in the chapters which are dedicated to the particular methods.

4.1 Satellite Geodesy as a Parameter Estimation Problem

The fundamental equation of satellite geodesy can be formulated as (see Fig. 4.1)

$$\boldsymbol{r}_S(t) = \boldsymbol{r}_B(t) + \boldsymbol{\rho}(t)$$

or

$$\boldsymbol{r}_j(t) = \boldsymbol{r}_i(t) + \Delta\boldsymbol{r}_{ij}(t). \tag{4.1}$$

For a solution to equation (4.1) we have to establish a relation between the observations, characterized by the vector $\Delta\boldsymbol{r}_{ij}(t)$, and the parameters which describe the satellite position $\boldsymbol{r}_j(t)$ as well as the location of the observation station $\boldsymbol{r}_i(t)$. In the estimation process either all parameters can be treated as unknowns, or some of the parameters are considered to be known, in order to stabilize and to simplify the solution.

In general, a nonlinear observation equation model between the observations and the parameters is introduced:

$$\boldsymbol{L} + \boldsymbol{v} = \boldsymbol{\Phi}(\boldsymbol{X}), \tag{4.2}$$

Figure 4.1. Basic relations for satellite observations

with

- $\boldsymbol{L}$ the vector of the observations,
- $\boldsymbol{X}$ the vector of the unknown parameters,
- $\boldsymbol{\Phi}$ a nonlinear vectorial function, and
- $\boldsymbol{v}$ the vector of the residuals, containing the unmodeled components of the total estimation process.

The observation equation (4.2) can be linearized when approximate values $\boldsymbol{X}_0$ are introduced for the unknown parameters. With

$$\boldsymbol{L}_0 = \boldsymbol{\Phi}(\boldsymbol{X}_0)$$

it follows that the vector of residual observations

$$l = L - L_0 \tag{4.3}$$

and the vector of residual parameters

$$x = X - X_0. \tag{4.4}$$

The linear form of (4.2) is then

$$l + v = Ax. \tag{4.5}$$

The *design matrix* A contains the partial derivatives of the observations with respect to the parameters, developed around the approximate point of expansion X_0:

$$A = \left(\frac{\partial \Phi(X)}{\partial X} \right)_0. \tag{4.6}$$

The system of equations (4.5) can be solved in a least-squares adjustment process, based on the minimization of the function

$$v^T P v \rightarrow \text{minimum}, \tag{4.7}$$

and yields a *best estimate* $\hat{X}$ of the unknown parameters. P denotes the weight matrix of the observations. For a full treatment of the subject see textbooks on adjustment or estimation techniques, for instance Wolf (1975); Pelzer (1985); Koch (1990); Niemeier (2002). Textbooks with special emphasis on GPS techniques are Leick (1995) and Strang, Borre (1997). The modern term *estimation* used in this book is equivalent to the classical term *adjustment*.

The parameters in equation (4.2) to (4.5) can be subdivided into different groups, for instance into:

(1) Parameters describing the geocentric motion of the observation station $r_B(t)$. The first of these are the geocentric station coordinates. Then there are geo-dynamic parameters, describing the relation between the Earth-fixed terrestrial reference system and the space-fixed inertial reference system, namely the po-lar motion and Earth rotation parameters. Also belonging to this group are the parameters used for the modeling of solid Earth tides and tectonic crustal deformations. Finally, the transformation parameters between geocentric and particular geodetic or topocentric reference frames may be considered.

(2) Parameters describing the satellite motion $r_s(t)$. These are, besides the satellite coordinates, the harmonic coefficients of Earth's gravity field, and parameters describing other gravitational or non-gravitational perturbations, like the solar radiation pressure.

(3) Parameters influencing directly the observations $\rho(t)$. These are e.g. atmospheric parameters, clock parameters, and signal propagation delays.

It is obvious that a complete solution, where all parameters are determined simultaneously, cannot be obtained by simple means. Certain conditions, or requirements, have to be introduced in order to avoid a singularity of the equation system. The combined determination of coordinates and gravity potential coefficients is often called a *satellite solution* or only a *solution*; the product is named an *Earth model* [12.2].

In general only a few particular parameters or groups of parameters are of interest, and the other parameters are considered to be known values. The satellite orbit is often treated as known when the coordinates of the observation stations are to be estimated. For the determination of polar motion parameters and universal time, the coordinates of the control stations and Earth's gravity field are usually introduced from other solutions.

The parameters of group (3), the so-called *bias parameters*, have a more technical meaning, and are mostly treated at a preprocessing stage.

Instead of estimating the parameters during the adjustment process, they can be eliminated through a suitable arrangement of observations. They are cancelled when simultaneous observations at different stations are differenced. This technique is often applied for the parameters of group (3). The satellite coordinates can also be eliminated, and need no longer be treated as unknown parameters, when observations are made simultaneously at a sufficiently large number of stations. The observation technique becomes more geometric in character [1.2]. This *geometrical concept* of satellite geodesy was used particularly with camera observations [5.1].

Geometrical methods are by their nature *relative methods*, and they do not provide geocentric coordinates. This is why some problems in satellite geodesy cannot be solved with the geometrical methods.

Dynamical methods of satellite geodesy are used, when a force model is required for the description of the satellite motion. The orbit must either be known from external sources, for instance from an ephemeris service, or the orbit must be determined within the adjustment process, be it completely, or partially. Different concepts are in use.

When simultaneous observations are available from at least two stations, corrections to some particular orbital elements or "degrees of freedom" (up to six in number) can be estimated for a short portion of the orbit, together with the other adjustment parameters. This procedure is also called "adjustment with a relaxed orbit", or the *semi-short-arc method*. No particular forces acting on the satellite are considered in this technique; rather some of the degrees of freedom in the satellite orbit (orbital parameters) are eliminated via a geometric procedure.

A further step is that for a short portion of the orbit, based on a given force field, some orbital elements are estimated as parameters from the observations. This procedure is called the *short-arc method*. For even longer orbits, of several revolutions, a larger number of perturbation parameters must be estimated (cf. Fig. 4.2). The term for this technique is the *long-arc method*. The method is used e.g. for the parameter estimation of polar motion, Earth rotation, solid Earth tides, and the anomalous gravity field of Earth.

Fixed orbit	Semi Short Arc (relaxed orbit)	Short Arc	Long Arc
0	3 − 6	> 6	degrees of freedom
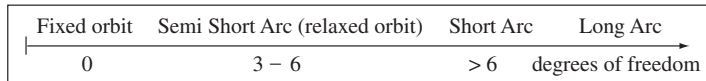
Figure 4.2. Role of the satellite orbit in the parameter estimation process

Long-arc methods are mostly used for the analysis of scientific problems. For operational tasks of applied satellite geodesy, the orbit is either taken as known, or a limited number of parameters are estimated for the orbit improvement within the adjustment process. For global radio navigation systems, like GPS, in general no orbit improvement is necessary because very precise orbits are available through particular services like the *International GPS Service* (IGS) [7.8.1].

With observations from a single station the parameter estimation process is usually restricted to the determination of the station coordinates only. The number of parameters can be increased when simultaneous observations are available from several stations; corrections to the satellite orbit and observation biases may then be estimated. For the solution of a general and global parameter estimation problem, observations to a large number of different satellites are required from many globally distributed stations. Fig. 4.3 contains a schematic representation concerning the process of observation and parameter estimation.

Satellite observations	**Transform raw data into observations**
Preprocessing	**Corrections for signal propagation, time, relativity, ambiguities, normal points**
Computed values from observations	**Numerical analysis, Geodesy, Orbital mechanics, Coordinate transformations**
Parameter estimation	**Station coordinates, Gravity field coefficients, Satellite positions, Polar motion, Earth rotation, Ocean and solid Earth tides, Geodynamic parameters, Atmospheric parameters Observation biases**
Estimation of accuracy and reliability	**Statistics, Reliability, Accuracy measures**

Figure 4.3. Functional scheme for the use of satellite observations

4.2   Observables and Basic Concepts

The observation techniques used in satellite geodesy can be subdivided in different ways. One possibility has been already introduced in [1.2], namely a classification determined by the location of the observation platform

- Earth based techniques (ground station → satellite),
- satellite based techniques (satellite → ground station),
- inter-satellite techniques (satellite → satellite).

Another classification follows from the observables in question. A summary of the most important operational techniques is given below. References to the specific artificial satellites are included. A graphical overview is given in Fig. 4.4. The detailed presentation and discussion of the individual observation methods follows in later chapters.
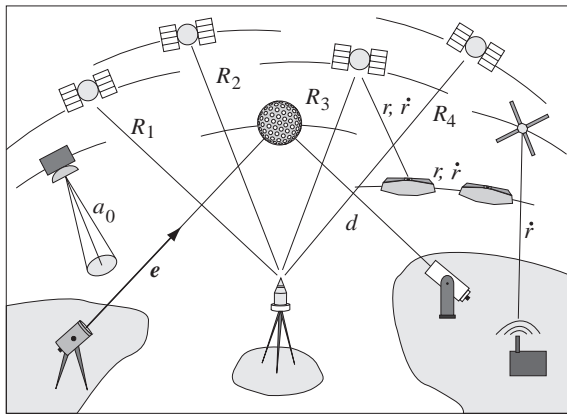
Figure 4.4. Overview of observation techniques in satellite geodesy

4.2.1   Determination of Directions

Photographical methods are almost exclusively used for the determination of directions. An artificial satellite which is illuminated by sunlight, by laser pulses, or by some internal flashing device, is photographed from the ground, together with the background stars. The observation station must be located in sufficient darkness on the night side of Earth. The stars and the satellite trajectory form images on a photographic plate or film in a suitable tracking camera, or on a CCD sensor [5.2]. The photogram provides rectangular coordinates of stars and satellite positions in the image plane, which can be transformed into topocentric directions between the observation station and the satellite, expressed in the reference system of the star catalog (equatorial system, CIS).

Two directions, measured at the same epoch from the endpoints of a given base-line between observing stations, define a plane in space whose orientation can be determined from the direction cosines of the rays. This plane contains the two ground-stations and the simultaneously observed satellite position. The intersection of two or more such planes, defined by different satellite positions, yields the inter-station vector between the two participating groundstations (Fig. 4.5). When more stations are involved, this leads to regional, continental, or global networks [5.1]. Note that these networks are purely geometric.

Direction measurements have also been used for orbit determination, and they were introduced into early comprehensive solutions for Earth models (gravity field coefficients and geocentric coordinates) [12.2]. Some satellites are equipped with laser-reflectors. In such cases the directions and ranges can be determined simultaneously, and provide immediately the vector $\rho(t)$ between the ground station and the satellite.

Figure 4.5. The use of directions with satellite cameras

Initially passive balloon satellites were used as targets, such as the first experimental telecommunication satellites, ECHO-1 and ECHO-2. In 1966 a dedicated geodetic balloon satellite, PAGEOS (*PA*ssive *GEO*detic *S*atellite), with a diameter of 30 m was launched for the observation of the BC4 World Network [5.1.5]. The satellite was placed into an orbit of about 3000 to 5000 km altitude and was used for about six years. Today, laser pulses can be reflected by satellites equipped with corner cube reflectors, and be used for the determination of precise directions. However, the achievable accuracy is very low when compared with modern ranging methods. This is why today the measurement of directions only plays a minor role in satellite geodesy. A certain revival took place with the launch of the Japanese Geodetic Satellite AJISAI in 1986 (Sasaki, 1983; Komaki et al., 1985), see [4.3.2]. A modern application is the use of powerful CCD-techniques for the directional observation of geostationary satellites [5.2.3].

Directional information can also be derived by analysis of electromagnetic signals transmitted from a satellite. The related techniques, which have been realized so far, yield only a rather low accuracy [4.4]. Very Long Baseline Interferometry (VLBI), on the other hand, is one of the most accurate observation techniques in geodesy (cf. [11.1]), and provides precise directions to extragalactic radio sources. The application of the VLBI technique using radio signals from artificial satellites is under discussion [11.1.4].

4.2.2 Determination of Ranges

For the determination of distances in satellite geodesy the propagation time of an electromagnetic signal between a ground station and a satellite is measured. According to the specific portion of the electromagnetic spectrum we distinguish between optical systems and radar systems.

Optical systems are weather-dependent. Laser light is used exclusively, in order to achieve the required signal strength and quality. *Radar systems* are weather-independent; wavelengths of the centimeter and decimeter domain are used. The propagation behavior, however, is significantly affected by atmospheric refraction.

We distinguish the one-way mode and the two-way mode. In the *two-way mode* the signal propagation time is measured by the observer's clock. The transmitter at the observation station emits an impulse at epoch t_j. The impulse is reflected by the satellite at epoch $t_j + \Delta t'_j$, and returns to the observation station where it is received at epoch $t_j + \Delta t_j$. The basic observable is the total signal propagation time Δt_j. Without considering relativistic effects, we find

$$\Delta t_j = 2\,\Delta t'_j.$$

Following Fig. 4.6, with c being the signal propagation velocity, we obtain the basic equation for the distance measurement in the two-way mode:

$$|r_j - r_i| = |\Delta r_{ij}| = \frac{1}{2}\,c \cdot \Delta t_j. \qquad (4.8)$$

One typical example for this mode is the *Satellite Laser Ranging* (SLR) technique [8]. The observer's clock can also be placed in the satellite, e.g. for spaceborne laser [8.7], or radio systems like PRARE [4.3.3.3].

In the *one-way mode* we assume that either the clocks in the satellite and in the ground receiver are synchronized with each other, or that a remaining synchronization error can be determined through the observation technique. This is, for instance, the case with the *Global Positioning System* (GPS) [7]. Equation (4.8) simplifies to

Figure 4.6. Concept of range measurements to satellites

$$|\Delta r_{ij}| = c \cdot \Delta t_j. \qquad (4.9)$$

Further we distinguish between either *impulse* or *phase comparison* methods. When a clear *impulse* can be identified, as is the case in satellite laser ranging, the distance is calculated from the signal propagation time using equation (4.8). With the

phase comparison method, the phase of the carrier wave is used as the observable. In the two-way mode the phase of the outgoing wave is compared with the phase of the incoming wave. In the one-way mode the phase of the incoming wave is compared with the phase of a reference signal generated within the receiver. In both cases the observed phase difference, $\Delta\Phi$, corresponds to the residual portion, $\Delta\lambda$, of a complete wavelength (cf. [2.3]). The total number, N, of complete waves between the observer and the satellite is at first unknown. This is the *ambiguity problem*. The corresponding equation in the two-way mode is

$$|\Delta \boldsymbol{r}_{ij}| = \frac{1}{2}(\Delta\lambda + N\lambda). \tag{4.10}$$

The factor $\frac{1}{2}$ disappears in the one-way mode. Different methods are used for the solution of the ambiguity term N, for example:

- measurements with different frequencies (e.g. SECOR [4.4.1]),
- determination of approximate ranges with an accuracy better than $\lambda/2$ (e.g. GPS with code and carrier phases),
- use of the changing satellite geometry with time (e.g. GPS carrier phase observations),
- ambiguity search functions (e.g. GPS).

The different methods are treated in more detail in the sections dealing with the specific observation techniques. In particular very powerful ambiguity search techniques have been developed for the precise use of GPS.

Impulse and phase comparison methods also show differences with respect to the signal propagation in the high atmosphere. Distances derived from phase measurements depend on the phase velocity v_p of the particular frequency, whereas distances derived from impulse measurements are based on the respective group velocity v_g. The ionosphere is a dispersive medium, and hence the phase velocity is greater than the group velocity [2.3.1.2].

Distinct satellites are used for the different methods of range measurements. For satellite laser ranging (SLR) the satellites are equipped with suitable reflectors. Dedicated laser satellites are for example LAGEOS-1,2 (*LA*ser *GEO*dynamic *S*atellite; orbital height about 5900 km), STARLETTE and STELLA (orbital height about 800 km). There are now about 70 satellites equipped with laser reflectors [8.2].

Those satellites designed to operate in the radio frequency domain must carry suitable electronic equipment and antennas. For the two-range mode this equipment is sometimes called a *transponder*. When ranges are determined simultaneously between three known ground stations and at least three different satellites the unknown coordinates of a station relative to the stations whose coordinates are known can be derived (cf. Fig. 1.2, p. 3). This procedure has been used on a regular basis with the SECOR system [4.4.1], and it can also be used in satellite laser ranging (SLR) under favorable atmospheric conditions. Similarly, the unknown coordinates of ground stations can be derived from three satellites whose positions are known in the two-way mode, and from four satellites whose positions are known in the one-way mode (e.g. GPS, see Fig. 7.2, p. 211).

In many cases the often high data rate of range observations, (e.g. one observation per second) is not required in the subsequent analysis process. The data are hence compressed to so-called *normal points*. For details of normal point generation see [8.4.2].

4.2.3 Determination of Range Differences (Doppler method)

Fig. 4.7 illustrates the geometrical principle of position determination from range differences between one observer and two pairs of satellite positions. The satellite positions at epochs t_1, t_2 and t_3 are taken as known.

Each of the observed range differences $(\overline{Bt_1} - \overline{Bt_2})$ and $(\overline{Bt_2} - \overline{Bt_3})$ define a hyperbolic surface in three-dimensional space. The observer (e.g. on a ship) is situated at the intersection of the hyperbolic surfaces with Earth's surface. It becomes evident from geometrical considerations that with a single satellite pass only a two-dimensional position can be determined. For a three-dimensional solution several satellite passes are required.
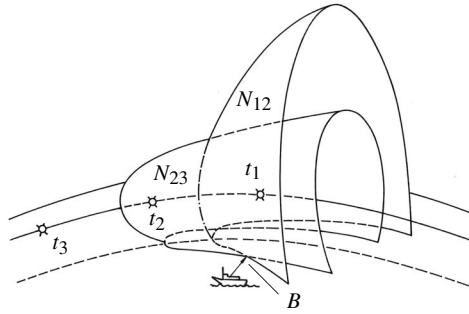
Figure 4.7. Geometrical interpretation of the positioning with range-differences

The range differences are derived from the measurement of the frequency shift caused by the change of range between the observer and the satellite during a given satellite pass [6.1]. The satellite transmits a signal of known frequency f_s which is tracked by a ground receiver. The relative motion ds/dt between the receiver and the transmitter causes the received frequency $f_r(t)$ to vary with time as

$$f_r(t) = f_s \left(1 - \frac{1}{c} \frac{ds}{dt} \right). \tag{4.11}$$

This is the well-known *Doppler effect*. The frequency shift in a given time interval t_j, t_k is observed, and is scaled into a range difference Δr_{ijk} (cf. Fig. 4.6). The related observation equation is

$$|\boldsymbol{r}_k - \boldsymbol{r}_i| - |\boldsymbol{r}_j - \boldsymbol{r}_i| = |\Delta \boldsymbol{r}_{ik}| - |\Delta \boldsymbol{r}_{ij}| = |\Delta \boldsymbol{r}_{ijk}|. \tag{4.12}$$

The observation of the Doppler effect is frequently used in satellite geodesy. The technique is always applicable when a satellite, or a ground-beacon, transmits on a stable frequency. The orbital elements of the very first satellites were determined by observing the Doppler-shift of the satellite signals. The most important application of the Doppler method in geodesy has been with the Navy Navigation Satellite System

(TRANSIT) (cf. [6]). A current space system based on the Doppler technique is *DORIS* [6.7].

The Doppler effect can also be used for the high precision determination of range rates $|\Delta\dot{\boldsymbol{r}}_{jk}|$ between satellites. This method is named *Satellite-to-Satellite Tracking* (SST) [4.2.5], and it can be applied to the mapping of a high resolution Earth gravity field [10.2], [12.2].

4.2.4   Satellite Altimetry

This is a specific form of ranging, where the vertical distance between a satellite and Earth's surface, in particular the ocean surface, is measured. Satellite altimetry was the first operational satellite-borne observation technique in satellite geodesy. The satellite carries a radar-altimeter; no remote station on the ground is required. The altimeter height, a_0, above the sea surface is determined from the two-way travel-time of a radar impulse which is transmitted from the satellite and reflected from the sea surface. The data are communicated in a suitable way from the satellite to the user.

With a knowledge of the satellite orbit, the satellite altitude, h, above the Earth ellipsoid is also known and provides via the simplified relation

$$M = h - a_0 \tag{4.13}$$

the separation M between the mean sea level and the ellipsoid (Fig. 4.8). M approximately equals the geoid height; hence satellite altimetry can be used to determine the geoid over the oceans. A more sophisticated evaluation requires some corrections, in particular to model the separation between the instantaneous sea surface and the geoid, and to model the deviations between the true satellite trajectory and the computed orbit.

GEOS-3 and SEASAT-1 were the first two satellites to carry radar altimeters and have been used extensively for geodetic purposes [9]. Further altimeter satellites were GEOSAT, ERS-1, ERS-2, TOPEX/POSEIDON, GFO, JASON and

Figure 4.8. Simplified principle of satellite altimetry

ENVISAT (cf. [4.3], [9.2]), providing significant contributions to geodesy, geophysics, and oceanography. The launch of additional satellites equipped with radar altimeters is expected in the next few years (cf. Table 9.2). Satellite altimetry with laser systems is under preparation [8.7]. A first laser altimeter will be flown on ICESAT.

4.2.5   Determination of Ranges and Range-Rates (Satellite-to-Satellite Tracking)

This method can be used for mapping the high frequency components of Earth's gravity field [10.2]. The observables are the range and the range rate, i.e. the rel-

ative velocity between two satellites. Different concepts have been proposed and are already realized or at the stage of realization. In the *low-high configuration* a low orbiting satellite (LEO, a few hundreds of kilometers orbital altitude) is combined with a satellite in a high (e.g. MEO or GEO) orbit. The advantage is that a rather long trajectory of the low orbiting satellite, which is particularly affected by high frequency components of the terrestrial gravity field, can be "seen" from the high orbiting satellite (Fig. 4.9). One example is CHAMP with GPS [10.2.2].
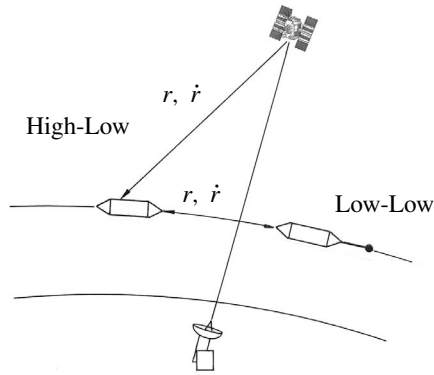
Figure 4.9. Principle of satellite-to-satellite tracking, high-low and low-low configuration

In the *low-low configuration* two satellites occupy the same low-altitude orbit separated by 100 to 300 km. A higher resolution of the gravity field is expected with this technique. The low-low configuration is used with GRACE [10.2.3].

4.2.6 Interferometric Measurements

The basic principle of interferometric observations is shown in Fig. 4.10. A_1 and A_2 are antennas for the signal reception. When the distance to the satellite S is very large compared with the baseline length b, the directions to S from A_1 and A_2 can be considered to be parallel. From geometric relations we obtain

$$d = \overline{A_1 P} = b \cdot \cos \theta. \tag{4.14}$$

If λ is the wavelength of a continuous signal from the satellite, then the phase difference Φ, caused by the range difference d, can be observed at both antennas.

The observed phase difference is uniquely determined only as a fraction of one wavelength; a certain multiple, N, of whole wavelengths has to be added in order to transform the observed phase difference Φ into the range difference d. The basic interferometric observation equation is hence
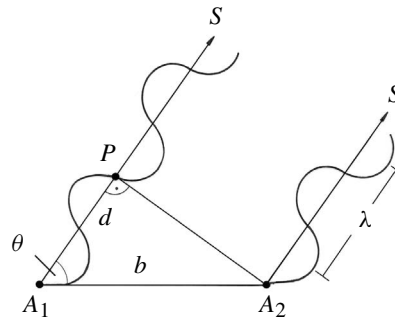
Figure 4.10. Interferometric measurements

$$d = b \cdot \cos \theta = \frac{1}{2\pi} \Phi \lambda + N \lambda. \tag{4.15}$$

The interferometric principle can be realized through observation techniques in very different ways. Equation (4.15) contains different quantities which can be used as derived observables, namely

  − the baseline length b between the two antennas,
  − the residual distance d between the antenna and the satellite, and
  − the angle θ between the antenna baseline and the satellite.

In each case it is necessary to know, or to determine, the integer ambiguity term N. The determination is possible through a particular configuration of the ground antennas, through observations at different frequencies, or through well defined observation strategies.

One example for the determination of directions to satellites by interferometric measurements is with the classical *Minitrack-System* [4.4.2], where the individual antenna elements are connected with cables. The achievable accuracy, however, is not sufficient for modern requirements in satellite geodesy.

With increasing baseline lengths the antennas cannot be connected directly with cables. The phase comparison between the antennas must then be supported by the use of very precise oscillators (atomic frequency standards). This is, for instance, the case with the *Very Long Baseline Interferometry* (VLBI) concept [11.1].

When natural radio sources (e.g. quasars) are observed with the VLBI technique, the range difference d is not determined through methods of phase comparison, but by the correlation of the signals obtained at both antennas. The signal streams are registered together with precise timing signals at both antenna positions, and they are later shifted, one against the other, within a correlator, until the maximum correlation is obtained. The time delay τ corresponds to the signal travel time between P and A_1, and can be scaled to a range difference

$$d = \tau \cdot c. \tag{4.16}$$

When artificial Earth satellites are used in the VLBI technique, it cannot be assumed that the directions from the antennas to the satellites are parallel. Instead, the real geometry has to be introduced by geometric corrections; e.g. the wavefronts must be treated as curved lines [11.1.4].

The interferometric principle has been widely used in the geodetic application of the GPS signals. Both methods described above for the determination of the range differences d are possible:

(a) The signals from the GPS satellites can be recorded at both antenna sites without any a priori knowledge of the signal structure, and later correlated for the determination of the time delay τ. A very large instrumental and computational effort is required, so the method is not really suitable for operational applications. It is, however, used to some extent in modern GPS receiver technology, in order to access the full wavelength of L2 under "Anti-Spoofing" (A-S) conditions [7.2.3].

(b) The phase of the carrier signal at both antenna sites can be compared, and the difference formed. These so-called *single phase differences* can be treated as the

primary observables. The method is now widely used for processing GPS observations [7.3.2.1].

There are different opinions in the literature, on the extent to which method (b) (i.e. the use of phase differences) belongs to the class of interferometric observation techniques. Both terms, namely "interferometric techniques" and "phase differences", were often used as synonyms in the early GPS literature.

4.2.7 Further Observation Techniques

Besides the observables and observation techniques already mentioned other methods were proposed, or are still in use, or are planned for forthcoming satellite missions. In many cases combinations of different observables are employed. One of the concepts proposed and now under development is *satellite gradiometry* [10.3]. A gravity gradiometer measures directly the second derivatives of Earth's gravitational potential. It is very hard to achieve the required resolution with the available instrumentation. The same is true for the application of *accelerometers* in the satellite. A first mission based on this concept will be GOCE, planned for launch in 2006 [10.3].

Earth observation satellites or *remote sensing satellites* carry a large quantity of sensors for the optical and microwave frequency domain. Of particular interest to geodetic applications is the *Interferometric Radar* (InSAR) technique which can be used to detect small deformations of Earth's crust. A short overview of this technique is given in chapter [11.2]. In general, remote sensing techniques are not included in this book. For information see e.g. Leberl (1990); Cracknell, Hayes (1991); Lillesand, Kiefer (2000).

4.3   Satellites Used in Geodesy

4.3.1   Basic Considerations

Most of the satellites which have been used, and still are used, in satellite geodesy were not dedicated to the solution of geodetic problems; their primary goals are various. Typical examples of this group are the navigation satellites of the TRANSIT and of the GPS systems, and remote sensing (Earth observation) satellites carrying a radar altimeter. Examples of satellites which were exclusively, or primarily, launched for geodetic and/or geodynamic purposes are:

PAGEOS (*PA*ssive *GEO*detic *S*atellite)	USA 1966,
STARLETTE, STELLA	France 1975, 1993,
GEOS-1 to 3 (*GEO*detic *S*atellite 1 to 3)	USA 1965, 1968, 1975,
LAGEOS-1, 2(*LA*ser *GEO*dynamic *S*atellite)	USA 1976, 1992,
AJISAI (EGS, *E*xperimental *G*eodetic *S*atellite)	Japan 1986,
GFZ-1 (*GeoF*orschungs Zentrum)	Germany 1986,
CHAMP (*CHA*llenging *M*ini Satellite *P*ayload)	Germany 2000.

This group of dedicated satellites includes some which were used during the first years of the satellite era for the establishment of geodetic datum connections (e.g. SECOR [4.4.1], ANNA-1B (1962)). All satellites which are dedicated to a given observation technique will be treated in detail together with this technique (e.g. [5.1.1], [6.2], [6.7], [7.1.2], [8.2], [9.2], [10]). In this chapter more general aspects are discussed.

The orbital height of a satellite is mainly determined by the purpose of the mission. A satellite used for *gravity field determination* should have a rather low orbit (about 300 to 500 km), and it must carry highly sophisticated instrumentation. A satellite used for *precise position location* should have a rather high and stable orbit, and could be much simpler, from the technical point of view. This is why dedicated missions for the mapping of a fine structured Earth gravity field have only been realized recently, or are in the final stage of realization (cf. [10]). In order to separate gravitational and non-gravitational forces the satellites must be carefully designed. One possibility is to select a favorable *mass/area relation*, which minimizes the forces acting on the satellite surface. In another solution the surface forces are compensated by a thrusting system. This keeps the satellite centered on a "proof mass" which is shielded from the satellite surface forces (cf. DISCOS system [4.3.3.1], Fig. 4.13).

A frequently used distinction for the purposes of subdivision is *passive* and *active* satellites. Passive satellites are exclusively used as targets. They have no "active" electronic elements, and are independent of any power supply. Their lifetime is usually extremely long. Active satellites in most cases carry various subsystems like sensors, transmitters, receivers, computers and have a rather limited lifetime. Table 4.1 gives an overview of the most important satellites that are in use, or have been used, in satellite geodesy.

Table 4.1. Satellites used in geodesy

Passive Satellites		Active Satellites	
ECHO-1	ETALON-1	ANNA-1B	ERS-2
ECHO-2	ETALON-2	GEOS-3	TOPEX/POSEIDON
PAGEOS	GFZ-1	SEASAT-1	GFO (Geosat Follow On)
STARLETTE		NNSS satellites	CHAMP
STELLA		NAVSTAR satellites	JASON
LAGEOS-1		GLONASS satellites	ENVISAT
LAGEOS-2		GEOSAT	GRACE
EGS (AJISAI)		ERS-1	

Another possible subdivision is into:
− Geodetic Satellites,
− Earth Sensing Satellites,
− Positioning Satellites, and
− Experimental Satellites.

Geodetic satellites are mainly high targets like LAGEOS, STARLETTE, STELLA, ETALON, ASIJAI, and GFZ which carry laser retro-reflectors. They are massive spheres designed solely to reflect laser light back to the ranging system. The orbits can be computed very accurately, because the non-gravitational forces are minimized.

Earth sensing satellites like ERS, GFO, TOPEX, JASON, ENVISAT carry instruments designed to sense Earth, in particular to monitor environmental changes. Many of these satellites carry altimeters [9]. The satellites are rather large with irregular shape, hence drag and solar radiation forces are also large and difficult to model. Most are equipped with an orbit determination payload, e.g. PRARE, GPS, and/or DORIS [4.3.3.3]. In addition most satellites carry laser reflectors to facilitate precise orbit determination.

Positioning satellites are equipped with navigation payload. To this class belong the former TRANSIT, GPS, GLONASS, and future GALILEO satellites. Some of the spacecraft carry laser reflectors (e.g. GPS-35, -36, and all GLONASS satellites). The satellites are arranged in constellations of up to 24 and more to provide global or regional coverage.

Experimental satellites support missions with experimental character. They are used in the development of various other kinds of satellites, to test their performance in real space operations. A large number of experimental satellites have been launched for communication technology. Experimental satellites of interest to geodesy are mostly irregularly shaped and fly in low orbits. Precise orbit determination (POD) is supported by laser cube corner reflectors and/or a navigation payload like GPS. Examples include TiPs (Tether Physics and Survivability), and Gravity Probe B.

4.3.2   Some Selected Satellites

In this section some satellites and subsystems are described that are in use or have been used for different observation techniques in satellite geodesy, and that are not discussed later in detail in the context of a particular method.

GEOS-3
The third satellite of the GEOS series was launched by NASA on April 9, 1975. The initial orbital elements and some physical parameters are:

period	102 minutes,
apogee height	844 km,
perigee height	837 km,
inclination	115°,
weight	340 kg,
diameter	132 cm, and
length	81 cm.

The antennas are oriented toward Earth using a 19.5 m gravity gradient boom and a 45 kg boom end mass. Fig. 4.11 shows the configuration of the spacecraft and its main elements. The experiment package consists of the following:
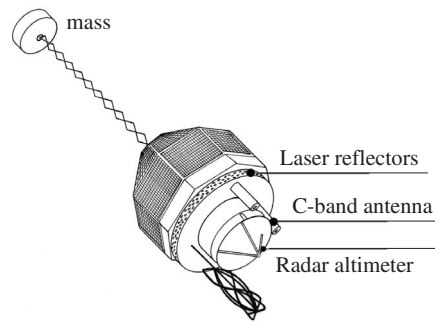
Figure 4.11. GEOS-3 spacecraft

- *Radar Altimeter* for the satellite-to-ocean surface height measurements [9.2]; 5.7 GHz; precision ±60 cm,
- *C-Band Transponder*, providing for range, range-rate and angle measurements in conjunction with appropriately equipped ground stations [4.4.2],
- *S-Band Transponder*, 2.1 and 2.2 GHz, for satellite-to-satellite tracking experiments [10.2],
- *Laser Retroreflector Array* with 264 quartz cube corner reflectors; design-accuracy ±10 cm [8.2],
- *Doppler System*, dual frequency (162 and 324 MHz), providing the determination of positions and position changes from ground stations [6].

GEOS-3, because of the well-equipped experiment package, became the geodetic satellite "par excellence". Many problems from science and practice could be solved with GEOS-3 data (cf. [12.2]). The most important subsystem from the geodetic point of view was the radar altimeter. It was in operation for more than three years until the end of 1978, without any considerable interruptions. The laser-reflector array can still be used.

SEASAT-1
The oceanographic satellite SEASAT-1 was launched on June 26, 1978, with an orbit similar to that of GEOS-3, namely a period of 109 minutes, an altitude of 800 km, and an inclination of 108° (cf. [9.2]).

The spacecraft carried several sensors for use in oceanography, and a radar altimeter with a ±10 cm resolution, which exceeded by far the design-accuracy of the GEOS-3 altimeter. Due to a break-down in the power system, SEASAT-1 only delivered altimeter data for five months. However, because of the much higher data rate, the size of the SEASAT-1 data set is similar to the GEOS-3 data set. The results of the SEASAT mission have contributed considerably to the progress of geodesy [9], [12.2].

ERS-1, ERS-2
The designation *ERS-1* stands for the *First European Space Agency (ESA) Remote Sensing Satellite*. It was launched on July 17, 1991. ERS-1 flies in a sun-synchronous

orbit at an altitude of about 800 km and an inclination of 98.5 degrees. The mission had the following main objectives (cf. [9.2]):

- − monitoring of the global oceans,
- − observing polar and sea ice,
- − monitoring regionally the land surface, and
- − supporting geodetic research.

From the geodetic point of view the two on-board systems of greatest interest were the radar altimeter (RA) and PRARE. The *radar altimeter* is a single frequency Ku-band (2 cm waves) instrument of the SEASAT type with an anticipated height-resolution of 0.1 m over sea and 0.4 m over ice [9.2]. PRARE [4.3.3.3] should have been used for precise orbit determination at the 10 cm accuracy level. Unfortunately the PRARE system could not be activated after launch. ERS-1 is also equipped with laser retro-reflectors providing the primary tracking of the spacecraft.

The remote sensing objectives of the mission were covered by a large variety of instruments, such as an *Active Microwave Instrument* (AMI) including a *Synthetic Aperture Radar* (SAR) and an *Along Track Scanning Radiometer* (ATSR), providing information on sea state, winds and waves. For a deeper study of techniques and objectives in remote sensing see the related literature (e.g. Maul (1985), Cracknell, Hayes (1991), Lillesand, Kiefer (2000)). For the use of altimeter data see [9.5].

A very powerful tool for geodetic deformation studies developed with the interferometric use of the SAR antenna, the *Interferometric SAR* (InSAR), see [11.2].

ERS-2 was the follow-on mission to ERS-1. It was launched on April 21, 1995 into an orbit similar to ERS-1. It carries similar instruments to ERS-1, as well as the *Global Ozone Monitoring Experiment* (GOME). Precision orbits were determined with PRARE and SLR. For a period of time, both satellites flew in the combined *tandem mission* (see [9.2]).

ASIJAI (EGS)

The Japanese *Experimental Geodetic Satellite* (EGS) was launched on August 12, 1986. The unofficial name is *AJISAI* (water-snake). The spacecraft is well suited for laser ranging and for photographic camera observations. It is polyhedron-shaped with an effective diameter of 2.15 m. It carries 318 mirror elements and 120 reflector assemblies for laser light (Fig. 4.12). The total weight amounts to 685 kg. The nearly circular orbit has an inclination of $i = 50°$ and a period of 115.7 minutes.

The orbital height is about 1500 km. The satellite initially rotated around its own axis at 40 revolutions per minute (rpm). The sunlight was thus reflected in such a way that an observer on the
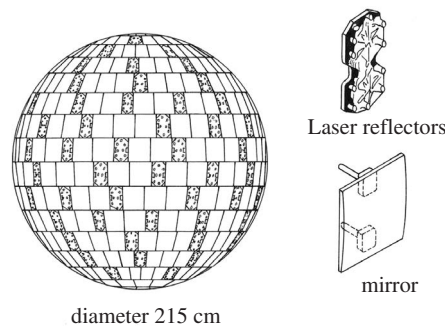
Figure 4.12. *AJISAI* (EGS), Japan

dark side of Earth could see about two short flashes per second with a duration of 5 ms and an apparent star magnitude of $1.^m5$ to $3.^m5$. These flashes could be photographed together with the background stars. In the meanwhile the rotation rate slowed down by about 0.00145 rpm/day and arrived at 33.9 rpm by March 1998 (Otsubo et al., 1999). The 120 reflector groups contain in total 1436 retroreflectors for laser light. The design accuracy of these reflectors corresponds to a range resolution of 1 to 2 cm (cf. [8.2]).

TDRSS

The *Tracking and Data Relay Satellite System* (TDRSS) provides tracking and data communication between low Earth orbiting (LEO) spacecrafts and ground-based control and data processing facilities. The space segment consists of seven Tracking and Data Relay Satellites (TDRS) located in geosynchronous orbits. The constellation provides global coverage. The system is capable of transmitting to and receiving from the spacecraft over 100% of their orbit. The ground segment is located near Las Cruces, New Mexico. The system has worked successfully since 1983 and supports a large number of scientific missions. Among these are the

 — Hubble Space Telescope,
 — Space Shuttle,
 — Landsat,
 — Ocean Topography Experiment (TOPEX),
 — Earth Observing System (EOS),
 — Space VLBI,
 — International Space Station (ISS), and
 — JASON.

4.3.3   Satellite Subsystems

4.3.3.1   Drag Free Systems

A drag-free satellite is built by isolating a proof mass within the satellite completely from the surrounding environmental influences. The spacecraft is equipped with a so-called *Disturbance Compensation System* (DISCOS). In its basic form a massive spherical *proof mass* is shielded from the forces on the satellite surface within a hollow ball. The ball is attached to the satellite and experiences all surface forces, such as drag and radiation pressure. The proof mass is only affected by gravitational forces. Changes in the relative position of the proof mass and hollow ball are measured, and allow separation of the gravitational forces from surface forces (Fig. 4.13). Two objectives can be achieved:

(a) Through a closed loop thrusting system the satellite can be kept centered on the proof mass and thus in a much more stable orbit. This is of particular importance for low-orbiting navigation satellites because, otherwise, the surface forces create large differences between the predicted (broadcast) orbit and the true satellite position. DISCOS was tested in the experimental navigation satellite TRIAD (launched 1972) and was later installed on the NNSS satellites of the new NOVA type [6.2], (Eisner, et al., 1982).
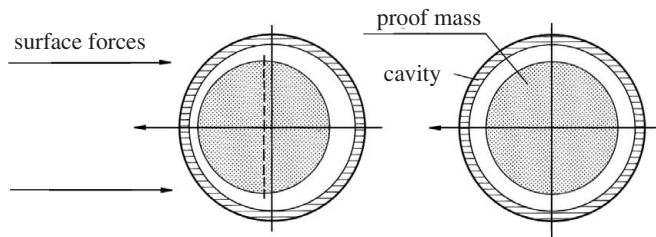
Figure 4.13. Displacement compensation system (DISCOS)

(b) The effects of the gravity field on satellite orbits can be analyzed in more detail, and the gravitational signal can be separated from the non-gravitational signal. This is of particular importance for low-orbiting satellites that are used for mapping a high resolution gravity field. Accordingly drag compensation systems form part of gravity field missions [10]. They are also foreseen for space probes testing the theory of relativity like GRAVITY PROBE B and STEP, or for gravitational wave astronomy projects. The SQUID (Superconducting Quantum Interference Device) is a very high resolution sensor to measure the changes in relative position. It was developed for the GRAVITY PROBE B and STEP missions (Karslioglu, 2000). Another field of study using DISCOS is aeronomy.

An alternative solution for counteracting drag is by the use of *accelerometers*. The position of a free-floating proof mass in a capacitive arrangement is sensed, and a force is exerted upon it to bring it back to a neutral position in its housing. The acceleration readout is obtained from the magnitude of the force required to maintain the proof mass at its neutral position. The measured forces are used to activate a system of micro-thrusters.

4.3.3.2   Attitude Control

For many purposes it is necessary to control the satellite with respect to translation and rotation. This is realized through an *attitude control system*. The *attitude* of a spacecraft may be defined as its rotational orientation in space with respect to an inertial frame (Quine, 1996). Some examples of control tasks for which this system is responsible are, as follows (Sidi, 1997):

— for orbital maneuvers in which the attitude of the space vehicle must be held in the desired Δv direction,

— a spin-stabilized satellite may be designed to point with its spin axis at a particular space-fixed direction,

— an Earth-observation satellite may be designed to track some defined targets on the surface,

— an astronomical satellite observing the sky must point its optical payload toward particular objects on the celestial sphere.

A basic distinction in attitude control concepts is between *passive* and *active* attitude control. Passive control requires less complicated and less expensive hardware. One example is *gravity gradient attitude control* realized through a boom end mass (see e.g. GEOS-3 [4.3.2], Fig. 4.11). Components of an active attitude control system may be:

– accelerometers,
– star sensors,
– gyros, and
– GPS arrays.

Accelerometers are most suitable for sensing translations. CCD star sensors are particularly suitable for monitoring the orientation of spacecrafts (Quine, 1996). Star positions are referred to inertial space, and they can be considered as point-like sources. With these characteristics star sensors allow attitude determination with accuracies in the arc-second range [5.3.1].

The use of mechanical gyros is sub-optimal because the fast rotating rotor elements produce high frequency vibrations that disturb the accelerometer readout (Karslioglu, 2000). The development of laser gyros may improve the situation. The use of GPS receivers or GPS antenna arrays gains importance with advanced receiver technology [7.6.2.9].

A detailed study of system behavior for attitude control is of high importance for sensor modeling with the new gravity field satellite missions, cf. [10]. An excellent reference for details on the subject is Sidi (1997).

4.3.3.3   Navigation Payload, PRARE

Three main systems are being used for precise orbit determination as active payload onboard spacecraft:

– GPS,
– DORIS, and
– PRARE.

GPS is flown on an increasing number of Low Earth Orbiters. Its use as navigation payload is discussed in [3.3.2.3] and [7.6.2.9]. DORIS is based on the Doppler technique; it forms part of several missions, in particular for Earth observation satellites. Several future missions are planned. DORIS is therefore treated within the chapter on Doppler techniques [6.7]. PRARE is a discontinued technology. The system has been flown on several platforms, but most probably will not be included in future space missions. Nonetheless from the conceptual and technological point of view PRARE is a very powerful and interesting system. Its main characteristics and features are therefore explained in more detail.

PRARE stands for *P*recise *R*ange *A*nd *R*ange-rate *E*quipment. The original concept was developed at the University of Stuttgart and the German Geodetic Research Institute (DGFI), Munich (Reigber, Hartl, 1989, 1990). PRARE was flown for the first time on the ERS-1 satellite, but could not be activated. Successful missions were with the Russian METEOR-3/7 satellite, and with ERS-2.

PRARE is an autonomous, spaceborne, two-way, dual-frequency microwave tracking system, consisting of three components:

− the *space segment*, a small unit containing all necessary instruments, including telemetry and data storage, but less the power supply, in a box measuring 400 mm x 200 mm x 100 mm, with a mass of 17 kg,

− the *control segment*, for system control and calibration, time control, communication with the space segment, preprocessing, distributing, and archiving the data,

− the *ground segment*, consisting of small, transportable, automated ground stations.

The observation principle of PRARE is demonstrated in Fig. 4.14. Two pseudo random noise (PRN) coded microwave signals are transmitted simultaneously from the space segment to the ground station; one signal is in the S-band (2.2 GHz), the other in the X-band (8.5 GHz). At the ground station the time delay in the reception of both signals is measured with an accuracy better than 1 ns, and transmitted to the space segment for a calculation of ionospheric correction. At the same time the atmospheric parameters at the ground station are transmitted for tropospheric modeling purposes.

The ranging signal in the X-band is transposed in the ground station to 7.2 GHz and retransmitted to the space

Figure 4.14. PRARE concept

segment. The two-way signal travel time is determined in the space segment via a correlation process and provides a measure for the two-way slant range between the satellite and the ground station. In addition to this, the Doppler-shifted carrier frequency is counted in the space segment, and is used to derive the relative velocity between the spacecraft and the ground station. Up to four ground stations can be tracked simultaneously. All data are collected and stored in the onboard memory, and can be transmitted to the ground control station whenever contact is made.

The overall noise of the preprocessed full range data (1 range measurement per second) and the 30 seconds Doppler count integration is 2.5 to 6.5 cm for range data (depending on multipath effects at the ERS-2 solar panels), and 0.1 mm/s for the range-rate (Doppler) data. The normal point noise is less than 1 cm and 0.015 mm/s, respectively (Flechtner1997). The main reason for the good range-rate measurement precision is the high carrier frequency (about 8 GHz).

A PRARE ground station consists of 3 separate units, easily carried by hand:

- an antenna unit with a 60 cm parabolic dish, based on an azimuthal mounting (the steering commands for the antenna mount are derived from the data signal, broadcast from the satellite),
- an electronic unit with RF-modules, station processor and power supply, and
- a monitor and computer as a user interface.

Up to 29 ground stations can be operated in a global network. Fig. 4.15 shows a block diagram of a PRARE ground station.
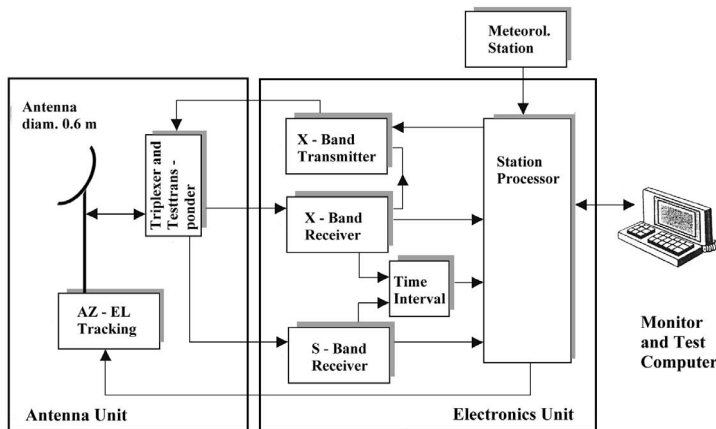
Figure 4.15. Block diagram of a PRARE ground station (Dornier)

The primary objective of the PRARE system is to provide precise orbit determination for satellite missions. Based on a global network of ground stations, a radial orbit accuracy of better than 10 cm has been achieved. Further contributions support, for example (Reigber, Hartl, 1990; Bedrich, 1998; Flechtner, 2000):

- absolute and relative position determination of the ground stations,
- studies of ice mass balance (e.g. Antarctica),
- determination of sea surface topography,
- refined modeling of Earth's gravity field,
- studies of the ionosphere (total electron content), and
- precise time transfer and clock synchronisation.

4.3.4 Planned Satellites and Missions

The time span between the first planning of a satellite mission and its final realization usually amounts to more than 10 years. During this period concepts are developed and thoroughly studied in several phases. They are re-evaluated, adapted to new

developments, and finally approved or rejected. Very few of the many proposals that are described and discussed in the literature and dedicated studies finally come to a realization.

Some planned missions of interest to geodesy are indicated in the following. More details are given in the respective chapters on particular observation techniques.

Navigation:

- GALILEO, European Satellite Navigation System, first launches planned for 2004, completely deployed by 2008 [7.7.3],
- GPS IIF, New generation GPS, launches begin after 2005, [7.1.2], [7.1.7],
- GPS III, Follow-up generation GPS, under design [7.1.7].

Altimetry:

- ICESAT, Ice, Cloud, and Land Elevation Satellite, carrying the *Geoscience Laser Altimeter System* (GLAS), launched on January 12, 2003 [8.7],
- CRYOSAT, part of ESA's Living Planet Program in the framework of the Earth Explorer Opportunity Mission; radar altimetry mission dedicated to the observation of polar regions; anticipated launch in 2004 or 2005 [9.2].

Gravity Field:

- GOCE, first satellite with an onboard gradiometer, launch planned for 2006, [10.3].

Astrometry:

- DIVA, astrometric satellite, to bridge the gap between HIPPARCOS and FAME, or GAIA, 15 million star positions; highly elliptical geosynchronous orbit (Pe: 500 km, Ap: 71 000 km); launch uncertain [5.3],
- FAME, astrometric satellite, positions, proper motions and parallaxes for 40 million stars with 50 microarcseconds, launch uncertain [5.3],
- GAIA, ESA's space astronomy mission; position and motion of more than 1 billion stars of our galaxy; 1 microarcsecond/year; launch around 2010–2012.

Remote Sensing:

- ADEOS-2, (Advanced Earth Observing Satellite), also named "Midori II"; research on global climate changes (mainly water changes and ozone layer); launched on December 14, 2002.
- RapidEYE, constellation of four mini-satellites carrying a CCD-based imaging system; resolution 6.5 meters; anticipated launches from 2004 onward.
- TerraSAR, two-satellite system in polar, sun-synchronous orbit with 12 minutes spacing; dual frequency SAR (L-band, X-band); launches after 2005 [11.2].

Experimental Satellites:

- GRAVITY PROBE B, test of relativistic theories; precise orbit determination with GPS and SLR; launch planned for July 2003,
- STEP, (Satellite Test of the Equivalence Principle); the experiment plans to fly several pairs of masses on a drag-free satellite in low Earth orbit; launch planned after 2006.

4.4   Some Early Observation Techniques (Classical Methods)

The satellite tracking methods of the first years after 1957 originated from before the launch of the first artificial satellites, or were based on existing techniques. This is true for the methods of *stellar triangulation*, which follow from astronomy with the Moon as a target, and for the visual, photographic, and electronic tracking of rockets. Only the satellite laser ranging technique can be regarded as an original development of the early satellite era (Henriksen, 1977).

After the launch of SPUTNIK-1 on October 4, 1957, the satellite signals, which were continuously transmitted on frequencies of 20 MHz and 40 MHz, could be received all over the world with existing antennas. The *Doppler shift* [6.1] of these signals was measured (mostly by observing the behavior of Lissajou figures with oscilloscopes), and could be used for tracking purposes. More precise radio-tracking systems (e.g. *Minitrack* [4.4.2]) were already under development for the anticipated national American space program, and were used after 1958 for the observation of a large number of satellites. For monitoring and tracking of *passive*, non-transmitting, satellites, powerful cameras were used, the so-called *tracking cameras*. The *Smithsonian Astrophysical Observatory* (SAO) initiated, within the *International Geophysical Year* (1957–1958), the development of the *Baker–Nunn Camera*, which could be used for the photography of small, sun-illuminated satellites (Pearlman, 1983).

The primary motivation during the first years of satellite observation was focussed on the development of improved models for the orbital motion of near Earth satellites, and to the determination of substantial geometrical and physical Earth models. Many observations of a large number of satellites were included in the determination of the early *Earth models* [12.2]. These data from the "classical period" of satellite geodesy still contribute to current Earth models. In this respect, the classical observation methods retain their importance, and are briefly discussed in this book. The present practical importance of the classical observation techniques in their original form, however, is very small.

The photographic determination of directions led to a remarkable early result in satellite geodesy, namely the establishment of the first worldwide geometric network [5.1.5]. Directional methods are still of high significance in satellite geodesy and found a remarkable new perspective with CCD technology. This is why the method is described in a particular chapter [5].

The TRANSIT technology also belongs to the classical observation techniques. The underlying Doppler method, however, is still an important observation tool in satellite geodesy, and has a modern realization in the DORIS concept. Furthermore, the methodology developed along with TRANSIT has considerably influenced the geodetic use of GPS. This is why the Doppler method is treated in a particular chapter [6].

4.4.1 Electronic Ranging SECOR

The development of electronic ranging techniques began rather early. To implement two-way ranging capability, the satellite had to be equipped with dedicated receivers and transmitters, so-called *transponders*. The SECOR technique was developed particularly for geodetic application. SECOR means *SE*quential *CO*llation of *R*anges. One of the first SECOR transponders was flown on ANNA-1B (1962). A total of sixteen satellites with SECOR equipment were launched into near polar orbits of 1000 to 4000 km altitude between 1964 and 1970 (NGSP, 1977, Vol.1, 221), among them GEOS-1 and GEOS-2.

The basic idea of SECOR is that four ground stations and one satellite form a group, the so-called *Quad*. Three of the four ground stations are considered to be at known positions, the fourth station is the new point N, to be located. This is the *trilateration* principle, a purely geometric method of coordinate determination which is illustrated in Fig. 1.2, p. 3. At least three well-selected satellite positions are determined through simultaneous ranging from the three "known" ground stations. Based on the three determined satellite positions the coordinates of the unknown station N are derived by spatial resection. Further "quads" of groundstations can be added to form larger networks, up to a worldwide girdle of stations. In addition to the purely *geometric simultaneous method*, the *orbital method* of SECOR was used (see Fig. 1.3). A short portion of the orbit (*short arc*) was determined from at least three known ground stations, and was then extrapolated for the determination of unknown ground stations. In practice, combined evaluations have also been used.

Note that the basic principles which have been developed for the technique of point positioning with SECOR are also applicable to modern ranging methods in satellite geodesy.

SECOR used a phase comparison technique for the determination of ranges. Modulated signals on a carrier frequency of 420.9 MHz were transmitted from the ground station to the satellite. They were transmitted back to the ground stations via satellite–borne transponders on two different frequencies (449 MHz and 224.5 MHz, for estimation of an ionospheric correction). The ranging signal had a frequency of 585 MHz, corresponding to a resolution of 25 cm. Three additional modulation frequencies were used for solving the ambiguities. One of the four stations was designated the master station, and synchronized all measurements. The interrogation period for all four participating stations was 50 ms.

The main purpose of the SECOR system was the geodetic connection of isolated local reference frames (datum connection), in particular between North America, Australia, Japan, and several islands in the Pacific (Rutscheid, 1972). To achieve this objective an equatorial network, consisting of 37 stations, was observed between 1964 and 1966. The pure SECOR solution showed a rather weak geometry and was affected by large systematic errors. The standard deviation of a single range measurement (internal accuracy) was about ± 3 m; however, the systematic differences, when compared with other solutions (BC4, Doppler), were up to 50 m (external accuracy). A combined final adjustment, including short arc techniques and orientation control from

BC4 azimuths, resulted in a position accuracy of $\pm$ 10 to 15 m (NGSP, 1977, Vol.1, 203).

4.4.2   Other Early Observation Techniques

During the first years of the satellite era some other electronic observation techniques were used for orbit control and orbit determination. Some of the results were used for geodetic purposes. The systems and techniques in question were:

GRARR   (Goddard Range and Range Rate) for the determination of ranges,

MINITRACK   for the interferometric determination of directions, and

C-BAND RADAR   for the simultaneous determination of directions and ranges.

Many of the satellites, that were launched between 1960 and 1970 carry the appropriate equipment for these techniques.

GRARR is a two-way ranging technique. A phase-modulated carrier frequency (2.27 GHz) is transmitted from the ground to the satellite, where it is shifted in frequency by a transponder, and sent back to the ground station on 1.70 GHz. The distances are determined from phase measurements with up to 8 modulation frequencies ($\lambda \approx 0.6$ to 37 500 km); the range rate is derived from the Doppler shift of the carrier. The precision of the ranging signals is about ± 10 m, and that of the range rate signals about $\pm$ 3 cm/s. The system was operated by the *Goddard Space Flight Center* (NASA) and was successfully used on GEOS-2 (NGSP, 1977, Vol.1, 433). The results have been included in the Goddard Earth Models (GEM) [11.2]. Orbital arcs up to 7 days duration were observed .

PRIME MINITRACK is an interferometric one-way technique; it was used by NASA for the orbit determination of many satellites. A beacon on the satellite transmits a continuous carrier signal at 136 MHz which is received at a pair of crosswise arranged antennas. Interferometric phase differences are measured [4.2.6] and transformed into direction information. Because of the rather long wavelength ($\lambda \approx 2.2$ m), compared with the extension of the interferometer (≈ 125 m), the angular resolution is only about $\pm 20''$. Minitrack observations of single orbital arcs have contributed to the GEM computations [12.2].

C-BAND RADAR uses the 5 to 6 GHz domain ($\widehat{=}$ 5 cm). It is a ground-based two-way technique. The radar signals are reflected from the satellite surfaces without using transponders. This is why the system is particularly suitable for the orbital control of satellites, rockets, and parts thereof. The ground station is rather large and uses an 8.8 m parabolic dish. The range is derived from pulse travel times. The orientation can be read at the two-axis mounting. The ranging accuracy is about $\pm$ 2 to 5 m and the angular accuracy about $\pm 20''$. With modified equipment coherent phases and phase changes can also be observed. When a transponder is used (as is the case on GEOS-3) the range and accuracy can be increased considerably. The C-band radar was intensively used for the determination of GEOS-3 orbits.

5 Optical Methods for the Determination of Directions

The determination of directions from the ground to satellites based on optical observations, is one of the early methods of satellite geodesy that led to remarkable results. In addition, optical tracking of satellites is of fundamental importance because it is the only technique in satellite geodesy which directly establishes access to the inertial reference frame (cf. [2.1.2.1]). All other methods (like GPS [7] or SLR [8]) only indirectly provide a link to the frame through the equation of motion.

Unfortunately the optical era in satellite geodesy came to a sudden end with the development of satellite laser ranging (SLR) and the use of the Doppler technique for positioning soon after about 1975. The reason is well understood. A directional accuracy of $\pm 0.''1$ corresponds to 3 m for a satellite at 6000 km (e.g. LAGEOS). The optical method was not competitive compared with the cm accuracy available with laser ranging

Recent progress made in the development of *Charge Coupled Device* (CCD) technology has led to a revival of optical satellite observations. This development is promising and interesting because directional observations, besides the direct link to the inertial frame, still provide important contributions to satellite geodesy and satellite tracking, e.g. (Hugentobler, 1998):

- optical observations are the most reliable and accurate source of information for small, passive, and remote objects, like inactive satellites or space debris, in particular in the geostationary belt,
- geostationary or GPS satellites show characteristic resonances with Earth's rotation which can be accurately determined with optical observations,
- other than VLBI or SLR, only optical observations from single stations can provide important information, and
- optical observations are an independent tool to control and calibrate other observation techniques.

The classical photographic determination of directions contributed significantly to the early development of satellite geodesy. The basic methodological foundations of this method, in particular the technique of *plate reduction*, are still of value and can also be applied to the analysis of CCD images. This is why a review of the photographic method is given first.

5.1 Photographic Determination of Directions

The principle of the method is based on taking photographs of illuminated, or flashing, satellites together with the star background [4.2.1].

Satellite directions are obtained when the individual images of the chopped satellite trail are interpolated into the framework of the background stars that serves as a field

of fiducial points. The necessary tools are: appropriate satellites, appropriate tracking cameras, precise star positions and appropriate methods of plate measurement and plate reduction. Furthermore, the observation epochs must be related with sufficient accuracy to a common time scale (e.g. UTC), in order to satisfy the geometric condition of simultaneity between stations.

5.1.1   Satellites used for Camera Observations

Satellites are required that are illuminated by the sunlight, illuminated by a laser from the ground station, or that are capable of emitting a sequence of self-generated flashes of light. Observations can only be made at night when the ground station is located within Earth's shadow. Targets must be bright enough to create images in the emulsion of the photographic plate or film. Initially passive *balloon satellites* were used, for example the early experimental communication satellites

ECHO-1 (1960–1968), $d = 30$ m, $h \approx 1600$ km, magnitude $-1.^{m}0$, and
ECHO-2 (1964–1969), $d = 40$ m, $h \approx 1200$ km, magnitude $-1.^{m}5$.

For observations of the *BC4 World Network* of the US Coast and Geodetic Survey (now National Geodetic Survey (NGS)) [5.1.5] a dedicated *PA*ssive *GEO*detic Balloon *S*atellite was launched on June 24, 1966:

PAGEOS (1966–1972), $d = 30$ m, $h \approx 2800 - 5600$ km, $i = 87°$, $e = 0.1356$, magnitude $+1.^{m}6$ (at 3000 km).

Also minor balloon satellites were used, such as

EXPLORER-19 (1963), $d = 3.0$ m, $h \approx 1300$ km, and
EXPLORER-39 (1968), $d = 3.6$ m, $h \approx 700 - 2500$ km.

Active satellites were able to emit a series of 6 to 8 flashes of about 1 millisecond length:

ANNA-1B (1962), $h \approx 1100$ km, $i = 51°$,
GEOS-1 (1965), $h \approx 1100 - 2300$ km, $i = 29.°5$, and
GEOS-2 (1968), $h \approx 1100 - 1600$ km, $i = 106°$.

After around 1975 very few camera observations were used in satellite geodesy; in most cases images of laser echos from satellites equipped with retro-reflectors were taken. Well defined targets were:

STARLETTE (1975), $d = 24$ cm, $h \approx 810 - 1100$ km, $i = 49.°8$, and
LAGEOS (1976), $d = 60$ cm, $h \approx 5900$ km, $i = 110°$.

The Japanese satellite AJISAI (EGS) [4.3.2]

AJISAI (1986), $d = 2.15$ m, $h \approx 1500$ km, $i = 50°$

is particularly suitable for camera observations because it generates flashes, by sunlight reflected off its rotating spherical polyhedron. In addition, laser ranging can be performed simultaneously, because the satellite carries retro-reflectors. With the launch of AJISAI, a certain "revival" of camera observations came about, in particular in Japan to establish geodetic datum connections between islands.

5.1.2 Satellite Cameras

Dedicated cameras are required to photograph artificial satellites. A special shutter is needed to divide the trails of stars and satellites into small separated images which can be measured on a comparator. The optical elements must be of high quality with respect to geometry (distortion) and luminous intensity. The following concepts have been realized (cf. Fig. 5.1):
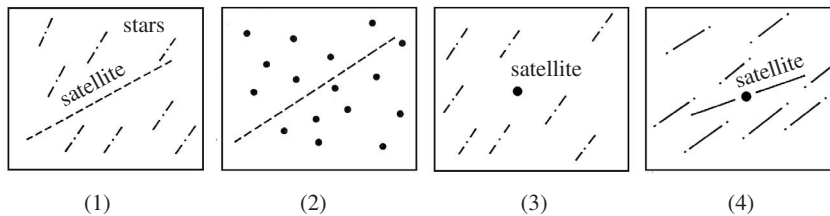
Figure 5.1. Concepts of satellite cameras

(1) *Azimuthal mounting*
The camera is fixed with respect to the ground and follows neither the stars nor the satellites. The stars (like the satellites) form trails on the photographic plate and have to be "chopped" with a shutter forming small, individual images. One characteristic example is the *BC4 Camera* (Fig. 5.2). This camera was developed by the manufacturer Wild (Heerbrugg, Switzerland), based on existing photogrammetric cameras. The objective lenses *Astrotar* ($f = 305$ mm), and *Cosmotar* ($f = 450$ mm) were especially developed; the plate format (18 cm $\times$ 18 cm) corresponds to a field of $24° \times 24°$ for the Cosmotar. The directional accuracy of a single camera observation is $\pm 0.''2$ to $0.''5$. The BC4 cameras have been used worldwide, in particular for the PAGEOS observations within the US NGS geometric satellite world network (Schmid, 1974, 1977; Böhler, 1972) [5.1.5].

(2) *Equatorial mounting*
This mounting permits rotation of the instrument about one axis, parallel to Earth's axis of rotation and thus allows for a compensation of the diurnal rotation. The star images are formed as points, but the satellite track has to be chopped into particular images by a rotating shutter (as in case (1)). One characteristic example is the *Ballistic Camera BMK*, produced by Carl Zeiss Company Oberkochen, Germany, with focal lengths of 45 cm or 75 cm. The high quality objective lens *Astro-Topar*, especially designed for this camera, has a nominal distortion of less than 5 μm. The plate format 18 cm $\times$ 18 cm corresponds to a field angle $22° \times 22°$ for the BMK 45. The directional accuracy of a single camera observation was found to be $\pm 0.''1$ to $0.''2$ (Seeber, 1972).

(3) *Three-axis mounting*
The camera can follow satellites (*tracking camera*). Also fainter satellites can be tracked. The camera is, however, not suitable for active satellites, laser echos and AJISAI echos. The characteristic example is the *Baker–Nunn Camera* (Fig. 5.2). This

camera was primarily operated by the Smithsonian Astrophysical Observatory (SAO) in its equatorial network (12 stations). The optical system was designed by *J.G. Baker*; the mounting and mechanical system by *J. Nunn*. The mirror optics combines a focal length of 50 cm with a relative aperture of 1:1. The field of view is $5° \times 30°$. The direction accuracy is about $\pm 2''$. Many of the observations have been used in the early "SAO Standard Earth" models [12.2]. For further reading see e.g. Pearlman (1977), Pearlman (1983).

(4) *Combined solution*

The camera follows either the satellites or the stars. Faint satellites can be identified within a field of faint stars. The characteristic example is the *Satellite Observation Instrument* SBG, developed by the Carl Zeiss Company Jena (former German Democratic Republic). The camera can track either the stars or the satellites. The 4-axis mount supports a Schmidt reflector. The focal length is 76 cm and the field of view $11.3°$. The photographic plate moves in the focal plane either with star or satellite velocity. Satellites up to a magnitude of 10^m can be recorded with 1s exposure time. The direction accuracy is about $\pm 1''$ to $2''$.

Note that many of these cameras are still available around the world, and can be used with CCD technology [5.2].
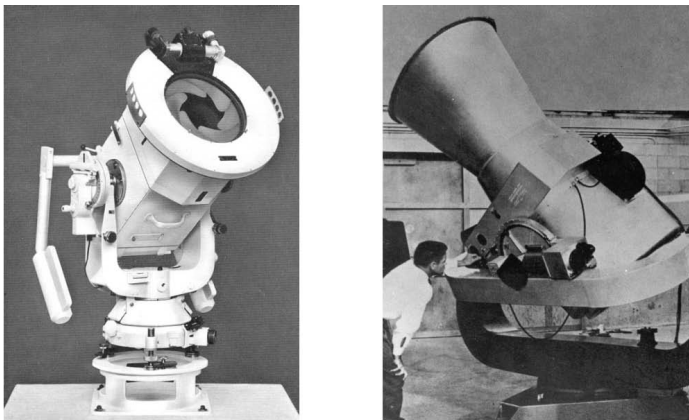
Figure 5.2. Satellite cameras; BC-4 (left), Baker–Nunn (right)

5.1.3   Observation and Plate Reduction

The organization, realization, and reduction of photographic satellite observations with cameras is extremely time consuming, and requires much effort. Only some basic considerations, and the fundamental steps are explained in this chapter. For a full treatment of this subject see e.g Schmid (1977), or Seeber (1972). The basic principles and algorithms, in particular of the plate reduction process, are also valid for the reduction of CCD images [5.2].

Camera observation

An appropriate observation epoch has to be determined, based on predictions of the satellite orbits, and on a computation of the satellite visibility for all participating stations. At least two stations have to observe simultaneously in order to contribute to the geometric solution [1.2] in satellite geodesy. The amount of valuable data increases with more than two participating stations. When passive, sunlight reflecting, satellites are used (like AJISAI), the satellite must be outside of the Earth's shadow (Fig. 5.3) and the Sun must be more than 18° below the horizon of the observation station (*astronomical darkness*). The related areas of satellite intervisibility are a function of the geographical locations of the observation stations, and of the respective orbital height.
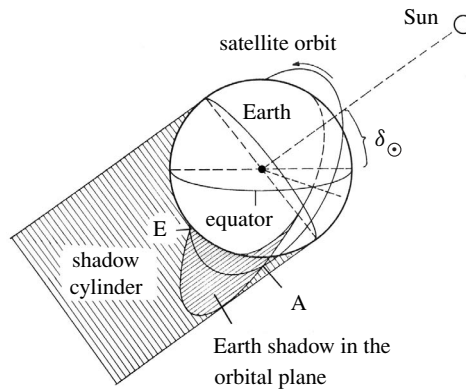
Figure 5.3.   Visibility conditions for passive satellites

Fig. 5.4 shows for three stations B_1, B_2, B_3 the visibility areas (zenith angles $z < 60°$), together with the sub-satellite track (cf. Fig. 3.25, p. 127). It becomes evident that only a small portion of the satellite orbit (S_A to S_E) is simultaneously visible from all three stations. The visibility conditions become more favorable with increasing orbital height and decreasing station separation. The best intersection conditions are present, from geometrical considerations, when the distance between the ground stations equals approximately the orbital height of the satellite used [5.1.4]. A successful observation from at least two stations is called an *event*.

The above-mentioned conditions explain the difficulties that arise with the realization of an observation project, in particular because fine weather conditions must be present simultaneously at all participating stations. These are the reasons for the long duration (several years) of all large projects which have been conducted in the past.
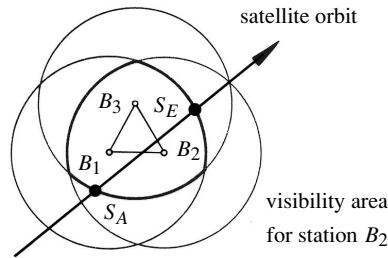
Figure 5.4. Visibility areas for three ground stations

Coordinate measurements and corrections

The exposed and developed plates or films (photograms) are measured on a comparator. The results are the rectangular coordinates x, y of the stars and of the satellite images, defined in the plane of the photographic plate; thus the analogous information of the

photogram is digitized. For well-defined images of stars and satellites the precision of the coordinate measurement is about $\pm 1\,\mu m$, corresponding to an angular resolution of about $\pm 0.''5$.

Within the further reduction process the measured coordinates of the fiducial stars are compared with the star positions from an appropriate star catalog, representing the fundamental reference system [2.1.2]. One such classical catalog is the *SAO Star Catalog*, that was compiled for the purposes of satellite geodesy and has been used for the adjustment of the BC4 satellite network. Fore more information on star catalogs see Eichhorn (1974), and for modern developments Walter, Sovers (2000).

Before starting the plate reduction process the measured coordinates may be corrected for

- radial and tangential distortion,
- astronomical refraction,
- satellite refraction,
- satellite aberration, and
- satellite phase.

The correction for *distortion* is possible when the coefficients K_i of a polynomial, which describes the distortion of the particular camera objective lens, are known. The effect of the *astronomical refraction* (z is the zenith distance, P: atmospheric pressure [HPa], t: temperature in centigrade)

$$\Delta z \;=\; R = A\tan z + B\tan^3 z,$$

$$A \;=\; 58.''294\left(\frac{P}{1013}\right)\left(\frac{283}{273+t}\right) \quad \text{and} \quad B = 0.''0668 \tag{5.1}$$

is not uniform for the whole photogram when a wide field camera is used. It is not the absolute amount of the refraction influence that is important but the variation within the field of view. Formulas for this *differential refraction* can be taken from astrometric literature (e.g. Seeber, 1972; Kovalevsky, 1990). For small field observations as in CCD astrometry [5.2], see e.g. Schildknecht (1994).

Stars are at infinite distance; the satellite, however, is often passing within the outer limit of the effective atmosphere. The problem is schematically pictured in Fig. 5.5. The astronomical refraction Δz_∞ and the portion of the refraction Δz which influences the light from the satellite, differ by the so-called *satellite refraction* σ (Schmid, 1977)

$$\sigma \;=\; \frac{\rho s\,\Delta z}{d\cos z} \tag{5.2}$$
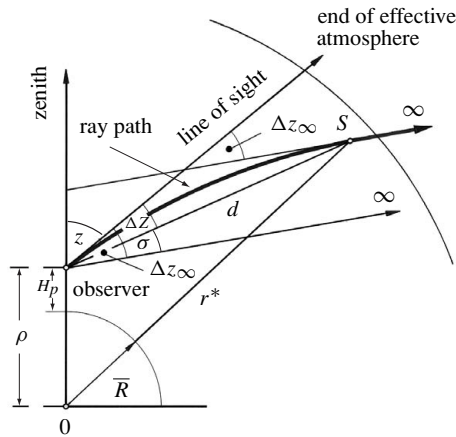
Figure 5.5. Satellite refraction

where

    ρ the geocentric distance of the observation station,

    s constant value $= 0.00125$, and

    d observer – satellite distance.

A new discussion of satellite refraction, after the advent of CCD astrometry, is given by Bretterbauer (2001).

The correction for *aberration* reduces the observation epochs for all participating stations to a common epoch defined at the satellite. The *phase correction* reduces the reflected images of the sun to the center of the satellite. Explicit formulas can be taken from Schmid (1977). Note that the last two geometrical corrections are also important for laser ranging to satellites [8.4].

Plate reduction

The photograph of the star field is nothing else but the projection of the astronomical sphere into a plane. If we assume ideal conditions, i.e. a rigorous central perspective projection without distortion, refraction etc., we can compute plane *tangential coordinates* ξ, η from the equatorial star coordinates α, δ with respect to a known camera orientation α_0, δ_0 (Fig. 5.6). The ideal tangential coordinates ξ, η differ from the measured coordinates x, y on the photogram only by random observation residuals v_x, v_y.
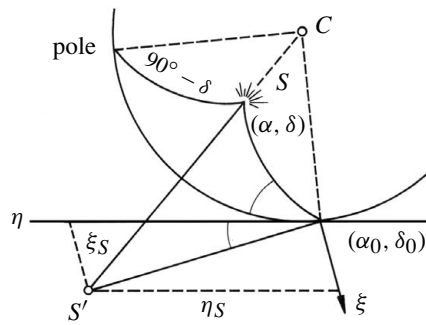
Figure 5.6. Tangential coordinates ξ, η

In practice such ideal conditions do not exist. Within the *plate reduction* process we try to find an adequate model for the relation between tangential star coordinates ξ, η and measured coordinates x, y. Once the parameters of this model are identified, they can be used to transform the measured satellite coordinates x_S, y_S via the tangential coordinates ξ_S, η_S into equatorial satellite directions α_S, δ_S.

Usually the plate reduction models are subdivided into astrometric methods and photogrammetric methods, because they are based on developments from both fields. The differences are, however, more in the formulation than in the results.

Within the *astrometric plate reduction* model the tangential coordinates ξ, η and the plate coordinates x, y are related through polynomials. The tangential coordinates ξ, η are determined with the formulas of the *gnomonic projection* (Green, 1985; Smart, 1977). Following Fig. 5.6 we introduce a quantity q with

$$q = \cot \delta \cos(\alpha - \alpha_0),$$

and obtain

$$\xi = \frac{\tan(\alpha - \alpha_0) \cos q}{\cos(q - \delta_0)} \quad \text{and} \quad \eta = \tan(q - \delta_0). \tag{5.3}$$

When the camera orientation is already well known we can use the simple linear relations:

$$
\begin{aligned}
v_x &= Ax + By + C - (\xi - x), \\
v_y &= A'x + B'y + C' - (\eta - y).
\end{aligned}
\tag{5.4}
$$

A more general form with quadratic terms, which allow for corrections in the camera orientation, is:

$$
\begin{aligned}
v_x &= Ax + By + C + Dx^2 + Exy + Fy^2 - (\xi - x), \\
v_y &= A'x + B'y + C' + D'x^2 + E'xy + F'y^2 - (\eta - y).
\end{aligned}
\tag{5.5}
$$

$A, B, C, \ldots$ are parameters in the adjustment process. In astrometry, the equations (5.4) and (5.5) are often named *Turner's formulas* (e.g. Smart, 1977).

Within the *photogrammetric plate reduction* model an attempt is made to formulate analytically as many influences as possible within the functional model. The relations are based on the general formulas for the perspective projection. From Fig. 5.7, with the approximative values

t_0, δ_0, κ_0 for the elements of exterior orientation: camera orientation angles and swing,

f_0, x_{H_0}, y_{H_0} for the elements of interior orientation: camera constant and principal point,
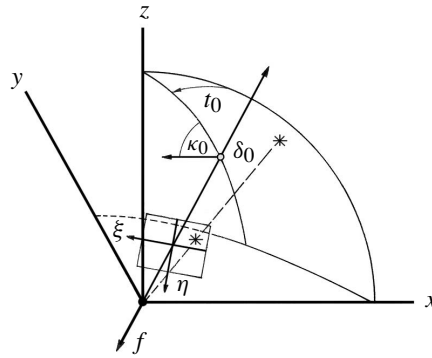
the tangential coordinates ξ, η of the stars can be expressed as

Figure 5.7. Perspective projection and camera orientation

$$
\xi = f_1(t_0, \delta_0, \kappa_0, f_0, x_{H_0}, y_{H_0}) \quad \text{and} \quad \eta = f_2(t_0, \delta_0, \kappa_0, f_0, x_{H_0}, y_{H_0}).
\tag{5.6}
$$

The basic model contains six parameters in the observation equations:

$$
\begin{aligned}
v_x &= \frac{\partial \xi}{\partial \delta}\Delta\delta + \frac{\partial \xi}{\partial t}\Delta t + \frac{\partial \xi}{\partial \kappa}\Delta\kappa + \Delta x_H + \frac{\partial \xi}{\partial f}\Delta f - (x - \xi), \\
v_y &= \frac{\partial \eta}{\partial \delta}\Delta\delta + \frac{\partial \eta}{\partial t}\Delta t + \frac{\partial \eta}{\partial \kappa}\Delta\kappa + \Delta y_H + \frac{\partial \eta}{\partial f}\Delta f - (y - \eta).
\end{aligned}
\tag{5.7}
$$

$\Delta\delta, \Delta t, \Delta\kappa, \Delta f, \Delta x_H, \Delta y_H$ are corrections to the approximate values in (5.6). The derivatives in (5.7) and the explicit formulas (5.6) can be taken from Schmid (1977). Because of the non-linearity in (5.7), the adjustment process requires several iterations. The model can be refined with additional parameters, e.g. for comparator biases, radial

and tangential distortion, or refraction biases. Equations with more than 20 parameters were used in the adjustment of the BC4 geometric world network (Schmid, 1977).

Not all individual satellite images on the photogram are required for further computations; the total information is condensed with a smoothing function into one or more selected points. Usually a polynomial in x and y is formulated as a function of time:

$$x' = a_0 + \sum_{i=1}^{k} a_i t^i, \quad y' = b_0 + \sum_{i=1}^{k} b_i t^i. \tag{5.8}$$

The order k of the polynomial representation depends on the particular satellite orbit and on the camera type. For PAGEOS and ZEISS BMK the order 5 to 6 was appropriate. The interpolated fictitious satellite points are converted into tangential coordinates and then into topocentric equatorial directions $\overline{\alpha}_S$, $\overline{\delta}_S$ to the satellite using the adjusted model parameters (5.7) and the reciprocal formulas of (5.6).

The accuracy of a single direction in the photogram was found to be $\pm 1.''6$, and for the smoothed central direction $\pm 0.''35$, based on analysis of more than 1000 observations in the BC4 worldwide network. With improved star catalogs, improved comparator techniques, and careful analysis of all existing error sources, the accuracy of adjusted directions to satellites based on photographic observations may be in the order of $\pm 0.''1$ to $\pm 0.''2$. This corresponds to a position accuracy of ± 0.7 m to ± 1.5 m for a satellite at 1500 km altitude (e.g. AJISAI).

5.1.4 Spatial Triangulation

The plate reduction gives directions in space between the ground station and satellite positions as a function of time. The directions are referred to the space-fixed stellar reference system CIS [2.1.2.1], because they are derived from star positions. The directions are represented by unity vectors e_{ij}. According to Fig. 5.8 we derive

$$e_{ij} = \begin{pmatrix} \Delta x_{ij} \\ \Delta y_{ij} \\ \Delta z_{ij} \end{pmatrix} = \begin{pmatrix} \cos\delta \cos\alpha \\ \cos\delta \sin\alpha \\ \sin\delta \end{pmatrix}. \tag{5.9}$$

The directions can be used as observables for orbit computations and orbit analysis [3.3], and hence contribute to dynamical satellite geodesy and to the determination of Earth models [12.2]. They can, however, also be used as elements in the construction of a purely geometric satellite network (cf. [12.1]). The latter procedure was, originally, proposed by the Finnish geodesist Väisälä (1946) and named *stellar triangulation*. Väisälä performed the first experiment in 1959, when he used radio-sonde balloons as optical targets. The realization of Väisälä's idea with artificial satellites led to the first global geometric geodetic network [5.1.5].

With simultaneous observations at P_1 and P_2 in Fig. 5.8, the unity vectors e_{11} and e_{21} intersect in S_1 and define a plane which is determined through its normal vector

$$n_1 = e_{11} \times e_{21}. \tag{5.10}$$

For a second satellite position S_2 (the same satellite at another position, or a different satellite) a second plane is defined in space:

$$n_2 = e_{12} \times e_{22}. \qquad (5.11)$$

With

$$g_{12} = \frac{n_1 \times n_2}{|n_1 \times n_2|} \qquad (5.12)$$

we obtain the unity vector, and hence the direction in space, between the two ground stations P_1 and P_2. The vector g_{12} gives the orientation be-tween two distant stations. Addi-

Figure 5.8. Spatial triangulation withdirections

tional ground stations can be included for the construction of larger networks (Fig. 5.9), or even global networks (Fig. 5.10). Several "events" are observed from each pair of ground stations, in order to increase the accuracy.

A network which consists only of directions or angles has a so-called *datum defect*; the scale and the origin are not fixed. At least one baseline has to be measured for the determination of the scale. This can, for example, be a laser range measurement between a ground station and a satellite position, or a distance determination with terrestrial techniques between two ground stations. Usually the latter solution has been adopted in satellite geodesy.

In order to fix the origin of the coordinate system, either the coordinates of

Figure 5.9. Satellite triangulation

one station have to be defined arbitrarily, or geocentric coordinates of one or several network stations are determined with alternative techniques, e.g. through Doppler measurements with known orbits ([6.6], [12.1]). In other words, the purely geometric techniques require additional dynamical information for the determination of a datum [4.1]. For a full treatment of the subject see Schmid (1977).

5.1.5   Results

The method of satellite photography was used frequently between 1964 and 1975 for the establishment of regional, continental and global geometrical networks, within national and international projects. Of historic importance is the geometric worldwide

satellite network of the National Geodetic Survey (NGS), the former US Coast and Geodetic Survey (Schmid, 1974, 1977). The dedicated balloon satellite PAGEOS was launched, and observations with BC4 cameras were carried out by 16 groups at 45 globally distributed stations (Fig. 5.10), between 1966 and 1970. The mean interstation distance was about 3000 to 4000 km. ECHO-1, ECHO-2 and GEOS-2
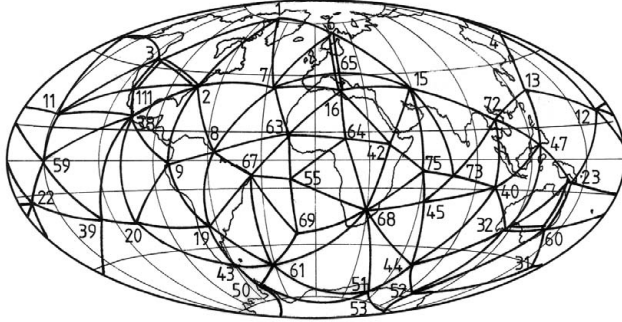
Figure 5.10. Geometrical world network, observed with BC4 cameras; terrestrial baselines are indicated as double-lines

were also observed. The observation period at a particular station could last more than one year before sufficient simultaneous events were obtained. The total number of plates in the reduction process amounts to 2350, with

856 2-stations events,
194 3-stations events,
 14 4-stations events.

These numbers illustrate the difficulties of having favorable observation conditions simultaneously at remote stations.

The scale was introduced into the network through seven terrestrial baselines, measured with a laser-geodimeter. The coordinates of the station Beltsville in Maryland were held fixed. The purely geometric adjustment provided three-dimensional Cartesian coordinates for the 45 stations in a homogeneous global reference frame with a mean standard deviation

$$\sigma_p = \sqrt{\sigma_\varphi^2 + \sigma_\lambda^2 + \sigma_h^2} = 4.53 \text{ m}.$$

The mean equatorial radius (semi-major axis) was found from this solution to be

$$a = 6\,378\,130 \text{ m}.$$

The transformation into a geocentric reference frame became possible via a set of geocentric coordinates from a Doppler solution [6.6] for some of the stations.

The historical importance of the worldwide BC4 network lies in the fact that, for the first time, a solution was obtained for the fundamental scientific problem of geometric geodesy, i.e. the determination of a global polyhedron (*Cage of Bruns*) [1.2].

Today photographic camera observations to artificial satellites are no longer used because of the great effort required to make and adjust the observations, and the rather low accuracy. Instead, CCD technology is replacing this traditional method. The basic methodological foundations, in particular the plate reduction techniques, however, are of continuing importance.

5.2   Directions with CCD Technology

In the last decade of the 20th century the fast development of electronic position sensors, in particular the *Charge Coupled Device* (CCD), initiated a revival of optical methods in astrometry and also satellite geodesy. The key factors, when compared with traditional photographic, or even visual methods are:

 − higher sensitivity, improved accuracy, shorter observation time,
 − the image information is available in digital form,
 − fully automatic data flow; no time consuming coordinate measurement necessary, and
 − availability of new star catalogs with sufficient and accurate reference stars.

As a consequence, in today's astrometry CCD technology is nearly universally applied, e.g. for the construction of ground based star catalogs (Ashford, 2001). In geodetic astronomy many classical observation techniques have been supplanted by new technology based on the use of CCD (Bretterbauer, 1997; Fosu, 1998; Gerstbach, 1999; Hirt, 2001). In the last decade various new applications have arisen in satellite geodesy (Schildknecht, 1994; Hugentobler, 1998; Ploner, Jackson, 1999). Satellites are fast moving objects and hence generate particular problems; however, much experience and many solution concepts can be taken from astrometry and classical photographic satellite tracking. The use of CCD technology in satellite geodesy will certainly grow and deliver significant results.

The basic objective is to determine the orientation of a camera with respect to the inertial frame. The camera may be either fixed to the ground (Earth based observatory) or to a space vehicle (satellite, rocket, platform, cf. Fig. 5.11). In both cases the orientation angles *declination* δ_0, *right ascension* α_0 (or *hour angle* t_0), and the *swing angle* κ_0 around the camera axis have to be determined (see Fig. 5.7). The process widely follows the procedure that has been developed in the photographic technique [5.1]. The main steps are depicted in Fig. 5.12.

5.2.1   Image Coordinates from CCD Observations

A digital image is composed of so called *pixels* (picture elements) that are arranged in the form of a matrix with r rows and c columns. The image information is represented by intensities (grey values), usually varying between 0 and 255. The position of a
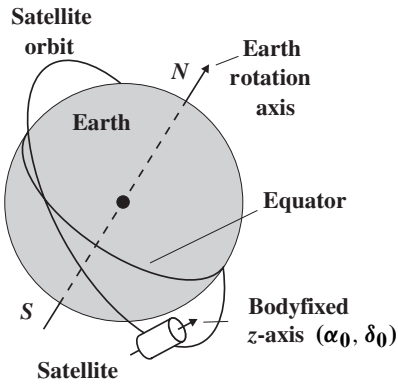
Figure 5.11. Orientation of a satellite in space
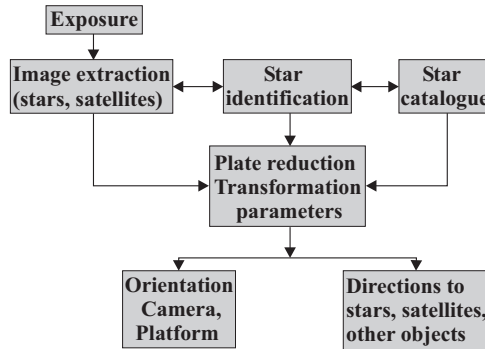
Figure 5.12. Process of CCD observations

pixel within the picture is defined by two dimensional coordinates (r, c), indicating the particular row and column. These are naturally discrete values. The origin of the image coordinate system is often transferred to the center of the sensor (Fig. 5.13). The measured coordinates are hence x, y, and correspond with the plate coordinates defined in [5.1.3]. The coordinates x_{S_i}, y_{S_i} of a star or satellite image covering several pixels can be determined with sub-pixel accuracy (see later). These coordinates hence are non-discrete continuous numbers.
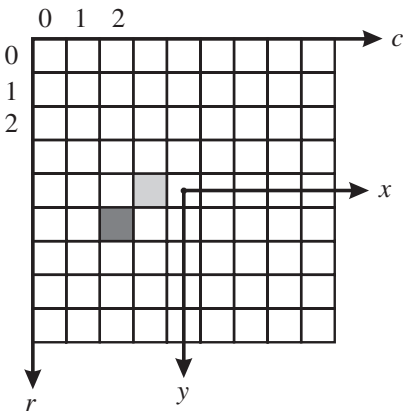
Figure 5.13. Image coordinate system
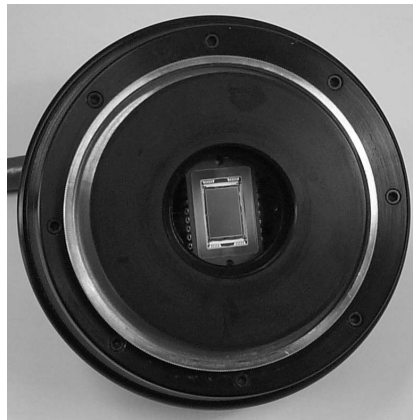
Figure 5.14. CCD camera Apogee KX2E

A CCD camera (Fig. 5.14) uses a CCD sensor instead of the photographic plate or film to store the image information. The technique was invented by 1970 in the U.S.A. The CCD sensor makes use of the photoelectric effect in silicon to convert photons into charges. The sensor or chip consists of a certain number of lines and columns

forming an array of pixels. To give an example, the CCD-chip Kodak KAF-1602E that is often used in small astrometric cameras has 1530×1020 pixels and measures 13.8 $\times$ 9.2 mm. The pixel size is 9μm $\times 9\mu$m. The corresponding field of view depends on focal length and is in most cases far below $1° \times 1°$. Arrays of about 1000×1000 pixels are standard. Larger arrays are available but are still rather expensive. The market, however, is developing very fast. For detailed information on CCD technology see the literature on digital photography. Photoelectricity in astrometry is discussed in detail by Kovalevsky (1995). A good overview with respect to requirements for taking fast moving objects (satellites) is given by Schildknecht (1994).

In order to obtain image coordinates for objects of interest the images of stars and satellites have to be recognized, and the coordinates of the image centers x_i, y_i have to be determined. This is the process of *image extraction*. The images are considered to be a group of pixels with similar properties; they differ from the background through significantly higher grey values. There exist a number of techniques of image extraction, developed in the field of digital image processing. With proper weighting centering algorithms may lead to accuracies of 0.1 to 0.2 pixels for the image centers. Depending on the camera's focal length this may correspond to 0.1 arcseconds or better. Significant improvements can be expected with progress in CCD technology. For details on algorithms see e.g. Schildknecht (1994); Hirt (2001).

For fast moving objects like satellites precise epoch registration is of particular importance. For solution concepts see e.g. Schildknecht (1994); Ploner, Jackson (1999).

5.2.2   Star Catalogs, Star Identification and Plate Reduction

In the next step star images in the digital photogram have to be identified and related to the equatorial positions given by a star catalog. Because of the small field of view and high sensitivity of CCD sensors, star catalogs with a very high number of precise star positions, down to apparent magnitudes of 15^m or fainter, are required. Such catalogs have only recently become available, or are still under construction. Traditional catalogs are by far insufficient. Table 5.1 gives an overview.

Table 5.1. Recent star catalogs and aptitude for CCD astrometry

catalogue	stars	magn.	stars/ $\Box°$	μ	$\sigma_{pos.}[^{\prime\prime}]$	aptitude
HIPPARCOS	118 000	12.4	3	yes	0.01	very low
TYCHO-2	2 500 000	14.5	60	yes	0.06	high
GSC	19 000 000	15.5	460	no	0.5–1.0	low
UCAC	80 000 000	16.0	2000	yes	0.02–0.07	very high

The *HIPPARCOS Catalog* is the main result of the HIPPARCOS mission (see [5.3.2]), one of the most important astrometric endeavors of the last century. HIPPARCOS delivered positions, proper motions, and parallaxes of about 118 000 stars

with an accuracy of milliarcseconds. The catalog is today the most important and most accurate realization of the Celestial Reference System (ICRS) at optical wavelengths (Walter, Sovers, 2000). It is, however, not suited for CCD astrometry because of the low density of only 3 stars per square degree. On most CCD images for the determination of directions to satellites no HIPPARCOS star would be available.

The *Tycho-2 Catalog* (Hog et al., 2000) was observed together with the Hipparcos mission but with lower accuracy. It contains 2.5 million stars over the complete sky. The proper motions were determined by comparison with old ground-based observations. The star density varies between 25 stars per square degree near the galactic poles and 150 stars per square degree near the galactic equator. The average position accuracy is $0.''06$, and for proper motions $0.''025$/year. Tycho-2 hence will be the most appropriate catalog for CCD astrometry until the complete publication of the UCAC.

The *Guide Star Catalog* (GSC) was compiled for the orientation of the *Hubble Space Telescope* (HST). The catalog contains a large number of star positions, but no proper motions. The position accuracy is rather low; the mean epoch of the ground based observations is 1983, hence the accuracy is rapidly decreasing. GSC positions are not suited for high precision work with CCD sensors.

The *U.S. Naval Observatory CCD Astrograph Catalog* (UCAC) (Zacharias et al., 2000; Sinnott, 2001) is under construction with the objective to provide a catalog of highest density for both hemispheres. The catalog contains positions and proper motions of stars between magnitudes 7.5 and 16 with a mean accuracy of $0.''02$ to $0.''05$. The results are related to the HIPPARCOS reference frame. The observations are ground based. The USNO twin astrograph has been equipped with a CCD chip of 4096×4096 pixels, covering about one square degree of the sky. The observations started in the southern hemisphere. A preliminary data set is already released. The observations of the northern sky will last until 2003. With about 2000 stars per square degree the UCAC will be the most appropriate catalog for the reduction of CCD images in astrometry and satellite geodesy.

For the process of *star identification* the approximate region of the photogram is delineated in the star catalog, and the equatorial star positions α_i, δ_i are converted to plane tangential coordinates ξ_i, η_i using (5.3) and the approximate camera orientation α_0, δ_0. The two point ensembles x_i, y_i and ξ_i, η_i are matched against one another with a suitable algorithm using translation, rotation, and scale until highest correlation is achieved. To start with, some arbitrary points from both ensembles are set to be identical. To accelerate the process, the search algorithm can be restricted to the brightest stars of the field (e.g. Quine, 1996).

The *plate reduction* itself follows the procedure as described in [5.1.3]. Because of the narrow field of view the astrometric model is appropriate (Schildknecht, 1994; Ploner, 1996; Hirt, 2001). Emphasis may be given to corrections for
  − astronomical refraction,
  − satellite refraction,
  − dispersion, and
  − aberration.

For details see [5.1.3] and Seeber (1993), Schildknecht (1994), Ploner (1996).

As a result, the orientation of the camera axis in inertial space (cf. [4.3.3.2]) and/or the directions from the camera position to space objects like satellites are obtained.

5.2.3   Applications, Results and Prospects

For CCD observation of fast moving objects like satellites two particular aspects are of importance, namely the angular velocity of the object with respect to the stars or to the camera, and the intensity of light crossing the pixels. Table 5.2 gives an idea. Depending on the telescope and the pixel size the *pixel crossing time* ranges between a few and several hundred milliseconds. For details see e.g. Schildknecht (1994).

Table 5.2. Observational characteristics for fast moving objects

	GEO	GPS	LAGEOS	ERS
Altitude (km)	36 000	20 000	6 000	780
Max. motion [arcs/s]	15	30	240	2000
Magnitude [m_v]	11	8 – 14	14	< 6

Telescopes can either follow the stars or the satellites, or be held fixed. Successful observations have been performed since about 1990 with existing satellite telescopes, e.g. with the ballistic camera Zeiss BMK 75 [5.1.2] in Austria (Ploner, 1996) or with the 0.5 m SLR telescope [8.3] in Zimmerwald, Switzerland (Schildknecht, 1994). In 1996 a combined *Laser Ranging and Astrometric Telescope* (ZIMLAT) was established in the fundamental station at Zimmerwald (Hugentobler, 1998). The mapping scale of the latter camera is about $0.''8$ / pixel.

Successful observations have been reported for a large number of geostationary satellites, as well as for GPS, LAGEOS and GFZ-1. The accuracy in orientation to GEO satellites was found to be about $0.''5$, and for GPS satellites $0.''1$ to $0.''2$ (Ploner, Jackson, 1999). Important results are, for example (Hugentobler, 1998):
- − determination of the resonant geopotential terms C_{22} and S_{22} from precise geostationary orbits,
- − control of space debris in high orbits and the calibration of alternative observation techniques, and
- − determination of the complete position vector to satellites of interest in geodesy.

With further developments in CCD technology, e.g. larger arrays, smaller pixel size and improved time tagging the optical determination of directions to satellites will again play a significant role in satellite geodesy.

5.3   Directions from Space Platforms

We distinguish two different tasks, (a) the determination of the orientation of a space platform, and (b) the determination of directions to objects (e.g. stars) from a space

platform, both with respect to the inertial frame. The first problem can be solved with star-trackers, the second is related to space astrometry using astrometric satellites. Both subjects have received tremendous support from the recent developments in CCD technology.

5.3.1   Star Tracker

The purpose of a star tracker or star sensor is to measure the direction to a star within the reference frame fixed to the sensor body. Star sensors are mainly used for orientation and attitude control of space vehicles [4.3.3.2]. The basic concept closely follows the procedure of CCD astrometry as explained in [5.2].

The star tracker is basically a digital camera which takes photographs of stars or sets of stars in the direction of its optical axis. The first step is to identify a star or set of stars, with reference to the onboard star catalog. Other than in space astrometry the approximate initial orientation of the star tracker cannot be taken as known. This is why powerful algorithms are required for matching the pattern of image stars against the catalog positions without any a priori knowledge of the sensor orientation (Quine, 1996). This type of sensor is also called *autonomous star tracker*.

The second step is to track the star(s), and to control the orientation of the platform. The last step is to refine the orientation with respect to the required frame. Star sensors have excellent long-term stability and provide accuracy in attitude control of down to 1 arcsecond. For details on instruments and algorithms see e.g. Quine (1996); Sidi (1997); Renken (1999).

5.3.2   Astrometric Satellites, HIPPARCOS

Astrometric satellites have opened a new era in astrometry. The unique features of a satellite as observation platform, if compared with ground based astrometry, are:
-   the satellite is able to observe the entire sphere from its location in space,
-   the absence of atmospheric disturbances, and
-   high instrumental stability brought about by the absence of gravitational instrumental flexure.

These features initiated a dramatic improvement in the accuracy of star positions over three orders of magnitude from 1 arcsecond to below 1 milliarcsecond (mas). Further improvement down to 1 microarcsecond (μas) is expected.

The first astrometric satellite was HIPPARCOS. The acronym stands for *HI*gh *P*recision *PAR*allax *CO*llecting *S*atellite. The principle of HIPPARCOS was invented by *Lacroute*, a French astronomer, in 1966 (Kovalevsky, 1995). After several modifications it was accepted as an ESA mission in 1980, and eventually launched on August 8, 1989, into a highly elliptical orbit instead of the planned geostationary one. This was due to a failure in the apogee booster. As a consequence the perigee is about 500 km, and the apogee close to 36 500 km. Nevertheless the satellite, after a complete modification of the observation program, worked successfully for about 3.5 years until March 1993.

The primary objective of the mission was to determine positions, annual proper motions, and trigonometric parallaxes (distances) of about 118 000 carefully selected stars, at the level of 1 mas. The results are compiled in the HIPPARCOS catalog (Kovalevsky et al., 1997), published by ESA in 1997 in a seventeen-volume document (also available on CD-ROM). In addition the lower-accuracy TYCHO catalog with more than one million stars and 20 to 30 mas in astrometric results was released.

The basic observing principle is to scan the sky and measure a wide angle of about 58° between pairs of stars in such a way that the number of observations of a star is maximized. The angle is realized through a complex beam-combining mirror (see Fig. 5.15) which reflects the light from two fields of view into the Schmid telescope. In the focal surface is a grid consisting of 2688 slits. When the satellite rotates, the incident light from the stars is modulated by the grid and the photon counts are registered according to the satellite scan.
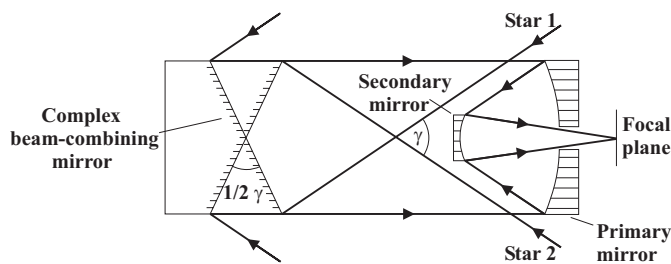
Figure 5.15. Observation principle of HIPPARCOS

HIPPARCOS was designed to spin by one revolution each two hours, together with a slow change in the orientation of the rotation axis. In this way the telescope could scan the whole sky several times during the mission (see Fig. 5.16). For details of the technical design and the whole mission see e.g. Kovalevsky (1995); Walter, Sovers (2000) and the ESA publication of the HIPPARCOS catalog .

5.3.3   Planned Missions

After the overwhelming success of the HIPPARCOS mission, plans are designed for several follow-on missions by NASA and ESA.

FAME
FAME stands for *F*ull Sky *A*strometric *M*apping *E*xplorer. It is a project under the leadership of the U.S. Naval Observatory (USNO) and was originally planned for launch in 2004 as a NASA mission. FAME will have a geosynchronous orbit and work on a similar principle to HIPPARCOS. The telescope will have two fields-of-view separated by a 65° basic angle. The spacecraft will rotate with a 40 minute period, the telescope sweeping out a great circle on the sky. The spacecraft rotation axis, at a 35° angle to the sun, will perform, a precession about the sun-spacecraft line
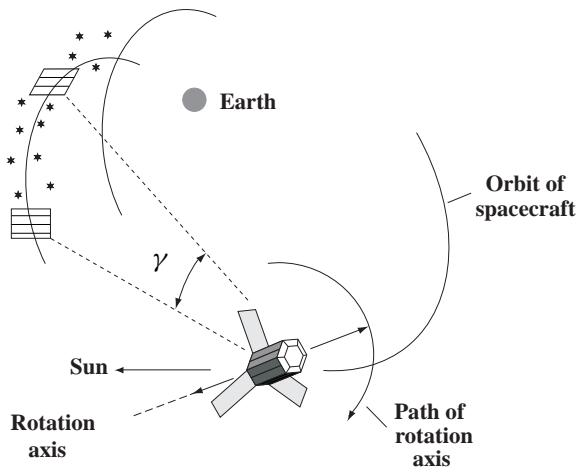
Figure 5.16. HIPPARCOS scanning the sky

with a 20 day period (Fig. 5.17) due to solar radiation pressure. The orbital motion of Earth around the sun, along with the spacecraft rotation and precession will result in complete sky coverage after 3 months. The telescope will includes 24 large format CCDs. The images of stars will continuously traverse the CCD array and be time
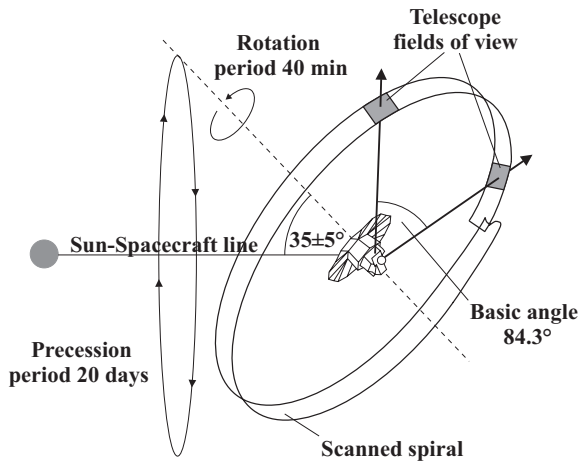
Figure 5.17. FAME observing concept

tagged as the spacecraft rotates. For technical details see publications of the USNO or e.g. Seidelmann et al. (1998). The objective is to measure the positions, parallaxes, and magnitudes of 40 million stars brighter than $15^{\rm m}$ with an accuracy of 50 μas for

the brighter stars and up to 300 μas for the fainter stars. The realization of the mission is now questionable.

DIVA

To bridge the gap until the launch of GAIA a small astrometric satellite DIVA (*D*ouble *I*nterferometer for *V*isual *A*strometry) has been proposed (e.g. Bastian, 1998). It is aimed to measure positions, proper motions and parallaxes of at least 30 million stars with an accuracy about five times better than HIPPARCOS. The basic principle is similar to that of HIPPARCOS but realized with advanced technology at much lower cost. The satellite shall be launched into a highly eccentric geosynchronous orbit (apogee: 71 000 km, perigee: 500 km). It will scan the sky and measure an angle of 100 degrees using the interferometric principle and CCD sensors. The mission is, however, not yet approved.

GAIA

The European Space Agency (ESA) has designed a next generation mission named GAIA. GAIA stands for *G*lobal *A*strometric *I*nterferometer for *A*strophysics. It combines the basic principle of HIPPARCOS with some new features. GAIA will not be orbiting Earth but operate at the Lagrangian point L_2, which is located some 1.5 million kilometers from Earth in the direction away from the sun (see Fig. 3.29, p. 134). Here the satellite will be free from eclipses and be in a stable thermal environment.

GAIA will be 2 to 3 orders of magnitude more accurate than HIPPARCOS, and will provide positions, distances and proper motions of about 1 billion stars at the 10 microarcseconds accuracy level. The spacecraft should be launched around 2010 to 2012 and will be operated for five years (Perryman, Pace, 2000).

6   Doppler Techniques

The *Doppler effect* or *Doppler shift*, named after the Austrian physicist *Christian Doppler* (1803–1853), denotes the

> *difference between the frequency of the radiation received at a point and the frequency of the radiation at its source, when observer and source are moving with respect to each other* (e.g. NGS, 1986).

The effect is well known from the fact that an acoustic signal, emitted by a vehicle passing the observer with high speed, shifts suddenly from a higher to a lower tone. When the vehicle is approaching, the observer receives a higher frequency compared to a non-moving source; and with increasing range a lower frequency is observed (Fig. 6.1). It is evident that the frequency shift depends on the relative velocity between source and observer.
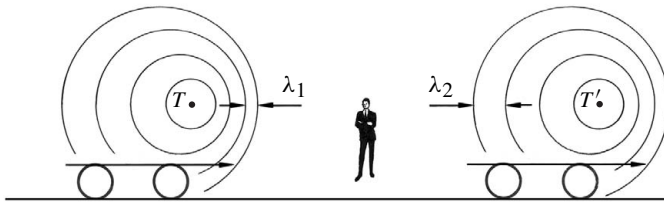
Figure 6.1. Doppler effect

Observation techniques based on the Doppler principle are widely used in science and technology for the determination of velocities. In astronomy the radial velocity of distant cosmic objects can be derived from the Doppler shift of spectral lines (e.g. red shift of galaxies). In navigation velocity over the ground can be obtained by using the Doppler effect (*Doppler log*). In space technology and satellite geodesy Doppler observations played an important role from the beginning, because most of the early satellites transmitted on a stable frequency. Velocity changes, and hence orbital elements, could be derived from the measured frequency shift with rather simple equipment. Using these techniques the orbital elements of the very first artificial satellites were determined and published without any other information (Priester, Hergenhahn, 1958).

The frequency shift caused by the Doppler effect can be described as a function of time (Fig. 6.2), and yields a characteristic curve. The *Doppler curve* of an artificial satellite lies in between curves (2) and (3), depending on the satellite's range and altitude. If the observer were directly in the path of the satellite the Doppler shift Δf would stay constant at all times, the only change being a sudden jump (curve (4)) from $f + \Delta f$ to $f - \Delta f$ as the satellite passed the observer. With increasing distance of the observer from the orbit, the frequency change becomes smoother and is described by

an *S*-shaped curve. In case (1) the observer is so far away that no significant change of range occurs; the curve deforms into a straight line. The turning point of the curvature corresponds to the *time of closest approach* (TCA) of the satellite with respect to the observer.
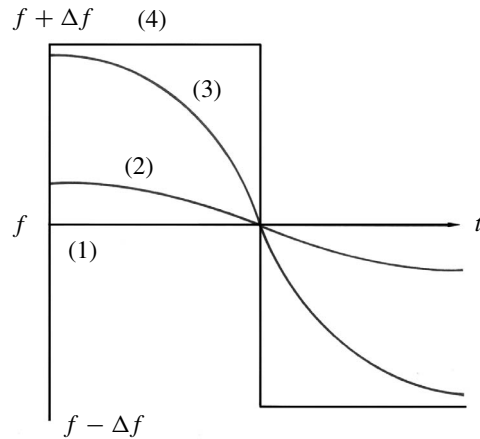
Figure 6.2. Doppler curves for different distances between observer and radiation source

It was recognized early on that the method of orbit determination based on Doppler measurements from a known tracking station on the ground was reversible, i.e. that unknown observer positions could be derived from the measured Doppler shift when the satellite positions are known (Guier, Weiffenbach, 1960, 1998). This idea led to the development of a global satellite-based navigation system, the *Navy Navigation Satellite System* (NNSS), also known as the *NAVSAT* or *TRANSIT* system [6.2]. The concept was designed in 1958; the system was declared operational for the American Navy in 1964, and it was released for world-wide civil use in 1967. Because of its continuously increasing accuracy, the TRANSIT system became of interest for geodetic applications about 1970. For a geometric interpretation of the basic principle see [4.2.3].

All satellites that transmit continuously on stable frequencies can be used for Doppler measurements. Many such satellites were launched in the early years of the space age (NGSP, 1977, Vol. I, p. 64). The observations were primarily used for orbit analysis, and they were also introduced into the solutions of Earth models.

When the Doppler principle is to be used for the establishment of a precise, global navigation system with real-time capability, the satellite system must meet at least the following requirements:

- global distribution of satellite orbits, and
- real-time transfer of information about satellite positions and time to users.

The availability of at least two carrier frequencies is of advantage for estimation of the ionospheric propagation delay. Furthermore, the user segment can be much sim-

pler with one-way observation techniques, i.e. when the satellites are the only active transmitters.

The principle works in reverse, i.e. ground beacons transmit on stable frequencies, and a receiver onboard the satellite measures the Doppler count. This idea has been realized in the current DORIS system [6.7]. Note that the Doppler effect is also used in GPS data reduction [7.3].

6.1 Doppler Effect and Basic Positioning Concept

The *Doppler effect*, or the *Doppler equation*, for an electromagnetic wave can be written as (e.g. Wells, 1974; Vaníček, Krakiwsky, 1986)

$$\frac{f_r}{f_s} = \frac{1 - \frac{v}{c}\cos\theta}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}}, \tag{6.1}$$

with (cf. Fig. 6.3)

f_s stable frequency transmitted from the satellite S,

f_r received (Doppler shifted) frequency at the observation station P,

v satellite velocity,

c velocity of light,

θ angle between the velocity vector of the satellite and the observer-satellite line of sight, and

r distance between observer and satellite.
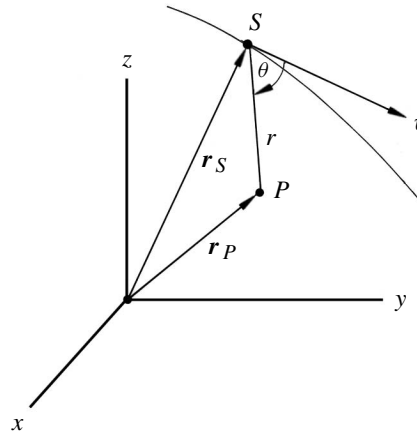
Figure 6.3. Explanation of the derivation of the Doppler effect

The relative velocity is

$$\dot{r} = \frac{dr}{dt} = -v\cos\theta. \tag{6.2}$$

Using (6.1), it follows that

$$f_r = f_s\left(1 - \frac{\dot{r}}{c}\right)\left(1 + \frac{v^2}{2c^2} + \frac{3v^4}{8c^4} + \cdots\right). \tag{6.3}$$

When higher order terms are neglected:

$$f_r = f_s \left(1 - \frac{1}{c} \frac{dr}{dt} \right). \qquad (6.4)$$

The simplification (6.4) is permitted when $v \ll c$. The neglected term

$$\Delta_R = \frac{v^2}{2c^2} + \frac{3v^2}{8c^4} + \cdots$$

is a measure of the *transversal Doppler effect* (e.g. Kleusberg, 1984; Schneider, 1988) and shows, according to general relativity, that the Doppler effect does not vanish completely for motions perpendicular to the line of sight. The relativistic corrections have to be considered for very high accuracy requirements [6.4.4].

Basically, the Doppler shift can be determined from the difference between the transmitted and the received frequency, following equation (6.4). In practice, the changing difference between the frequency of the received signal f_r and a stable reference frequency f_g, generated within the receiver, is measured during a given time interval, because the instantaneous value of a frequency cannot be observed directly. Actually, the zero crossings of the difference frequency $(f_g - f_r)$ are counted giving the *integrated Doppler count*:

$$N_{jk} = \int_{T_j}^{T_k} (f_g - f_r) \, dT \qquad (6.5)$$

with
   f_g the stable reference frequency (generated in the receiver),
   f_r the received (shifted) frequency, and
   $T_{j,k}$ time marks for start and stop of the counting interval.

The difference $(f_g - f_r)$ is also called the *beat frequency* f_b. For technical reasons the frequencies are selected such that the difference $(f_g - f_r)$ is always positive (cf. Fig. 6.4).

When satellites carrying a Doppler beacon emit time marks, e.g. every even minute of UTC, such as the TRANSIT satellites [6.2], these timing signals can be used to define the counting interval in (6.5). In Fig. 6.5 the satellite has the position S_j at epoch t_j when it emits a time signal. This signal reaches the receiver at epoch T_j and starts the counting interval. Similarly, for the end of the counting interval we denote t_k, S_k and T_k. The following relations are valid, without considering propagation delays and relativistic effects:

$$T_j = t_j + \frac{r_{ij}}{c}, \qquad T_k = t_k + \frac{r_{ik}}{c}. \qquad (6.6)$$

Accordingly, we obtain from (6.5)

$$N_{jk} = \int_{t_j + \frac{r_{ij}}{c}}^{t_k + \frac{r_{ik}}{c}} (f_g - f_r) \, dt. \qquad (6.7)$$
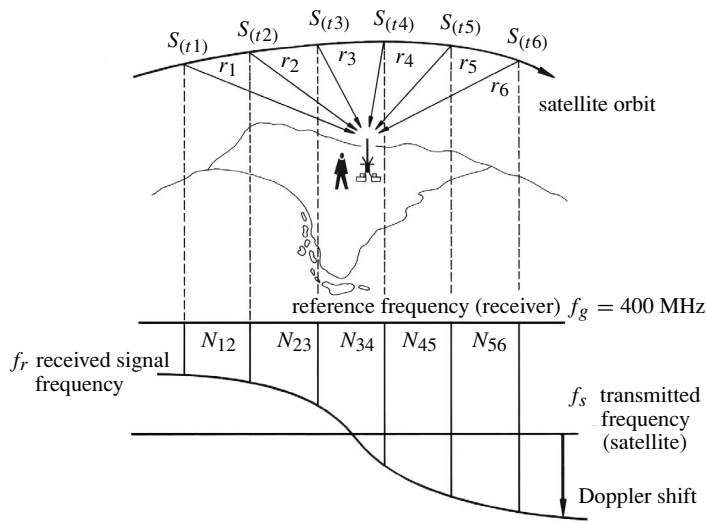
Figure 6.4. Measuring the Doppler shift with the TRANSIT system
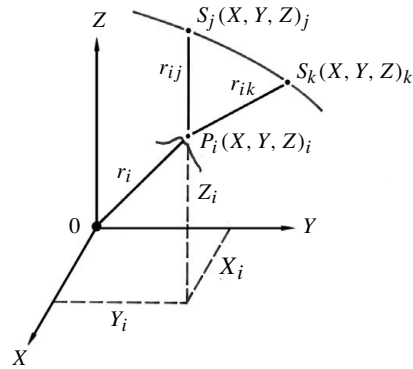
Figure 6.5. Determination of coordinates from Doppler observations

Note that the relation between the observable N_{jk} and the change of range is obtained through the limits of integration.

The number of cycles received between the time marks T_j and T_k must equal the number of cycles transmitted between t_j and t_k, hence

$$\int_{t_j}^{t_k} f_s \, dt = \int_{t_j+\frac{r_{ij}}{c}}^{t_k+\frac{r_{ik}}{c}} f_r \, dt. \qquad (6.8)$$

Substituting (6.8) into (6.7) and solving yields the observation equation

$$N_{jk} = (f_g - f_s)(t_k - t_j) + \frac{f_g}{c}(r_{ik} - r_{ij}). \tag{6.9}$$

The geometric interpretation of this equation is depicted in Fig. 4.7. Equation (6.9) can be written in terms of coordinates (cf. Fig. 6.5). By substituting in

$$r_{ik}^2 = (X_k - X_i)^2 + (Y_k - Y_i)^2 + (Z_k - Z_i)^2, \tag{6.10}$$
$$r_{ij}^2 = (X_j - X_i)^2 + (Y_j - Y_i)^2 + (Z_j - Z_i)^2, \tag{6.11}$$

we have the basic observation equation of Doppler positioning:

$$\begin{aligned}
N_{jk} = \frac{f_g}{c}(\{(X_k - X_i)^2 + (Y_k - Y_i)^2 + (Z_k - Z_i)^2\}^{\frac{1}{2}} \\
- \{(X_j - X_i)^2 + (Y_j - Y_i)^2 + (Z_j - Z_i)^2\}^{\frac{1}{2}}) + (f_g - f_s)(t_k - t_j),
\end{aligned} \tag{6.12}$$

with four parameters only, namely

$$X_i, Y_i, Z_i \quad \text{unknown station coordinates}$$
$$(f_g - f_s) \quad \text{unknown frequency difference.}$$

A more detailed observation equation is discussed later in [6.4] and [6.5].

6.2   One Successful Example: The Navy Navigation Satellite System (TRANSIT)

A very successful system, based on the Doppler principle and used for geodetic purposes, was the TRANSIT system. Other acronyms for the system are NAVSAT (Anderle, 1986), or Navy Navigation Satellite System (NNSS). It was developed in cooperation between the *Applied Physics Laboratory* (APL) of the Johns Hopkins University and the *U.S. Department of Defence* (DOD). The first operational TRANSIT satellite was launched in 1962. Released for civilian service in 1967, TRANSIT has provided reliable positioning and navigation information for nearly 30 years. The TRANSIT program terminated navigation service on December 31, 1996. The former Soviet Union has developed a similar system, named *TSIKADA*; however very little information is available.

For three decades the use of TRANSIT considerably influenced geodetic positioning techniques. With hindsight, TRANSIT can be regarded as the forerunner of GPS. Most of the observation and adjustment techniques used in the geodetic GPS are based on early developments during the TRANSIT era. The development of GPS technology for geodetic positioning is much better understood with a knowledge of the TRANSIT history. For this reason a short overview of the TRANSIT architecture and its use in geodetic positioning is given. For a detailed treatment see e.g. the first editions of this book (Seeber, 1989b, 1993). An excellent overview of all aspects of the system was published by the APL (Pisacane (ed.), 1998).

6.2.1 System Architecture

The original requirements at the conception and implementation of the TRANSIT system were (Guier, 1962):

— real-time navigation accuracy of 200 m ($\widehat{=}$ 0.1 nautical miles),
— no active receivers,
— global coverage, and
— maximum time interval between two consecutive position determinations (*position fixes*), 2 hours.

The TRANSIT satellites were launched into nearly circular polar orbits at an altitude of approximately 1100 km (Fig. 6.6). About seven satellites were usually active in order to maintain global coverage. Depending on the particular distribution of orbital planes the mean time gap between two consecutive *satellite passes* (i.e. the portion of the orbit which can be observed from one station) reaches two hours at the equator and thirty minutes in polar regions. Each individual pass lasts 16–20 minutes, depending on the elevation of the satellite above the horizon.

The TRANSIT satellites were launched by *Scout* rockets. Because of the rather low aiming accuracy of the launchers, which is about $\pm 0.°5$, the orbital inclinations are not exactly 90°. One of the consequences is that the orbit planes experience different precession
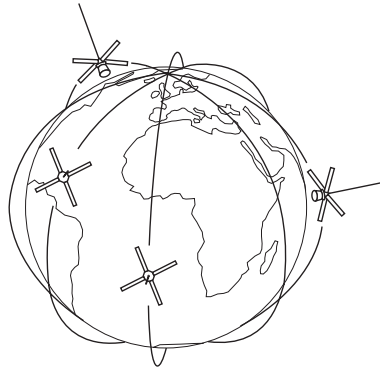
Figure 6.6. TRANSIT orbits

rates (cf. [3.2.2]), and tend to migrate across each other.

Three types of information were transmitted by each satellite:

— two stable frequencies (for the observation of the Doppler shift),
— timing signals every two minutes universal time (UTC), and
— predicted orbital elements (*broadcast ephemeris*), for the determination of satellite positions.

Two types of TRANSIT satellites have been launched: the older type OSCAR and the newer NOVA satellites. Fig. 6.7 shows a functional diagram of the OSCAR satellites. The main components are the 5 MHz quartz oscillator, which generates both carrier frequencies ($\approx$ 150 MHz and $\approx$ 400 MHz), a command receiver with data decoder for the information provided by the control segment, and a memory element for the navigational data. The broadcast navigation information was transmitted as phase modulation on both frequencies. The OSCAR satellites turned out to be extremely reliable. OSCAR-13, launched in 1967, worked for nearly 22 years, before its power system failed.

The NOVA satellites are a new generation of TRANSIT satellites, stemming from the *TRANSIT Improvement Program* (TIP). The original idea was that the OSCAR
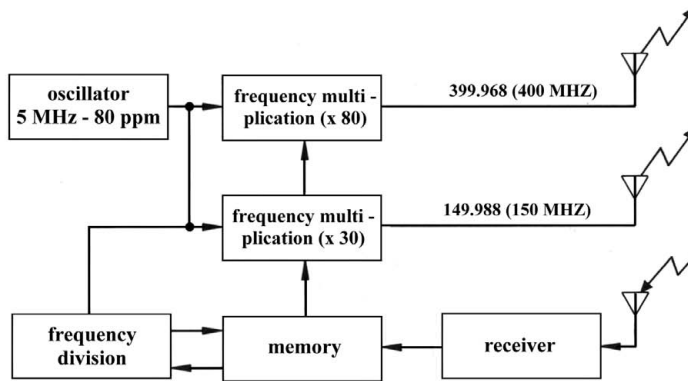
Figure 6.7. Functional diagram of OSCAR satellites

satellites should be replaced completely by the NOVA type spacecraft, in order to achieve an overall improvement of the TRANSIT system. Due to the concurrent development of the GPS system, and the decision to replace TRANSIT completely by GPS after 1994, only three NOVA satellites were built. One of the main new features of the NOVA satellites was the disturbance compensation system *DISCOS* [4.3.3.1] which compensates for the effects of atmospheric drag and radiation pressure. Predicted ephemerides are thus valid much longer.

As of January 1, 1997 after termination of the navigation service, several TRAN-SIT satellites remained operational and could be used for *ionospheric tomography*. The satellite system has been known since then as the *Navy Ionospheric Monitoring System* (NIMS). The message modulation, however, does not carry time and position information. The satellites are being used as dual-frequency beacons by ground data collection sites to determine the free electron profile of the ionosphere (Tucker, 1998).

6.2.2 Broadcast and Precise Ephemerides

Two kinds of orbital information have to be distinguished, the pre-determined (ex-trapolated) *broadcast ephemeris*, derived from observations made at four US tracking stations, and the post-processed (interpolated) *precise ephemeris*, derived from obser-vations made at about 20 globally distributed stations.

The broadcast ephemerides are communicated to the user as the navigation mes-sage. For a detailed explication see e.g. Seeber (1993, p. 167ff). An appropriately equipped observer is able to decode the orbit information, to convert the elements into Earth-fixed satellite coordinates, and to solve the observation equation (6.12) for the user position. Smoothing algorithms [3.3.3.2] can be used to interpolate orbit positions between the given epochs.

The broadcast ephemerides are generated from an orbit integration process [3.3.2.2], based on observations at the tracking stations of the *OP*erational *NET*work

(*OPNET*). In the case of TRANSIT these were the four stations at Maine, Minnesota, California, and Hawaii (Fig. 6.8).
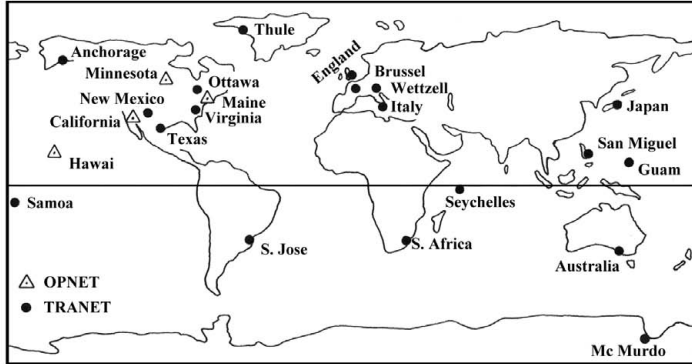
Figure 6.8. Control stations of the OPNET (triangles) and the TRANET (solid circles)

Precise orbit determination requires knowledge of the gravity field. The gravity field parameters were determined alongside the development of the TRANSIT system, and are partly based on observations with the early, experimental TRANSIT satellites. Fig. 6.9 shows the evolution of the orbit prediction accuracy over more than twenty years. A significant advance was achieved with the introduction of the gravity field
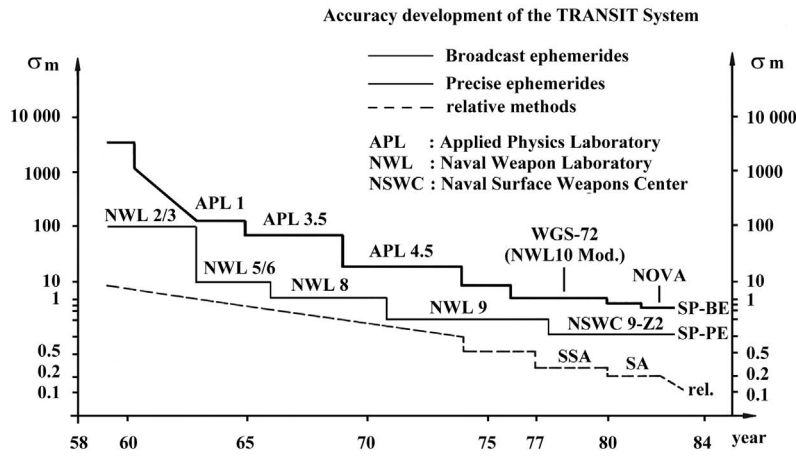
Figure 6.9. Evolution of the accuracy of the TRANSIT system

APL 4.5 (up to degree and order 15) in 1969. The gravity field of the *World Geodetic System 1972* (WGS 72), with potential coefficients up to (20,20), was introduced 1975, and the transition to the *WGS 84* occurred in 1988 [2.1.6]. Note that early GPS orbits were also based on WGS 72 and only since 1987 on WGS 84.

The main sources of error, inherent in the predicted ephemerides of LEO spacecrafts like the TRANSIT satellites, are not introduced by the gravity field model but by unmodeled along-track forces: *atmospheric drag* and *solar radiation pressure*. In particular, during periods of high solar activity the predicted orbit and the real orbit may show large differences. In order to improve the positioning accuracy, about 30 to 50 satellite passes had to be observed at each station. The accuracy of the geocentric coordinates, obtained with a broadcast ephemeris, was not better than $\pm$ 2 to 5 m, because the reference system, at that time, could not be realized to a better level of accuracy (cf. [12.1.1]) through observations over a limited time span.

Precise ephemerides were computed for some of the TRANSIT satellites based on observations at about 20 globally distributed stations. The accuracy was much higher than for the broadcast ephemerides, because the orbits were not predicted, but derived from measurements. The network of the participating tracking stations was operated by the Defence Mapping Agency (DMA); however civil organizations also participated in the observations. The network was called *TRANET* (*TRA*nsit *NET*work, see Fig. 6.8). With TRANET much experience was gained which was later incorporated into the organization and operation of a civil network for precise GPS orbits, namely CIGNET (Scheneweck, et al., 1990) and then IGS [7.8.1]. Since January 1987 the precise ephemerides have been based on the WGS 84 Earth model.

The accuracy of the precise ephemerides was about 1–2 meters. The ephemerides were delivered to the user as a set of 3D coordinates and velocities for each full minute. Coordinates for epochs in between could be determined with appropriate smoothing algorithms [3.3.3]. Geocentric coordinates, based on several days of observations and precise ephemerides, could be determined with an absolute accuracy of ± 0.5 to ± 1.0 m. Many geodetic control points all over the world, which are still in use today, have been determined with point positioning techniques and precise ephemerides [6.5.2], [6.6.1].

6.3   Doppler Receivers

6.3.1   Basic concept

The equipment consists of the antenna with preamplifier, and the receiver with data processor, data logger, power supply, and (optionally) sensors for obtaining meteorological data. Built-in microprocessors control the whole system, support the field planning, and provide a navigation solution, or even the complete data reduction, in the field.

Doppler receivers can be roughly subdivided into *single channel* and *dual channel receivers*. Single channel sets only track one frequency (e.g. 400 MHz). For applica-

tions in geodesy only dual frequency receivers are of interest, because of the need to correct for ionospheric refraction.

Fig. 6.10 shows the functional diagram of a generic geodetic Doppler survey set as used for TRANSIT observations. The satellite signal is received at the antenna
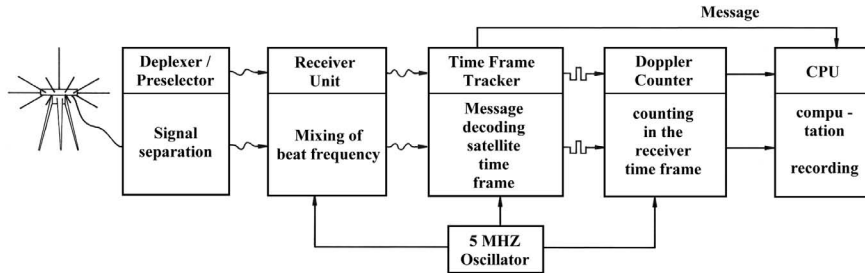
Figure 6.10. Functional diagram of a geodetic doppler receiver

and amplified in the preamplifier, which is usually integrated within the antenna body. For dual frequency receivers the incoming signal is then separated into the two (e.g. 150 MHz and the 400 MHz) signals in the *deplexer/preselector* module. Each signal is transferred to the *receiver module* and mixed with the reference oscillator frequency, generating the Doppler shifted beat frequency $f_b(t)$ (6.5) which reaches 32 KHz.

Both beat frequencies are passed to the *time frame tracker*. Here the satellite information (message) is decoded, and the satellite time frame is generated from the satellite time signals. The beat frequencies are digitized and transferred to the *Doppler counter*. The Doppler count can either be obtained with respect to the satellite time frame, given by the timing signals from the satellite, or with respect to the receiver time frame, generated by the internal 5 MHz oscillator. All internal functions and signal processing are controlled by the built-in *microprocessor* (central processing unit *CPU*).

The central module of the receiver is the 5 MHz *precision oscillator*. The ultimate accuracy depends essentially on the quality of the oscillator because instabilities during the satellite pass (16–18 minutes) affect directly the Doppler count [6.4.3]. Usually a precise quartz oscillator is used. In most cases an external oscillator (rubidium, cesium) can be connected to the receiver.

For post-processing purposes all data (Doppler counts, message, results, meteorological data, station identification data) are recorded with *data logging devices*.

The basic observable is the integrated Doppler count N_{jk} (6.5). The counting interval, defined by the satellite timing signals (e.g. two minutes for TRANSIT satellites) may be too long for geodetic applications. A higher resolution can be obtained from the signal structure of the broadcast message, e.g. 4.6 second word length in the TRANSIT message (*short Doppler count*). The improved resolution of a single wavelength of the Doppler frequency f_b is called the *fractional Doppler count*.

Two different modes are used to count the Doppler frequency $f_b(t)$ during the satellite pass (Fig. 6.11). In the first mode the counter is set to zero by each satellite
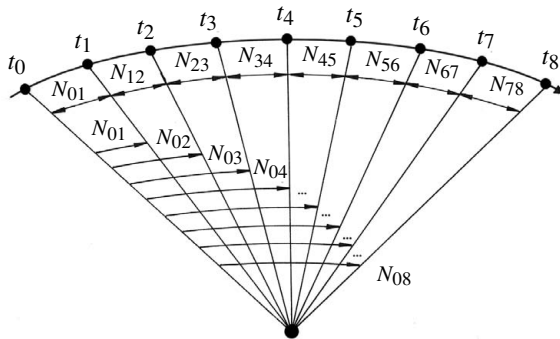
Figure 6.11. IID and CID Dopplercounts

timing signal. The procedure is named *intermittently integrated Doppler count* (IID). In the second mode the counter runs continuously, and is read out with each incoming timing signal. This is the *continuously integrated Doppler count* (CID). The two versions correspond to the differences between the terrestrial measurements of angles and directions.

One advantage of the Doppler method over range measurements is that it is far less sensitive to errors in the determination of time. It is the velocity of the transmitter in the satellite and not the velocity of the signals that counts. This results, for the TRANSIT system, in a reduction of the effect of timing errors by a factor of $\approx 5 \cdot 10^4$ (Hatch, 1982a).

6.3.2 Examples of Doppler Survey Sets

In 1967 *Magnavox* initiated the geodetic use of Doppler techniques with the launch of the *GEOCEIVER* AN/PRR-14. Since then the name *Geoceiver* has been synonymous with precise geodetic Doppler equipment. The instrument was mainly used by U.S. military and governmental agencies. Only precise ephemerides [6.2.2] were used, because the receiver was not equipped for decoding and recording the satellite message with broadcast ephemerides.

Later, Magnavox developed some commercial versions of the Geoceiver finishing, in 1977, with the *MX 1502 Satellite Surveyor*, Fig. 6.12, (Stansell, 1979). The instrument is quite compact, very rugged, battery operated, and has a weight of 19 kg. The satellites were tracked automatically. Real-time and accumulated single station solutions were displayed on the front panel. With additional boards the data from two simultaneously observing receivers could be processed in the field translocation solution, cf. [6.5.3].

Figure 6.12. Magnavox MX 1502
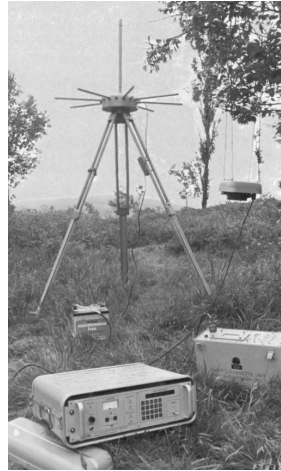
Figure 6.13. Marconi CMA 751

In 1977 *Canadian Marconi* delivered the microprocessor-based Doppler receiver set CMA 751. A pre-selection of satellite passes was possible, and also data reduction in the field. The tracking data were recorded on digital cassettes with a separate digital cassette unit. The atmospheric data could also be recorded on digital tape cassettes by the *environmental sensor unit* (ESU). Fig. 6.13 shows a typical field survey configuration.

Several other manufacturers built geodetic Doppler receivers between about 1973 and 1985. All receiver systems came with powerful software for coordinate transformation, prediction of satellite passes (*alerts*), single or multi-station solution. All these instruments had about the same accuracy and capacity, and they were able to meet all requirements until the TRANSIT-program was phased out in 1996. For DORIS receivers see [6.7].

6.4 Error Budget and Corrections

The basic observation equations (6.9) and (6.12) from section [6.1] are valid under ideal conditions. The real observation situation is different, mainly for the following reasons:

- − the predicted ephemerides differ from the true satellite positions [6.4.1],
- − the signal propagation is not in vacuum [6.4.2], and
- − the signal processing in the receiver electronics is not stable [6.4.3].

Furthermore, we have to consider that Earth, together with the receiver antenna, rotates while the signal is propagating (*aberration*), and that the transmitter and the receiver are moving relative to each other in differing gravity fields (*relativistic effects*) [6.4.4]. Finally, for non-stationary observers, the proper motion of the user antenna plays a role [6.4.5].

Some of these influences can be compensated through corrections to the observations. Some of the remaining influences can be modeled as parameters in the adjustment process. The residual unmodeled influences are regarded as observational or system noise; they contribute to the error budget of the individual satellite pass.

6.4.1   Satellite Orbits

Orbital accuracy is of importance when Doppler beacons on satellites are used in an operational satellite-based navigation system like TRANSIT. The situation is different for a system like DORIS [6.7], where the Doppler technique is primarily used for the orbit determination of spacecrafts. The predicted orbital positions of TRANSIT-type satellites (*broadcast ephemerides*) differ from the true satellite positions for three main reasons (cf. [6.2.2]):

   − uncertainty in the gravitational model,
   − limited accuracy of the orbit representation, and
   − inadequately modeled surface forces.

The relation between the degree and order of the Earth gravity field model used and the position error of a satellite at 1000 km orbital height is illustrated in Fig. 6.14. *Kaula's rule of thumb* shows that the position uncertainty amounts to 4–6 m for a potential development up to degree and order of 20. This corresponds to the effect of the inaccuracies of the WGS 72 Earth model on positions of a TRANSIT or typical Earth observation satellite, as is illustrated in Fig. 6.9. With the introduction of the WGS 84 gravity field, the effect has been reduced.

By far the largest portion of the orbit prediction error comes from unmodeled surface forces on the satellite like solar radiation pressure [3.2.3.4] and atmospheric drag [3.2.3.3]. The latter effect is particularly important for navigation and Earth observation satellites in low Earth orbits and with an unfavorable surface/mass ratio. The perturbations are greatest in the along track direction, and may reach several tens of meters. The effects can be much higher with high solar activity.

Figure 6.14. Kaula's "rule of thumb" for 1000 km orbital height

The situation is considerably better for satellites which have a compensation system for surface forces (DISCOS) [4.3.3.1], [6.2.1]. The remaining discrepancies from the predicted orbit, caused by unmodeled drag and radiation pressure, was only about

2–5 m for NOVA satellites (Eisner, et al., 1982). Another solution, as used for GPS satellites, is to develop a tailored solar radiation model (cf. [7.4.3.1])

Orbital errors can also be modeled as unknown parameters in the adjustment process [6.5.3].

6.4.2 Ionospheric and Tropospheric Refraction

The *ionosphere* is a dispersive medium for frequencies in the radio parts of the electromagnetic spectrum, i.e. the refraction index depends on the transmission frequency [2.3.3.1]. The first order effect on the Doppler-shift of a satellite signal can be eliminated by use of two phase-coherent signals on different frequencies, which are combined in the evaluation process.

Following Wells (1974) the influence of ionospheric refraction on the Doppler count is

$$IC = \Delta N(f) = N - N_b(f) = \frac{a_1}{f} + \frac{a_2}{f^2} + \frac{a_3}{f^3}. \tag{6.13}$$

N_b is the observed and N the corrected Doppler count. *IC* is the *ionospheric correction term*. The coefficients a_i are independent of the frequency, but they depend on the state of the atmosphere:

a_1/f is the first-order contribution of the ionospheric refraction,
a_2/f^2 is the second-order contribution; typically about 1% of the a_1 term,
a_3/f^3 is the third-order contribution.

The second-order term can be neglected. The third-order term equals, for the 150 MHz (e.g. TRANSIT) frequency, about 10% of the first-order term. The influence becomes smaller at higher frequencies. Because of this, in most cases only the first-order term of the ionospheric refraction is used for the evaluation of Doppler observations. When observations on both frequencies are available we find the following development for the corrected Doppler count N (cf. Egge, 1985). With

$$K = f_2/f_1 \tag{6.14}$$

it follows for (6.13) that

$$N - N_b(f_1) = a_1/f_1 \tag{6.15}$$

and

$$KN - N_b(f_2) = a_1/f_2, \tag{6.16}$$

or, because of (6.14)

$$KN - N_b(f_2) = a_1/(K f_1). \tag{6.17}$$

Dividing (6.15) by K, and then subtracting (6.17):

$$\frac{1}{K}N - \frac{1}{K}N_b(f_1) - KN + N_b(f_2) = 0. \tag{6.18}$$

It follows that

$$N = \frac{1}{(\frac{1}{K} - K)}\left(\frac{1}{K}N_b(f_1) - N_b(f_2)\right). \tag{6.19}$$

For the TRANSIT system, $f_1 = 400$ MHz and $f_2 = 150$ MHz, hence $K = 3/8$, and equation (6.19) gives for the corrected Doppler count N:

$$N = \frac{24}{55}\left(\frac{8}{3}N_b(400) - N_b(150)\right).$$
(6.20)

For the DORIS frequencies [6.7], $f_1 = 2000$ MHz and $f_2 = 400$ MHz, we find that

$$N = \frac{5}{24}\left(5N_b(2000) - N_b(400)\right).$$
(6.21)

Tropospheric refraction produces a signal propagation delay and thus causes an increase in the observed range. The total effect is subdivided into a dry and wet term (cf. (2.111)):

$$\Delta r = \Delta r_d + \Delta r_w.$$

Models for the computation of Δr are discussed in [2.3.3.2]. The Doppler count between two consecutive time marks t_j and t_{j+1} is influenced, according to

$$TC = \frac{1}{\lambda_s}(\Delta r_{j+1} - \Delta r_j),$$
(6.22)

where λ_s is the transmitted wavelength. The amount of tropospheric refraction depends on meteorological conditions and on the path length through the troposphere. Hence, low elevation satellite passes are much more affected than passes at high elevation angles. If passes with low elevations $5° < E < 10°$ have to be used adequate refraction models are required [2.3.3.2]. In most cases the use of "standard meteorological data" in the reduction models is completely sufficient.

6.4.3 Receiver System

Signal processing within the electronic parts of the receiver is mainly affected by three error sources:

- variation of the antenna phase center,
- variation of the signal propagation delay (time jitter), and
- instability of the oscillator.

The reference mark for Doppler observations is a point in the vertical axis of the antenna which represents a mean position of the phase center and which is usually defined by the manufacturer. The actual position of the phase center is not stable during the observation period. It depends on the satellite elevation, the signal strength and the multipathing in the vicinity of the antenna.

The variations occur mainly in the vertical direction. They can be averaged out through a large number of observed satellite passes. The noise can be reduced, when the antenna is shielded against reflections and multipath with a large horizontal ground plane (Seeber, Egge, 1981).

For receivers that time tag the Doppler counts with respect to the receiver time frame, the differing signal propagation time within the electronic parts is of importance. The *receiver delay* ε_j is not necessarily constant. If ε_0 is the receiver delay at the beginning of a Doppler count and ε_1 at its end, the mean delay is

$$\varepsilon_m = \frac{\varepsilon_0 + \varepsilon_1}{2} \tag{6.23}$$

and the effect on the Doppler count (Egge, 1985) is

$$DC = \frac{\dot{r}_1 - \dot{r}_0}{\lambda_s} \varepsilon_m. \tag{6.24}$$

Here $\dot{r}_0$ and $\dot{r}_1$ are the respective relative velocities between the satellite and the receiver. The mean delay ε_m can be introduced as a parameter into the observation equation (cf. [6.5.1]).

The *oscillator quality* plays a key role, because instabilities in the period of the satellite pass (about 18 minutes for TRANSIT satellites) have a direct influence on the Doppler count. For high accuracy solutions the frequency drift can be modeled with a particular parameter $\dot{\Delta}F$ in the observation equation. Also, an external oscillator can be used for reduced observation noise. For DORIS ultrastable oscillators are used in the satellite package and in the ground beacons [6.7].

6.4.4 Earth Rotation and Relativistic Effects

In general, the integrated Doppler count is a "four points function", i.e. it depends on the position vectors

$$\begin{aligned} & \boldsymbol{r}_S(t_j), \boldsymbol{r}_S(t_k) && \text{of the satellite antenna, and} \\ & \boldsymbol{r}_i(T_j), \boldsymbol{r}_i(T_k) && \text{of the receiver antenna.} \end{aligned}$$

With respect to an Earth-fixed reference frame and with a stationary receiver antenna on the ground, the position vector of the antenna is time invariant, so

$$\boldsymbol{r}_i(T_j) = \boldsymbol{r}_i(T_k) = \text{ constant.} \tag{6.25}$$

The Doppler count is hence a "three point function" (cf. Fig. 6.5), and the adjustment model simplifies. Furthermore, the effect of Earth rotation during the signal travel time $\Delta t = r/c$ has to be considered, because the slant range r between the satellite and the receiver changes with the *aberration effect* (Egge, 1985):

$$\delta_S = \frac{1}{c}(\boldsymbol{\omega}_e \times \boldsymbol{r}_S)(\boldsymbol{r}_S - \boldsymbol{r}_i), \tag{6.26}$$

where $\boldsymbol{\omega}_e = (0\ 0\ \omega_e)^T$ is the Earth rotation vector.

$$\boldsymbol{r}_S = (X_S, Y_S, Z_S)^T, \quad \boldsymbol{r}_i = (X_i, Y_i, Z_i)^T$$

are the position vectors $\boldsymbol{r}_S$ of the satellite position S_j (and also S_k), and $\boldsymbol{r}_i$ of the receiver antenna P_i, referred to the geocentric Earth-fixed Cartesian system. Applying a vector operation, we obtain the changes in the slant ranges r_{ij} (or r_{ik}) as

$$\delta r_{ij} = \frac{\omega_e}{c}(X_{S_j}(Y_{S_j} - Y_i) - Y_{S_j}(X_{S_j} - X_i)). \tag{6.27}$$

The corresponding *aberration correction* of the Doppler count is then

$$AC = \frac{1}{\lambda_S}(\delta r_{ik} - \delta r_{ij}). \tag{6.28}$$

The range correction (6.27) amounts to only a few cm.

  Relativistic effects have two sources (Schlüter, Pesec, 1982):
  − the transmitted frequency is observed to be lower due to the relative motion (*special relativity*), and
  − the transmitter operates in a field of different gravitational potential; the Earth-bound observer receives a higher frequency (*general relativity*).

Summing both influences we obtain the *relativistic frequency correction*:

$$\Delta f_{s_{\mathrm{rel}}} = f_s \left(\frac{1}{2}\frac{v_S^2}{c^2} + \frac{GM}{c^2}\left(\frac{1}{|\boldsymbol{r}_S|} - \frac{1}{|\boldsymbol{r}_i|}\right)\right); \tag{6.29}$$

v_s is the relative velocity of the transmitter with respect to the receiver antenna. The relativistic effect on the integrated Doppler count is (Egge, 1985):

$$RC = \left(-GM\left(\frac{1}{|\boldsymbol{r}_i|} - \frac{1}{|\boldsymbol{r}_S|}\right) - \frac{|\dot{\boldsymbol{r}}_S|^2 - |\dot{\boldsymbol{r}}_i|^2}{2}\right)\frac{t_k - t_j}{c\,\lambda}. \tag{6.30}$$

The influence on the range difference, caused by the relativistic effect, is for satellites of TRANSIT type and short Doppler counts (e.g. 4.6 s):

$$\Delta r_{\mathrm{rel}} = 0.28 \text{ m},$$

and amounts to about 21 m after 6 minutes (Schlüter, Pesec, 1982). The relativistic effect is nearly completely absorbed by the parameter $\Delta F = (f_g - f_s)$ in the basic adjustment model.

6.4.5 Motion of the Receiver Antenna

The integrated Doppler count can only be treated as a "three points function" when the user antenna is Earth-fixed. This precondition is not valid for navigational applications, or on moving surfaces like buoys or ice sheets. In such cases the observed integrated Doppler count also contains influences from the motion of the receiver antenna. The measured Doppler count can be corrected for motion, when the relative changes of the antenna position with respect to the ground are known from independent sources, as is the case with *Doppler-Sonar* techniques in marine applications.

When insufficient information on the antenna motion during a satellite pass is available the observed integrated Doppler count contains systematic effects that lead to systematic position errors. Fig. 6.15 demonstrates for TRANSIT satellites how an error of 1 m/minute in the observer's velocity propagates into the position. It becomes clear that, without velocity information, considerable position errors may be present in the navigation solution.
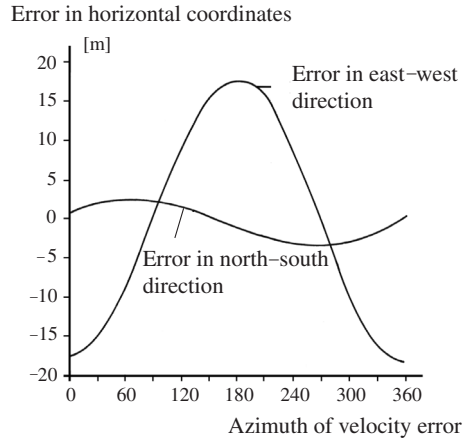
Figure 6.15. Effect of a velocity error (1m/min) on the single pass position fix (TRANSIT)

6.5   Observation Strategies and Adjustment Models

6.5.1   Extended Observation Equation

A simplified observation equation has been already introduced in (6.9), according to Fig. 6.5 (p. 185):

$$N_{jk} = (f_g - f_s)(t_k - t_j) + \frac{f_g}{c}(r_{ik} - r_{ij}).$$

This basic equation is mostly used in navigation. It is very suitable for application to simple receivers because the timing interval $(t_k - t_j)$ for the Doppler count is provided from the space segment, and no time measurement in the receiver is necessary. The geodetic form of the observation equation requires some refinement and extension, based on the discussions in sections [6.4] and [6.5].

In particular, a much more stable time measurement is required of the internal oscillator used in a geodetic receiver. The transition to the receiver time frame is imperative, and a variable receiver-delay has to be assumed for the beginning and the

end of the Doppler count. Equation (6.9) thus expands to (Egge, 1985)

$$N_{jk} = (f_g - f_s)(t_k - t_j) + \frac{f_g}{c}(r_{ik} - r_{ij}) + \left((f_g - f_s) + \frac{\dot{r}_m}{\lambda_s}\right)\Delta\varepsilon + \frac{\dot{r}_{ik} - \dot{r}_{ij}}{2}\varepsilon_m,$$

(6.31)

where

$$\begin{aligned} \Delta\varepsilon &= \varepsilon_k - \varepsilon_j & \text{the delay difference,} \\ \varepsilon_m &= \frac{\varepsilon_k + \varepsilon_j}{2} & \text{the mean receiver delay, and} \\ \dot{r}_m &= \frac{\dot{r}_{ij} + \dot{r}_{ik}}{2} & \text{the mean relative velocity,} \end{aligned}$$

(6.32)

during the observed interval of integration. $\lambda_s = c/f_s$ is the wavelength of the carrier frequency.

The second term in (6.31) can be replaced with $\Delta F = (f_g - f_s)$ as

$$\frac{1}{\lambda_s}(r_{ik} - r_{ij}) + \frac{\Delta F}{c}(r_{ik} - r_{ij}).$$

(6.33)

With the substitution $\Delta T_s = (t_k - t_j)$ equation (6.31) can be developed into a well structured form. The observation equation (6.31) is now formulated as a function of the wavelength λ_s that is determined by the satellite control segment

$$N_{jk} = \Delta F \Delta T_s + \frac{1}{\lambda_s}(r_{ik} - r_{ij}) + FC + PC + DC.$$

(6.34)

Here we find that the *frequency correction*,

$$FC = \frac{\Delta F}{c}(r_{ik} - r_{ij}),$$

(6.35)

only depends on the frequency difference between the transmitter and the receiver. Further, the *partial cycle correction*,

$$PC = \left(\Delta F + \frac{\dot{r}_m}{\lambda_s}\right)\Delta\varepsilon,$$

(6.36)

gives the relation between the transmitter time frame and the receiver time frame, and the *delay correction* (cf. (6.24)),

$$DC = \frac{\dot{r}_{ik} - \dot{r}_{ij}}{\lambda_s}\varepsilon_m,$$

(6.37)

describes the influence of the receiver delay on the Doppler count. We add the previously introduced corrections

(6.28) aberration correction AC,
(6.30) relativistic correction RC,
(6.13) ionospheric correction IC,
(6.22) tropospheric correction TC,

and obtain the extended observation equation (Egge, 1985)

$$N_{jk} = \Delta F \Delta T_s + \frac{1}{\lambda_s}(r_{ik} - r_{ij}) + FC + PC + DC + IC + TC + AC + RC. \quad (6.38)$$

Rearrangement provides (for example, for $f_1 = 400$ MHz)

$$
\begin{aligned}
N_{jk} &= N_b(400) - IC - TC - FC - PC - AC - RC \\
&= \Delta F \Delta T_s + \frac{1}{\lambda_s}(r_{ik} - r_{ij}) + \frac{1}{\lambda_s}(\dot{r}_{ik} - \dot{r}_{ij})\varepsilon_m.
\end{aligned}
\quad (6.39)
$$

The first line of equation (6.39) contains all those corrections that generate the *ideal Doppler count* N_{jk} from the observed Doppler count N_b (400); the ideal Doppler count satisfies the functional model of the second line. Finally the observation noise v_{jk} has to be added. The corrections are computed from a priori knowledge, or they are directly measured in the receiver. Remaining components can be added to the observation equation as model parameters; this holds, for example, for the tropospheric scale factor k which is separately estimated for each satellite pass.

Several alternative formulations of the observation equation are in use. The Doppler-count can also be regarded as a *pseudorange*, i.e. a range measurement with a constant bias, instead of a range difference. This model corresponds to the continuously integrated Doppler count (CID) [6.3.1]. It was proposed very early on by Brown (1970) as a precise adjustment model.

In order to extract the information from the Doppler observations in the best possible way, particular observation and adjustment techniques have been developed. A first subdivision is into

    − single station techniques, and
    − multi-station techniques.

The multi-station techniques can further be subdivided into

    − translocation methods,
    − semi-short-arc methods, and
    − short-arc methods.

The single station methods are also named *absolute position determination* or *point positioning*; the multistation techniques are referred to as *relative position determination*. The related absolute accuracy is in general much less than the relative accuracy (cf. [12.1.1]). Note that this distinction is also used for GPS positioning [7].

6.5.2 Single Station Positioning

This is the classical method of Doppler point positioning. The satellite coordinates are usually held fixed and are assumed known parameters. Geocentric coordinates are determined, referred to the reference frame of the particular satellite

ephemeris. Starting with the approx-
imate position x^0, a correction vector
dx is derived from the adjustment of
all observations (Fig. 6.16). A certain
number of satellite passes, depending on
the particular satellite system, has to be
observed in order to achieve a reason-
able accuracy. For TRANSIT and broad-
cast ephemerides about 30 to 50 passes
were required for an absolute accuracy
of 2–3 m with respect to WGS 84. For
DORIS see [6.7].

With precise ephemerides the point
positioning result may reach an accuracy
of ±1 m, or better. In several countries
the coordinates from TRANSIT point
positioning with precise ephemerides
still form the basis of the national net-
work of control points.
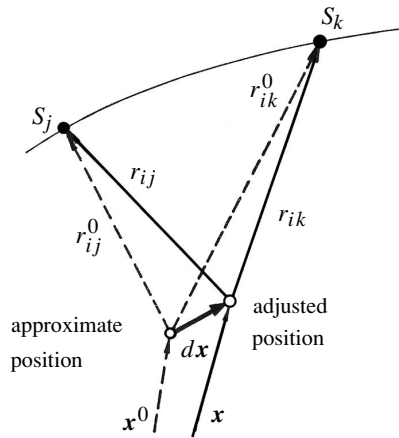
Figure 6.16. Single station determination (point positioning)

6.5.3 Multi-Station Positioning

Some of the biases are identical for nearby stations, and cancel when differences are
formed between the observations. This is why simultaneous observations at two or
more stations is the standard procedure in Doppler surveying. In the case of two stations
the technique is named *translocation*. Note that experiences with the translocation and
multi-station techniques in TRANSIT data analysis have considerably influenced the
DGPS technology [7.5].

In principle, in the translocation
technique, some of the parameters are
eliminated instead of being estimated
(cf. Fig. 6.17). This is true for the sys-
tematic orbit errors and, to some extent,
for the propagation errors. The increase
in accuracy obviously only applies to the
relative coordinates between stations A
and B, not to the absolute coordinates of
both stations. This is why translocation
techniques are suitable when a new sta-
tion B is connected to an existing point
A. The same principle is valid for GPS
[7.5].

The parameters for orbit errors and
propagation delay can also be estimated,
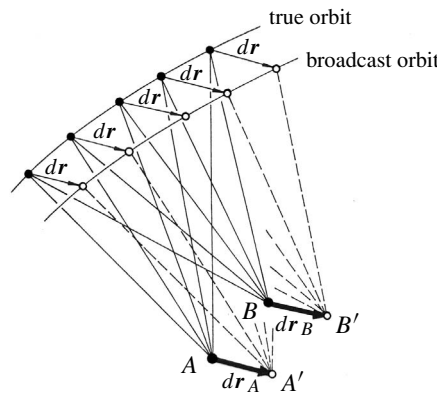
Figure 6.17. Translocation technique; orbit er-
rors are cancelled in the difference

instead of being eliminated. This procedure introduces greater flexibility into the adjustment model and provides a better adaptation of the observations to the real situation. However, simultaneous observations from more than two stations are required.

The basic model is as follows. The observed portion of the orbital arc is given three degrees of freedom within certain limits; the broadcast orbit arc is thus only translated in space, with no change in orientation or shape (Fig. 6.18). This method is sometimes referred to as the *relaxed orbit technique*.

Another possibility is to allow for corrections to the six Kepler parameters, or for 3 translations and 3 velocity components in the broadcasted satellite positions. Such models are named *semi-short-arc methods*. They must be distinguished from the *short-arc methods*, where a small portion of the orbit is completely re-computed within the adjustment process. The broadcast ephemerides are only used for generation of the start values. A new orbital
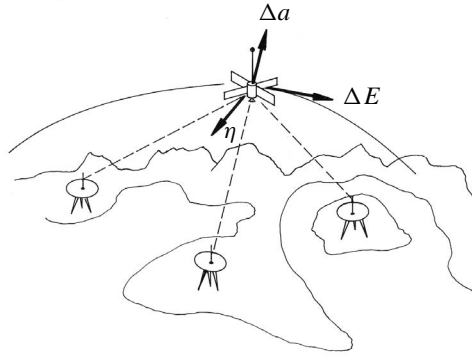
Figure 6.18. Multistation technique; the orbit is given three degrees of freedom

arc is determined from numerical integration, based on a gravity field development and all available observations from the participating stations.

For TRANSIT a relative accuracy of ±0.2 m to ±0.5 m was achievable with multi-station techniques when the observation and adjustment process was carefully controlled. The best results, documented for smaller networks, were ±0.1 m to ±0.2 m (Schenke, 1984). With higher frequencies and advanced technology (e.g. with DORIS [6.7]) sub-decimeter accuracy can be reached.

6.6 Applications

In the following a short review is given of the most important applications of satellite Doppler technology during the TRANSIT era. During this era most of the concepts of modern GPS [7] and satellite Doppler technology (e.g. DORIS [6.7]) were founded.

Satellite Doppler techniques found broad applications in geodetic work between about 1970 and 1990. In many countries the Doppler technique at that time was considered to be one of the most important tools for geodetic control surveys, in particular those with an insufficient cartographic infrastructure. Some of the main fields of applications were (see also [12]):
- geodetic control,
- control points for photogrammetry, hydrography, geophysics,
- marine and polar geodesy, and
- polar motion.

6.6.1   Applications for Geodetic Control

Several objectives can be distinguished:

(1) establishment of fundamental geodetic control,

(2) densification of existing geodetic control,

(3) analysis and improvement of existing geodetic control, and

(4) contribution to geoid determination.

(1) The first objective included the *establishment and realization of a geodetic datum* and the installation of a *basic control network*. Usually the World Geodetic System WGS 84 was selected as the geodetic datum. The accuracy of the realization was ±0.5 m with precise ephemerides and ±1 to 2 m with broadcast ephemerides. The relative accuracy between datum points was, however, much better and reached ±0.2 m with 50 to 100 accepted satellite passes.

(2) The *densification and expansion* of an existing network was achieved with different procedures. The easiest method, from the logistical point of view, was the establishment of single stations, each with a precise ephemeris. This procedure was widely used in the geodetic development of vast areas like Brazilian Amazonia (Fig. 6.19). The coordinates were first determined in the satellite datum, and then
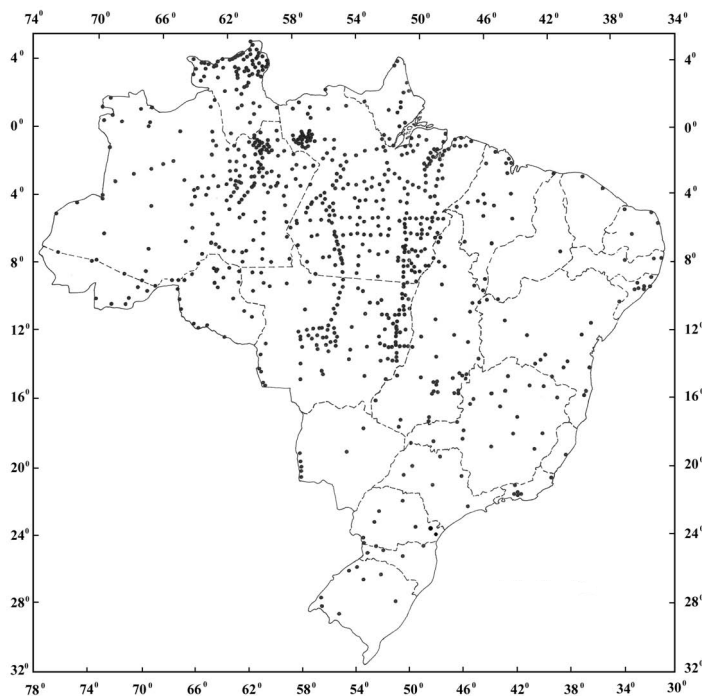
Figure 6.19. Doppler control points in Brazil (source: IBGE 1987)

converted with conventional transformation formulas to the particular national datum, e.g. the South/North American Datum SAD 69 or NAD 83. Accuracy was ±0.5 m to ±1.0 m. Those coordinates define and realize the geodetic datum in the newly developed region.

(3) Doppler observations were also used to *control, analyze and improve* existing geodetic networks. Examples in Europe from the TRANSIT era are the

   EDOC-1,2   *European Doppler Campaign* (Boucher et al., 1979)
   DÖDOC   *German-Austrian Doppler Campaign* (Rinner et al., 1982)
   RETDOC   *European Triang. Network Doppler Campaign* (Wolf, 1984).

In Africa ADOS (*African Doppler Survey*) was organized in international cooperation (Chodota, 1987, 1990). More than 300 stations in 46 African countries were included in the final computation. Examples from smaller areas are the German Fundamental Network (DHDN) (Seeber, Seeger, 1984), one part of the Venezuelan network (Hoyer, 1982), and the fundamental network in Southern Brazil (Campos, 1987).

The German–Austrian Doppler Campaign (DÖDOC) resulted in positions with standard deviations of about 15 cm and showed discrepancies to the existing classical network at the order of about 1 m. This example demonstrated that Doppler observations over distances of several 100 km are able to meet or even exceed the accuracy standard of classical triangulation networks.

(4) Doppler satellite observations provide three-dimensional Cartesian coordinates that can be converted to ellipsoidal latitude, ellipsoidal longitude, and *ellipsoidal height h*. With known levelled orthometric height H the *geoid undulation N* can be determined for the observation point using the simple relation (cf. Fig. 7.82, p. 366):

$$N = h - H. \tag{6.40}$$

This procedure has been used in many countries for the determination of regional geoids. One example is the project ALGEDOP (1980–1983) for the European Alps (Seeger, 1984). Comparisons with existing geoid computations showed differences of less than 1 to 2 m. Another example is the Doppler geoid in Africa based on ADOS results (Fashir, Abdalla, 1991). The same relationship, however with much higher accuracy, is used in GPS geodesy for regional and local geoid determination [7.6.2.3].

6.6.2 Further Applications

Satellite Doppler surveying has been widely used to determine control points in various applications besides navigation. In the following some typical examples from the TRANSIT era are presented. The same types of application are possible with the DORIS system [6.7], although in most cases today the GPS technology is preferred.

Control and fiducial points
Fiducial points in *small scale photogrammetry* or for *geophysical surveys*, (e.g. for the determination of seismic shot points) have been frequently established in the single

station or the translocation mode with few satellite passes. The translocation technique could be used to reduce the necessary observation time.

For *hydrographic surveys* with conventional short-range radio navigation (e.g. Syledis, Hifix, Trisponder, Miniranger) the coordinates of the shore-based antennas often were determined with Doppler translocation techniques. An accuracy of ± 1 m to ± 3 m was usually sufficient.

Another important application was *inertial surveying* (Seeber, 1979; Schwarz, 1980, 1983). In a first step, control points with inter-station distances of 100 to 150 km were determined with Doppler techniques. The densification was done, in the second step, with inertial surveys. This technique has been widely used in Canada for the determination of control points in unsurveyed areas (Webb, Penney, 1981). Multi-station Doppler techniques were required because inertial surveying determines the coordinates with a relative accuracy of ± 0.1 m to ± 0.2 m.

Marine and polar geodesy

The increasing economic importance of the seas and the sea floor caused increasing accuracy demands. The demarcation of marine boundaries and the location of drilling positions for the exploitation of marine hydrocarbons and ocean mining require a position accuracy of as high as ± 1 m with respect to a global datum. *Fixed* structures in the marine environment, like oil production platforms, can be surveyed with the same geodetic methods used for land-based objects. This is why Doppler techniques were extensively used since about 1970 for positioning marine platforms, e.g. in the North Sea (Leppard, 1980), cf. Fig. 12.9, p. 523.

The situation is more difficult for *moving objects* (ships, buoys) because the relative motion of the object during the satellite pass decreases the achievable accuracy (Egge, 1982; Seeber, Egge, 1981), cf. [6.4.5]. With translocation measurements to a fixed reference station the best possible accuracy was found to be ± 10 m (Seeber, 1983). TRANSIT Doppler receivers were an essential component of all integrated navigation systems.

Doppler observations were applied successfully to determination the *velocity of ice sheets* in the Arctic or Antarctic (Drew, 1983; Seeber, Hinze, 1984). In the Antarctic, the rates of shelf-ice motion are between about 3 m/day near the coast to about 0.1 m/day for the inland ice (Hinze, 1990).

Polar motion

Satellite orbits are computed with respect to an inertial reference frame; the Doppler observations are related to an Earth-fixed reference frame. The difference between the two systems contains among other effects the influence of polar motion [2.1.2.3]. In order to use the satellite observations in orbit computation they must either be corrected for polar motion, or the components x, y of polar motion have to be estimated as additional parameters in the orbit adjustment program. This second solution has been used successfully since 1969 (Anderle, 1973, 1986) with the computation of the precise ephemeris of the TRANSIT satellites, based on the observations within the global TRANET network [6.2.2].

Doppler-derived polar motion data have been published weekly by the U.S. Naval Observatory and were used in the regular polar motion service of the former *Bureau International de l'Heure* (BIH) until 1987. Comparisons over many years demonstrated a good agreement with other techniques [8.5.5], [12.4.2]. The Doppler results were two to four times more accurate than classical astrometric methods. Because of the even better accuracy with SLR and VLBI techniques the Doppler polar motion data derived from TRANSIT observations were not used by the IERS service [12.4.2], that superseded BIH on January 1, 1988. Since 1994, however, DORIS observations have been included in the computation of Earth orientation parameters and other IERS products [6.7].

6.7   DORIS

DORIS (*D*oppler *O*rbitography and *R*adio Positioning *I*ntegrated by *S*atellite) is a French development which, like TRANSIT, uses the Doppler concept, but in a reverse mode: a stable frequency is emitted by ground beacons and the measurement of the Doppler count is made in the satellite.

The DORIS system was developed by the French Space Agency CNES (Centre National d'Études Spatiales) in cooperation with IGN (Institut Géographique National) and GRGS (Groupe de Recherches de Géodésie Spatiale) with the objective to support precise orbit determination for low-altitude Earth satellites. A first realization of the system was in 1990 on the remote sensing satellite mission SPOT-2.

The system is based on the measurement of Doppler shifts in radio signals, transmitted by ground beacons and received by the DORIS onboard package as the satellite passes overhead. The ground beacons broadcast continuously and omnidirectionally at frequencies of 2036.25 MHz and 401.25 MHz (cf. Fu et al., 2001, p. 75f.). A receiver onboard the satellite receives the signal and measures the Doppler shift over a short count interval, e.g. 10 seconds. The data are time tagged with respect to an ultra-stable onboard crystal oscillator. The precise Doppler measurement is made on the 2 GHz signal; the use of the second frequency allows for the removal of the effects of ionospheric refraction [6.4.2]. The ground beacons are also equipped with stable oscillators and sensors to provide in situ meteorological data. The average precision of the range rate observations is about 0.3 to 0.5 mm/s. The DORIS system comprises:

- the onboard package, consisting of a receiver for both frequencies (total mass 17 kg), an ultrastable crystal oscillator (stability class $5 \cdot 10^{-13}$), and an omnidirectional antenna,

- a global network of permanent ground beacons (Fig. 6.20), at about 50 sites, each consisting of two transmitters, an ultrastable oscillator, microprocessor, power supply, meteorological sensors, and an antenna,

- dedicated location beacons, functionally identical to the permanent beacons, and

- a control segment (master beacon and control center in Toulouse, France).
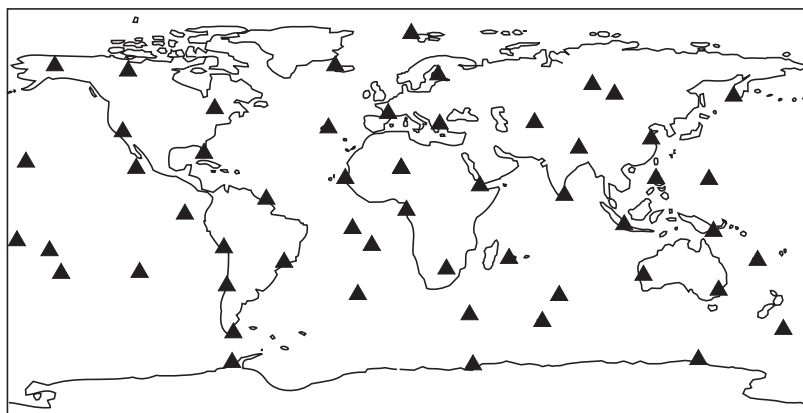
Figure 6.20. DORIS orbitography beacon network (status 2002)

The network of beacons is distributed evenly around the globe. This is a key factor in the high quality of DORIS products. The network also provides very good coverage of the oceans since nearly half of the stations are on islands. All data are collected by the onboard receiver on a time-sharing basis, stored in the onboard memory, and downloaded via suitable data links at regular intervals to the mission control center in Toulouse. The DORIS onboard package so far has been installed on seven satellites, and it will certainly be flown on more.

Table 6.1. Satellites carrying DORIS payload

Satellite	Launch date	End of mission
SPOT-2	January 22, 1990	
TOPEX/POSEIDON	August 10, 1992	
SPOT-3	September 26, 1993	November 14, 1996
SPOT-4	May 26, 1998	
JASON	December 7, 2001	
ENVISAT	March 1, 2002	
SPOT-5	May 5, 2002	

DORIS is primarily an orbit determination system, however it also contributes to studies of other geodetic and geodynamic problems (Lefebvre et al., 1995). In the following a short overview on DORIS applications is given.
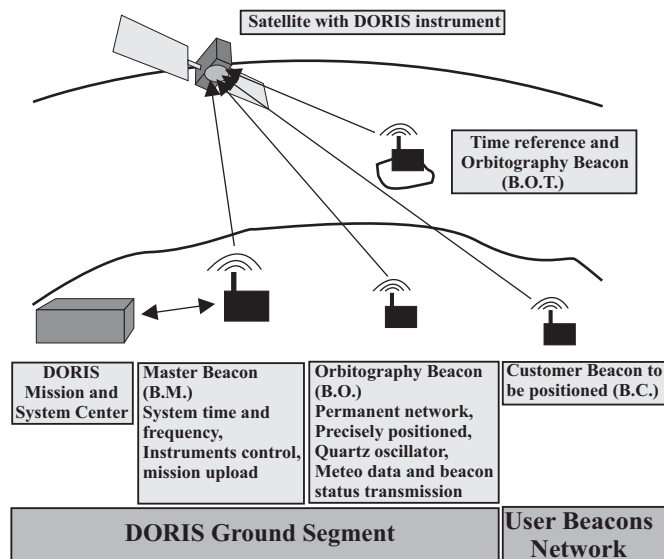
Figure 6.21. DORIS System Overview

Orbit determination

The original design accuracy for orbit determination was about 10 cm for the radial component in post-processing mode after about 1 month. With improved network configuration, better theory, and error modeling, the current accuracy is about 2.5 cm for post-processed data. For JASON the aim is to measure the satellite's altitude to within 1 cm. This high performance will contribute considerably to the use of altimetry in geodesy and oceanography [9.5], (Fu et al., 2001).

With SPOT-4 in 1998 the real-time autonomous on-board orbit determination capability DIODE was established, providing real-time orbits with an accuracy within a few meters. DIODE stands for *Détermination Immédiate d'Orbite par Doris Embarqué* (immediate onboard orbit determination by DORIS). One important objective of the DIODE navigator is to deliver in real-time to SPOT image users all necessary information for a rectification of the SPOT scenes.

A further improvement is expected with a second generation receiver onboard JASON, ENVISAT, and SPOT-5, that has reduced size and weight, and is capable of communicating with two beacons at the same time. For more information see e.g. Agnieray (1997); Costes (1997); Jayles et al. (2000). DORIS offers three different orbit products:

   − *Real-time*, with an accuracy of several meters, and sub-meter accuracy in the future with the new generation system,

   − *Operational*, with submeter accuracy after 48 hours ($<$ 20 cm on the radial component), and

   − *Precise*, with sub-decimeter accuracy after one month (cm for radial component).

The continuous evolution of the system promises improved products in the future.

Positioning
Once the satellite trajectory is known, the exact position of a DORIS station anywhere in the world can be calculated. In practice, the more satellite passes are used the better is the positioning accuracy. There are two services. In operational geodesy with *dedicated location beacons*, any point on Earth at any time can be determined with about 20 cm accuracy after a one-day measurement time, and 10 cm after 5 days. This commercial service, offered by a CNES subsidiary in Toulouse, may be of interest for geophysical exploration or similar work. The beacons have their own power supply and only transmit when a satellite is within view. They operate unattended for several months. The automatic monitoring of the beacons make the system well suited for use in high risk areas (e.g. land slide, volcanic activity).

The *permanent beacon network* delivers high precision 3D coordinates for geodetic and geodynamic applications. Positions and motions are available to better than 1 cm and 1 mm/year, respectively. Due to the dense and homogeneous global beacon network DORIS significantly contributes to the realization and maintenance of the ITRS [2.1.2.2]. Around 30 of the more than 50 beacons are collocated with other space techniques (SLR, VLBI, GPS) (Fagard, Orsoni, 2000), i.e. they can be included in combined ITRF computations, and provide important information on the stability of the individual solutions. Tectonic plate deformations derived from permanent DORIS observations correspond very well with the geological NUVEL-1 model (cf. [12.4.1]).

Geodesy, Geodynamics, and related fields
Further contributions to geodesy and geodynamics are only mentioned:
  − Gravity field improvement,
  − Motion of the geocenter,
  − Vertical displacement near tide gauges,
  − Polar motion, Earth rotation,
  − Ionospheric studies, TEC models, and
  − Atmospheric drag.

For more information see the proceedings of the regularly organized *DORIS Days*, e.g. CNES (2000), or documents of the future (status 2002) *International DORIS Service* (IDS). Because of the increasing importance of DORIS data and products for the geodetic community and related disciplines the IUGG in 1999 decided to start the *DORIS Pilot Experiment*, with the objective to assess the need and feasibility of an International DORIS Service. The IDS will be an international scientific service under the auspices of the IUGG, similar the IGS (International GPS Service), ILRS (International Laser Ranging Service) and IVS (International VLBI Service), (Tavernier et al., 2000).

After more than 10 years of operation DORIS has developed into one of the key technologies in geodetic space techniques.

7 The Global Positioning System (GPS)

7.1 Fundamentals

7.1.1 Introduction

The NAVSTAR GPS (*NAV*igation *S*ystem with *T*ime *A*nd *R*anging *G*lobal *P*ositioning *S*ystem) is a satellite-based radio navigation system providing precise three-dimensional position, navigation, and time information to suitably equipped users. The system is continuously available on a world-wide basis, and is independent of meteorological conditions. GPS has been under development in the U.S.A. since 1973, and is primarily a military system, with limited access to civil users [7.1.6]. It has been used for the solution of geodetic problems since about 1983. In its final configuration, available since 1995, the system nominally consists of 24 satellites placed in orbits of about 20 200 km altitude above the Earth's surface (Fig. 7.1). The arrangement of satellites has been planned in such a way that at least four satellites are simultaneously visible above the horizon, anywhere on Earth, 24 hours a day.
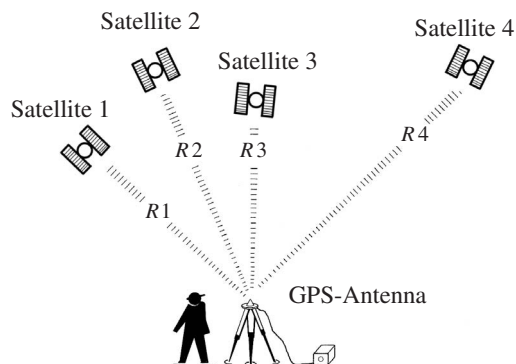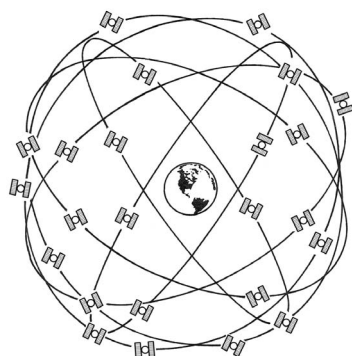
Figure 7.1. The Global Positioning System (GPS), 24 satellites configuration

Figure 7.2. Basic principle of positioning with GPS

GPS is primarily a navigation system. The fundamental navigation principle is based on the measurement of so-called *pseudoranges* [7.1.4], [7.3.1] between the user and four satellites (Fig. 7.2). Starting from the known satellite coordinates in a suitable reference frame the coordinates of the user antenna can be determined. From the geometrical point of view three range measurements are sufficient. A fourth observation is necessary because GPS uses the one-way ranging technique [4.2.2], [7.1.4], and the receiver clock is not synchronized with the satellite clock. This synchronization error is the reason for the term "pseudorange".

Unlike the NNSS TRANSIT system [6.2], GPS continuously provides navigation data in real-time on a global basis. Technological advances over about twenty years also mean that a much higher accuracy is achieved than that for TRANSIT. Some characteristic features of NNSS and GPS are compared in Table 7.1.

Table 7.1. Characteristics of GPS and TRANSIT

Features	GPS	NNSS (TRANSIT)
orbital height	20 200 km	1000 km
period	12 h	105 min
frequencies	1575 MHz	150 MHz
	1228 MHz	400 MHz
navigation data	4D: X, Y, Z, t; velocity	2D: φ, λ
availability	continuously	$\approx$ 15–20 min per pass
accuracy	15 m	30–40 m depending on
	0.1 m/s	velocity error
constellation	21–24 SV	4–6 SV
geometry	repeating	variable
satellite clocks	rubidium, cesium	quartz

GPS has been designed to provide, at best and for authorized users, a real-time navigation accuracy of ± 10 m to ± 15 m. It was recognized early on, however, that GPS can also support geodetic positioning with great accuracy. Anderle (1979) predicted, during a symposium on satellite geodesy in Athens, that a relative accuracy of ± 10 cm over a distance of 2000 km would be attainable. The experience to date has proved that a broad variety of problems in geodesy and geodynamics find their solution with GPS with even higher accuracy. Fig. 7.3, from the early eighties, demonstrates that the advent of GPS filled a gap in capabilities between the terrestrial surveying tools and the existing satellite techniques.

The description of the GPS system follows the division that is customary for navigation satellites:
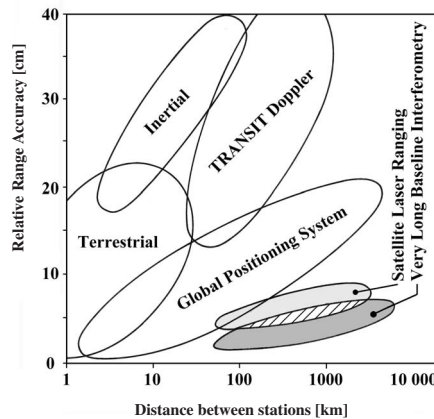
Figure 7.3. Accuracy in geodetic positioning techniques, status about 1985

Space Segment	with active satellites [7.1.2],
Control Segment	for system and time control, and for the prediction of orbits [7.1.3],
User Segment	with different receiver types [7.2].

A voluminous and fast-growing literature exists on GPS and the geodetic use of GPS. For basic technical information on the GPS system see the collection of articles from the U.S. *Institute of Navigation* (ION) as a primary reference (Janiczek (ed.), 1986), as well as the two-volume handbook edited by Parkinson et al. (1996). Recent developments are recorded in the proceedings of the annual "International Technical Meeting of the Satellite Division of the ION" *ION GPS*, or the journals "GPS World" and "GPS Solutions". Geodetic aspects are documented in the proceedings of various IAG or FIG symposia, and in all journals of geodetic relevance. Monographs and textbooks with particular emphasis to GPS are e.g. Wells (ed.) (1986), Leick (1995), Kaplan (1996), Teunissen, Kleusberg (1998) Misra, Enge (2001), Hofmann-Wellenhof et al. (2001).

GPS is not the only satellite-based navigation system. The Russian Federation is building up GLONASS which is very similar to GPS. The system briefly had a full constellation of 24 operating satellites in 1996. At the time of writing (2002), however, only a few satellites are operational. GLONASS is considered to be a valuable complementary system to GPS for future application. For more details on GLONASS see [7.7.1]. The European Union (EU) together with the European Space Agency (ESA) is planning to build up GALILEO [7.7.3] as a civil alternative to GPS.

The general name given to these systems is *Global Navigation Satellite System* (GNSS). For the next about five years, however, GPS most probably will remain the only operational system of its kind.

7.1.2 Space Segment

The basic constellation, when fully implemented, consists of 24 space vehicles. The satellites are placed in almost circular orbits in six orbital planes, with an orbital inclination of 55 degrees. The orbital height is about 20 200 km, corresponding to about 26 600 km for the semi-major axis. The orbital period is exactly 12 hours of sidereal time, and provides repeated satellite configurations four minutes earlier each day with respect to universal time.

The arrangement of satellites in the full constellation, the so-called *baseline constellation*, is demonstrated in Fig. 7.4. The orbital position of each satellite in one of the six orbital planes A to F is indicated by its plane position number, also named *slot*. Four slots are assigned to each plane. Six additional slots, $A5$ through $F5$, are provided on the basis of need for active spares. The separation in right ascension Ω between two orbital planes is $60°$. The position of a satellite within the particular orbital plane can be identified by the *argument of the latitude* (3.87),

$$u = \omega + v,$$

or the mean anomaly, M, for a given epoch. Although the baseline constellation
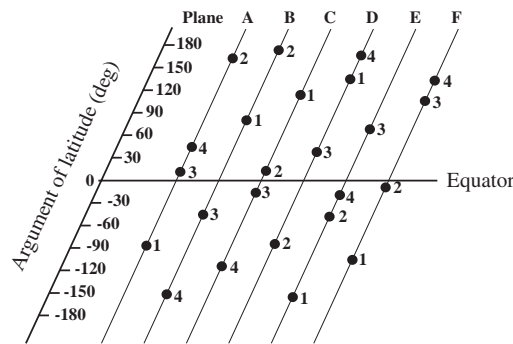
Figure 7.4. GPS 24 satellites baseline constellation

includes 24 satellites, the number of active satellites on orbit may vary due to failures, launches, or maintenance requirements, and since 1995 has exceeded 24. On January 1, 2003, the constellation comprised 28 satellites. With the augmented constellation, most users will have six to eight, or at times even more, satellites in view instead of the minimum of four satellites.

Three generations of satellites have been launched:

        Block I   development satellites,
        Block II/IIa   production satellites, and
        Block IIR   replenishment satellites.

Eleven Block I satellites, NAVSTAR 1 to 11, were launched between 1978 and 1985 into two orbital planes of 63° inclination. The design life of these prototype test vehicles was only five years, but has been exceeded in most cases. One advantage of the prototype satellites was that the navigation signals were not subject to deliberate corruption (cf. [7.1.6]). The fundamental software concepts for the geodetic use of the GPS signals have been developed based on data from these satellites.

The first Block II production satellite was launched in February 1989. A total of 28 Block II operational vehicles have been built and launched to support the 24 satellite configuration. The launching vehicle was the McDonnell Douglas Delta 2 booster. Beginning in November 1989, a slightly modified version, the upgraded (*advanced*) Block IIa, carrying more capable and reliable systems, was introduced. The design life-time of the operational Block II satellites is 7.5 years, but after more than 10 years of operation the real lifetime in orbit has turned out to be much longer.

Fig. 7.5 gives a schematic view of a Block II/IIa satellite. Electrical power is supplied by two solar energy collector plates with a surface area of 7.2 m^2 each. The large panels and momentum reaction wheels help stabilize the satellite. There is additional battery back-up to provide energy when the satellite moves into Earth's shadow (*eclipse period*). Each satellite weighs 845 kg, and has a propulsion system for positional stabilization and orbit maneuvers. As the supply of fuel is rather limited
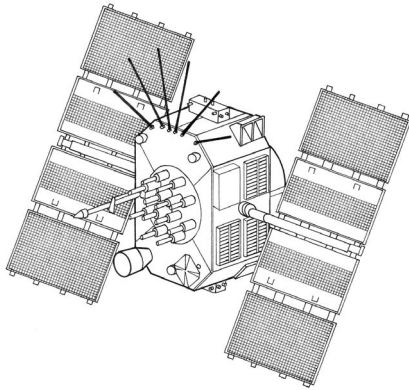
Figure 7.5. Schematic view of a Block II/IIa
GPS satellite
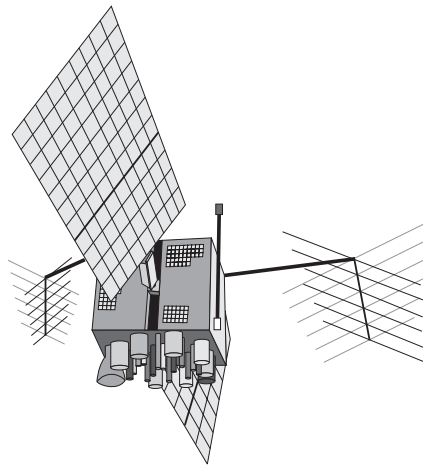
Figure 7.6. Schematic view of a Block IIR
GPS satellite

orbit changes usually last several weeks or months.

Each satellite carries high performance frequency standards with an accuracy of between 1×10^{-12} and 1×10^{-13}, forming a precise time base. The prototype satellites were partly equipped only with quartz oscillators. All Block II/IIa production satellites, however, have two cesium frequency standards and two rubidium frequency standards [2.2.5] (Van Melle, 1990).

The development and deployment of the next generation is underway. Twenty *replenishment satellites*, to be known as Block IIR satellites, will replace the current Block II satellites as necessary (Fig. 7.6). The satellites have an in-orbit mass of 1100 kg and carry Cesium and Rubidium clocks. Two of the new design features are the ability to measure distances between the satellites (*crosslink ranges*), and to compute ephemeris on-board (Kaplan, 1996). This *autonav* capability enables the satellites to generate their own navigation message for a period of 180 days. The signal and data transmission is identical to the Block II/IIa satellites. After a launch failure the first Block IIR satellite was successfully launched in July 1997. Probably from 2004 onwards a modified version of satellites, Block IIR-M, will include a new civil signal on L2 (see [7.1.7]). The Block IIR satellites will sustain the constellation at least until 2005. For details see e.g. Kaplan (1996), Parkinson et al. (1996, Vol. I, chap. 6) or Misra, Enge (2001).

A new generation of GPS satellites, the *Follow-On* Block IIF satellites, with improved facilities, is under construction. The first six satellites are ready for delivery in 2003 and will be "launched on need" (LON), probably after 2005. One important feature for civil use will be the inclusion of a third civil signal, $L5$. Plans for a new series of satellites, called GPS III, are underway, see [7.1.7].

Two carrier frequencies in the L-band are coherently derived from the fundamental

10.23 MHz frequency, generated by the onboard atomic oscillators:

$$L1 : 154 \times 10.23 \, \text{MHz} = 1575.42 \, \text{MHz} \, (\widehat{=} \, 19.0 \, \text{cm})$$
$$L2 : 120 \times 10.23 \, \text{MHz} = 1227.60 \, \text{MHz} \, (\widehat{=} \, 24.4 \, \text{cm}).$$

Each satellite transmits signals on both frequencies. These are the navigation signals (codes), and the navigation and system data (message). The codes are modulated on the carrier frequencies as so-called *pseudo random noise* (PRN) sequences. The L1 signal contains both the precise P-code and the less precise C/A-code. The L2 signal contains only the P-code. The signal structure is discussed in more detail in section [7.1.4].

The antenna beam of a Block II satellite is somewhat larger than the angle of the Earth, as seen from the satellite altitude (Fig. 7.7). This enables GPS receivers inside LEO satellites to receive GPS signals, if they are not blocked by Earth's shadow (Spilker, 1996b).

The GPS satellites are identified by different numbering schemes (cf. Table 7.5, p. 232), including the block-wise launch sequence number, the SVN (space vehicle number) or NAVSTAR number, and the PRN (pseudo random noise) or SVID (space vehicle identification) number that is related to the PRN segment allocated to the individual satellite. Usually the PRN number is taken for identification, and consequently it will be used throughout this book.

An overview of the main characteristics of the satellites currently in use or under construction is given in Table 7.2
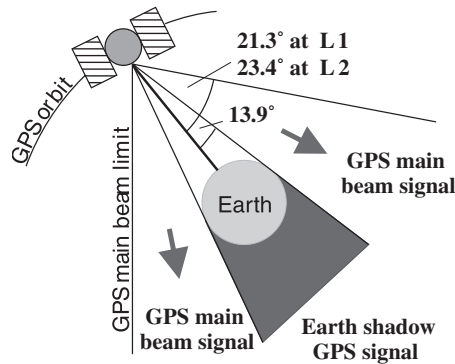
Figure 7.7. GPS signals with respect to Earth (not at scale), after Spilker (1996b)

(Misra, Enge, 2001). Remarkably the satellites' unit cost has decreased with increasing capability.

Table 7.2. Satellites forming the baseline constellation

	Block II/IIA	Block IIR	Block IIF
Number	28	21	12
First Launch	1989	1997	≈ 2005
Weight (kg)	845	1100	≈ 1700
Power/solar panel (W)	1100	1700	≈ 2900
Design life (years)	7.5	10	15
Unit cost (U.S. Dollars)	43 M	30 M	≈ 28 M

7.1.3 Control Segment

The tasks of the Control Segment are to (e.g. Russel, Schaibly, 1980; Misra, Enge, 2001)

– continuously monitor and control the satellite system,
– determine the GPS system time,
– predict the satellite ephemerides and the behavior of the satellite clocks,
– periodically update the navigation message for each particular satellite, and
– command small maneuvers to maintain orbit, or relocate to substitute an un-healthy satellite.

Within the Control Segment are the *Master Control Station* (MCS), several unmanned *monitor stations* (MS) located around the world, and *ground antennas* (GA) for uploading data to the satellites. The *Operational Control Segment* (OCS) for GPS consists of the MCS near Colorado Springs (U.S.A.), four monitor stations and co-located ground antennas in Ascension Island, Cape Canaveral, Diego Garcia and Kwajalein, and two more monitor stations in Colorado Springs and Hawaii (Fig. 7.8). The monitor stations and ground antennas are operated remotely from the Master Control Station.

▲ Air Force Monitor Station
● NIMA Monitor Station
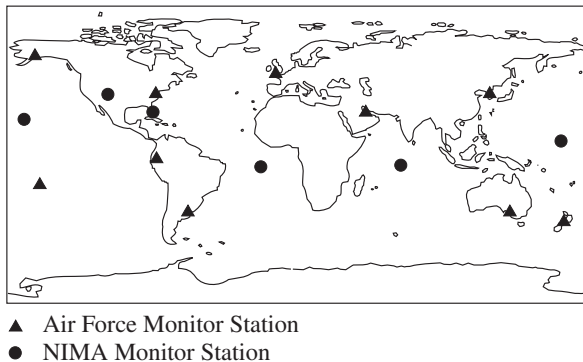
Figure 7.8. Control segment with observation stations

The monitor stations receive all satellite signals, from which they determine the pseudoranges to all visible satellites, and transmit the range data along with local meteorological data via data link to the Master Control Station. From these data the MCS precomputes satellite ephemerides and the behavior of the satellite clocks and formulates the navigation data (message). The message data are transmitted to the ground antennas and uplinked via S-band to the satellites in view. Fig. 7.9 shows this process schematically. Because of the global distribution of the upload antennas at least three contacts per day can be realized between the control segment and each particular satellite.

Signals transmitted by GPS satellites are based on *GPS System Time*. Until June 1990 this was the time given by the cesium oscillator at one of the monitor stations.
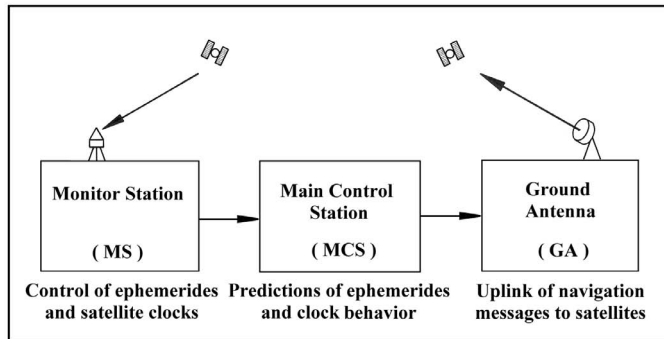
Figure 7.9. Data flow in the determination of the broadcast ephemeris

Since then the practice has been to obtain GPS time as the weighted mean (paper clock) of all operational monitor station and satellite clocks. GPS time is controlled over the long term to remain within one microsecond of the international time standard UTC [2.2.3], [7.1.5.3] (Langley, 1999a) without considering leap seconds.

The requirements of an operational navigation system are completely met by the geographical distribution of the monitor stations. The coverage, however, does not in all cases satisfy precise orbit determination requirements for geodetic, and in particular geodynamic, applications. Much denser networks of monitor stations, mostly under civil national and international responsibilities, have been built up and are operational. One eminent example is the *International GPS Service* (IGS) (cf. [7.4.3], [7.8.1]). The *National Imagery and Mapping Agency* (NIMA) runs its own network of monitor stations. Along with the DoD's GPS Accuracy Improvement Initiative (AII) [7.1.7] it is planned to include data from a subset of the NIMA monitor stations into the prediction of the broadcast ephemerides (see Fig. 7.8). The upload of navigation data, consisting of predicted orbits and clock corrections, is made to each satellite about once daily. The cross-link ranging capability of the Block IIR and IIF satellites will allow the satellites to update their broadcast ephemeris autonomously and operate over some period without contact from the control segment. For further details on GPS orbit computation and orbit representation, see [7.1.5], [7.4.3].

7.1.4   Observation Principle and Signal Structure

NAVSTAR GPS is a *one-way ranging system*, i.e. signals are only transmitted by the satellite [4.2.2]. The fundamental observable is the signal travel time between the satellite antenna and the receiver antenna. The signal travel time is scaled into a range measurement using the signal propagation velocity.

One-way ranging means that a clock reading at the transmitter antenna is compared with a clock reading at the receiver antenna. In general, it cannot be assumed that the two clocks are strictly synchronized. The observed signal travel time thus

contains a systematic synchronization error (time bias). Biased ranges are also called *pseudoranges*. Hence, the basic observation principle of GPS can be regarded as the determination of pseudoranges. Fig. 7.2 demonstrates that the simultaneous observation of four pseudoranges is required to derive the three coordinates of the user antenna and the clock synchronization error. As an additional requirement, it is also necessary to know the satellite position and the satellite time (cf. [7.3.1]).

GPS signals must provide a means for determining positions in real-time. This is achieved by modulating the carriers with *pseudorandom noise* (PRN) codes. These are sequences of binary values (zeros and ones, or $+1$ and -1) which appear to have random character, but which can be identified unequivocally. Their most important property is a low autocorrelation value for all delays except those that coincide exactly. The pseudoranges are derived from the travel time of an identified coded PRN signal. Two different codes are in use, the P-code and the C/A-code. P means *precision* or *protected*, and C/A means *clear/acquisition*.

The *P-code* has a frequency of 10.23 MHz, i.e. a sequence of 10.23 million binary digits or *chips* per second. This frequency is also referred to as the *chipping rate* of the P-code. The corresponding "wavelength" of one chip is about 30 m. The P-code sequence is extremely long; it only repeats after 266 days ($= 38$ weeks). Portions of seven days each are assigned to the various satellites. As a result, all satellites can transmit on the same frequency and can be identified by their unique one-week PRN-segment. This technique is also called *code division multiple access* (CDMA). The code segments are set back to zero each week at midnight (0^h UT) from Saturday to Sunday. The P-code is the principle code for navigation and available on both carrier frequencies L1 and L2. Note that with the implementation of *Anti-Spoofing* the P-code has been encrypted for non-authorized users (see [7.1.6]).

The *C/A-code* has a length of only one millisecond and is generated at a chipping rate of 1.023 MHz. The corresponding wavelength is about 300 m. The C/A-code is currently only transmitted on the L1 carrier. The epochs of both codes are synchronized. For detailed information on the structure and the generation of the codes, see e.g. Spilker (1980), Forsell (1991), Kaplan (1996) or Parkinson et al. (1996). Note that a complete alteration in the signal structure for civil use is expected with the launch of the modified Block IIR satellites, Block IIR-M, after 2003, and the launch of the Block IIF satellites after 2005. For details see [7.1.7] and e.g. Van Dierendonck, Hegarty (2000); Fontana et al. (2001).

To determine the signal propagation time, the user needs a copy of the code sequence in the receiver. This code sequence is phase-shifted in time step by step, and correlated with the received code signal until maximum correlation is achieved. The necessary phase shift in the two sequences of codes is a measure of the signal travel time between the satellite and receiver antennas (cf. [7.3.1.2]). This technique can be described as *code phase observation*.

For precise geodetic applications the pseudoranges have to be derived from phase measurements on the carrier signals because of the much better resolution. This technique requires, however, a solution to the problem of ambiguity determination,

and is discussed in more detail in section [7.3.2.3].

The third type of signal transmitted from a GPS satellite is the broadcast message (cf. [7.1.5.4]). The message is sent at a rather slow rate of 50 bits per second (bps), and repeats every 30 seconds. Both code chip sequences are separately combined with the stream of message bits by binary addition; i.e. the same value for code and message chip gives 0, and different values result in 1.

The main features of all three signal types used in GPS observations, namely carrier, code, and data signals, are given in Table 7.3. The signal structure permits both the phase and the phase shift (Doppler effect, cf. [6.1]) to be measured, as well as the direct signal propagation. The necessary bandwidth is achieved by phase modulation ($0°$ and $180°$) of the PRN-code (Fig. 7.10 and 7.11).

Table 7.3. GPS satellite signals (bps = bits per second)

Atomic clock (Cs, Rb) fundamental frequency	10.23 MHz
L1 carrier signal	154×10.23 MHz
L1 frequency	1575.42 MHz
L1 wavelength	19.0 cm
L2 carrier signal	120×10.23 MHz
L2 frequency	1227.60 MHz
L2 wavelength	24.4 cm
P-code frequency (chipping rate)	10.23 MHz (Mbps)
P-code wavelength	29.31 m
P-code period	266 days; 7 days/satellite
C/A-code frequency (chipping rate)	1.023 MHz (Mbps)
C/A-code wavelength	293.1 m
C/A-code period	1 millisecond
data signal frequency	50 bps
data signal cycle length	30 seconds

As a whole, the L1 signal has the following structure (Spilker, 1980; Wübbena, 1991):

$$S_{L1}(t) = A_p P_i(t) D_i(t) \sin(\omega_1 t) + A_c C_i(t) D_i(t) \cos(\omega_1 t), \qquad (7.1)$$

where

A_p	amplitude of the P-code,
$P_i(t)$	P-code sequence with state ± 1,
$D_i(t)$	data stream with state ± 1,
A_c	amplitude of the C/A-code,
$C_i(t)$	C/A-code sequence with state ± 1, and
$A \sin(\omega_1 t)$	carrier signal.

The index i stands for the i-th satellite.

The L2 signal has a much simpler structure because it does not contain the C/A-code:

$$S_{L2}(t) = B_p P_i(t) D_i(t) \sin(\omega_2 t). \qquad (7.2)$$

Here, $P_i(t)$ is again the P-code sequence for the i-th satellite and B_p is the P-code amplitude. The epochs of both codes and carriers are synchronized.

Fig. 7.10 shows how code and carrier are combined. The technique is called *binary biphase modulation* (also known as *binary phase shift keying* BPSK, see [7.2.1]). Because the PRN-codes (and the message) are binary data streams, only two states of phase modulation are possible. The state $+1$ or -1 leaves the carrier unchanged; a code transition from $+1$ to -1 or from -1 to $+1$ involves a phase shift of 180°.
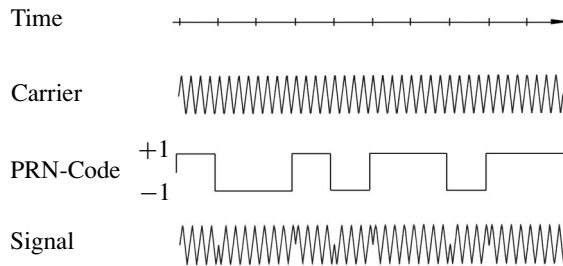
Figure 7.10. Structure of GPS satellite signals

The L1 channel has to carry both codes. This is accomplished by a technique named *phase quadrature*. The unmodulated L1 carrier is split off and shifted in phase by 90° before it is mixed with the C/A-code signal, and is then added to the modulated P-code signal. This procedure is implied in equation (7.1), and is demonstrated in Fig. 7.11, cf. Wells (ed.) (1986); Langley (1990).

The binary biphase modulation with a PRN-code sequence produces a rather broad bandwidth for the navigation signals. This technique is referred to as *spread spectrum technique* [7.2.1] and limits the interference from other signals (Spilker, 1980; Forsell, 1991; Parkinson et al., 1996; Misra, Enge, 2001). The P-code spectrum has a bandwidth of 20 MHz, corresponding to a resolution of 1 nanosecond $\hat{=}$ 30 cm for conditions with a good signal-to-noise ratio. The bandwidth of the C/A-code is 2 MHz, corresponding to a ten-fold reduction in signal resolution.

Direct access to the P-code is only possible for receivers that are precisely synchronized with the GPS system time, and located at a site with exactly known coordinates. This is why access to the P-code is, in general, achieved with the aid of the much shorter C/A-code via the *Hand Over Word* (HOW). The HOW contains the so-called *Z-count* and appears in every subframe of the data signal (cf. [7.1.5]). The Z-count is defined as the integer number of 1.5-second periods since the beginning of the GPS week [7.1.5.3] and thus identifies the epoch of a data record in GPS time. If one knows the Z-count, one can acquire the P-code within the next six seconds.
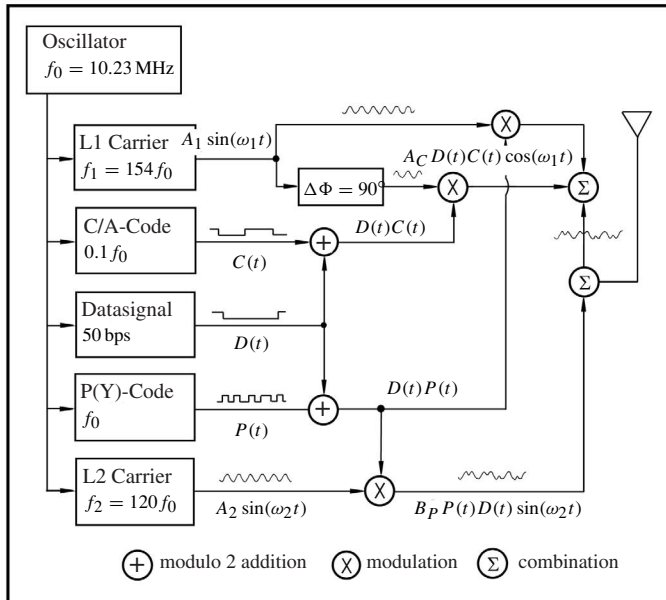
Figure 7.11. Generation of GPS signals

For more details on the rather complicated GPS signals see e.g. Forsell (1991); Parkinson et al. (1996); Misra, Enge (2001) and the *Interface Control Document* ICD-GPS-200C with actual revisions. An excellent elementary introduction is given by Langley (1990).

7.1.5   Orbit Determination and Orbit Representation

7.1.5.1   Determination of the Broadcast Ephemerides

In order to solve the navigation task the user must have real-time access to the satellite positions and satellite system time. This is made possible by the orbit information, the *navigation message*, that is contained in the data signal. The navigation message is determined by the Control Segment and "broadcast" to the users via the GPS satellites.

These *broadcast ephemerides* are generated in two steps. First, a *reference ephemeris*, based on several days of observations from the monitor stations, is generated (*off-line*) using a highly sophisticated software package for orbit determination (cf. [3.3.2.2]). In the second step (*on-line*) the discrepancies between the current observations at the monitor stations and the reference ephemeris are derived, and are processed in a linear *Kalman filter algorithm* to predict corrections to the reference ephemeris.

For this purpose, code-pseudorange and carrier observations are made of all visible satellites at all monitor stations. The data are corrected for ionospheric and tropospheric delays, for Earth rotation and for relativistic effects. The corrected measurements and carrier-aided smoothed observations are input into the Kalman filter process and are used to estimate the following states (Parkinson et al., 1996, chap. 10):

- satellite position at epoch,
- satellite velocity at epoch,
- three clock parameters per satellite,
- solar radiation pressure coefficients per satellite,
- y-axis acceleration bias,
- two clock parameters per monitor station, and
- one tropospheric scale factor per monitor station.

The estimated perturbations in the elements are used to correct the satellite reference ephemeris and to generate the broadcast ephemerides. In a similar way the satellite clock behavior is predicted and included in the data signal in the form of a second order polynomial.

Computation of the satellite trajectories is based on the gravity field parameters and the station coordinates of the *World Geodetic System* 1984 (WGS 84). In order to improve the accuracy of the ephemeris the WGS 84 station coordinates were replaced by ITRF 91 coordinates in 1994, and by ITRF 94 coordinates in 1996, cf. [2.1.6]. Earth orientation parameters are taken from the IERS Rapid Service [12.4.2]. The process of orbit determination is still based on the technology of the 1980s (Russel, Schaibly, 1980; Swift, 1985) but will be upgraded along with the *Accuracy Improvement Initiative* (AI I) [7.1.7].

7.1.5.2   Orbit Representation

The satellite positions estimated in the Kalman filter process are next represented in the form of Keplerian elements with additional perturbation parameters. Table 7.4 summarizes all parameters that describe the satellite orbit and the state of the satellite clock. The parameters refer to a given reference epoch, t_{0e} for the ephemeris and t_{0c} for the clock, and they are based on a four hours curve fit (ICD, 1993). Hence, the representation of the satellite trajectory is achieved through a sequence of different disturbed Keplerian orbits.

At present, a fresh data set is broadcasted every two hours, causing small steps between the different overlapping representations. These steps can reach a few decimeters but may be smoothed by suitable approximation techniques, e.g. Chebyshev polynomials [3.3.3.2].

The parameter set of Table 7.4 is used to compute the satellite time and the satellite coordinates. The unit "semicircles" can be converted to degrees (multiplication by 180), or to radians (multiplication by π). The first group of parameters is used to

Table 7.4. Representation of GPS broadcast ephemerides

	Time parameters
t_{0e}	Reference time, ephemeris parameters [s]
t_{0c}	Reference time, clock parameters [s]
a_0, a_1, a_2	Polynomial coefficients for clock correction (bias [s], drift [s/s], drift-rate (ageing) [s/s^2])
IODC	Issue of Data, Clock, arbitrary identification number
	Keplerian parameters
$\sqrt{A}$	Square root of the semi-major axis [m$^{1/2}$]
e	eccentricity [dimensionless]
i_0	inclination angle at reference time [semicircles]
Ω_0	Longitude of ascending node at reference time [semicircles]
ω	Argument of perigee [semicircles]
$\overline{M}_0$	Mean anomaly at reference time [semicircles]
IODE	Issue of Data, Ephemeris, arbitrary identification number
	Perturbation parameters
Δn	Mean motion difference from computed value [semicircles/s]
$\dot{\Omega}$	Rate of change of right ascension [semicircles/s]
$\dot{i}$	Rate of change of inclination [semicircles/s]
C_{us}	Amplitude of the sine harmonic correction term to the argument of latitude [rad]
C_{uc}	Amplitude of the cosine harmonic correction term to the argument of latitude [rad]
C_{is}	Amplitude of the sine harmonic correction term to the angle of inclination [rad]
C_{ic}	Amplitude of the cosine harmonic correction term to the angle of inclination [rad]
C_{rs}	Amplitude of the sine harmonic correction term to the orbit radius [m]
C_{rc}	Amplitude of the cosine harmonic correction term to the orbit radius [m]

correct satellite time. The second group determines a Keplerian ellipse at the reference epoch, and the third group contains nine perturbation parameters. These are:

Δn secular drift in $d\omega/dt$ due to the second zonal harmonic (C_{20}); also it absorbs effects of the Sun's and Moon's gravitation and solar radiation pressure over the interval of fit,

$\dot{\Omega}$ secular drift in right ascension of the node due to the second zonal harmonic; includes also effects of polar motion,

$\dot{i}$ rate of change of inclination, and

C_{us}, C_{uc} short periodic effects of C_{20}; also include higher order effects and

C_{is}, C_{ic} short periodic effects of lunar gravitation (during the closest approach

C_{rs}, C_{rc} of the space vehicle to the Moon); also absorb further perturbations.

Fig. 7.12 explains the Keplerian and the perturbation parameters. Note that the element Ω_0 in the GPS message is not measured from the vernal equinox, Υ, but from the zero meridian, X_T. In essence, Ω_0 is not a right ascension angle but a longitude. In recent literature the parameter is therefore designated as *longitude of ascending node (LAN)*.
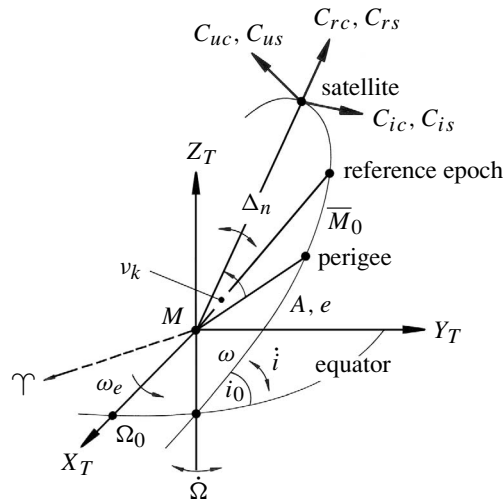
Figure 7.12. Keplerian and disturbance parameters in the broadcast message

7.1.5.3 Computation of Satellite Time and Satellite Coordinates

The GPS system time is characterized by a week number and the number of seconds since the beginning of the current week; the GPS time can hence vary between 0 at the beginning of a week and 604 800 at the end of a week. The initial GPS epoch is January 5, 1980 at 0^h UTC. This is why the GPS week starts at midnight (Universal Time) between Saturday and Sunday. The GPS week number is included in subframe 1 of the navigation message [7.1.5.4], and is represented by 10 bits. Hence the largest possible week number is 1023, and at the end of the week with the number 1023 the week number "rolls over" to zero (cf. ICD-GPS-200C). This event is called the *End of Week (EOW) rollover*. The first cycle of week numbers ended on August 21, 1999. The current second cycle runs from August 22, 1999 until April 6, 2019 (Langley, 1999a). Note that for some purposes a continuous numbering of the GPS week is used (no rollover), e.g. for RINEX data [7.3.3.2].

The GPS system time is a continuous time scale, and is defined by the weighted mean of the atomic clocks in the monitor stations and the satellites (cf. [7.1.3]). The leap seconds in the UTC time scale, and the drift in the GPS clocks mean that GPS system time and UTC are not identical [2.2.3]. The difference is continuously monitored by the control segment and is broadcast to users in the navigation message. On January 1, 2003, the difference was about 13 seconds (GPS time ahead).

Because of constant and irregular frequency errors in the satellite oscillators, the satellite clock readings differ from the GPS system time. The behavior of the individual satellite clocks (rubidium or cesium oscillator) is monitored by the control segment, and predicted in the form of a second degree polynomial. The polynomial coefficients

are included in the first parameter group of Table 7.4. The individual satellite time, t_{SV}, is corrected to GPS system time, t, using

$$t = t_{SV} - \Delta t_{SV}, \tag{7.3}$$

in which

$$\Delta t_{SV} = a_0 + a_1(t - t_{0c}) + a_2(t - t_{0c})^2; \tag{7.4}$$

t_{0c} is the reference epoch for the coefficients a_0, a_1, a_2. Following the Interface Control Document, the OCS shall control the GPS time to be within one microsecond of UTC (USNO) modulo one second. In practice, GPS time has been kept within about 10 nanoseconds. The term a_0 in the satellite message hence has only a value of a few nanoseconds, (see also [2.2.3]).

The time parameter, t, can be substituted in the further calculation by t_{SV} without loss of accuracy. Differentiating (7.4) with respect to time yields an expression for the drift of the satellite clock:

$$\dot{\Delta t}_{SV} = a_1 + 2a_2(t - t_{0c}). \tag{7.5}$$

The satellite coordinates X_k, Y_k, Z_k are computed for a given epoch, t, with respect to the Earth-fixed geocentric reference frame X_T, Y_T, Z_T (cf. [2.1.2]). The time, t_k, elapsed since the reference epoch, t_{0e}, is

$$t_k = t - t_{0e}. \tag{7.6}$$

A possible change of the week has to be considered. Two constants are required:

$$GM = 3.986005 \cdot 10^{14} \text{m}^3/\text{s}^2 \quad \text{WGS 84 value of the geocentric} \tag{7.7}$$
$$\text{gravitational constant,}$$

$$\omega_e = 7.292115 \cdot 10^{-5} \text{ rad/s} \quad \text{WGS 84 value of the Earth rotation rate.} \tag{7.8}$$

Also

$$\pi = 3.1415926535898 \quad \text{(exactly).}$$

Note that (7.7) is not the refined WGS 84 value for GM from 1994 [2.1.6], but the original WGS 84 value. Actually, the more recent GM value is used for precise prediction of GPS orbits in the OCS whereas the old value is used for the conversion from the predicted Cartesian state vectors into the quasi-Keplerian broadcast elements. Hence it should also be used for interpolation purposes to obtain satellite positions at a given epoch. For details on the subject see (NIMA, 2000).

Furthermore we use:

$$A = (\sqrt{A})^2 \qquad \text{Semi-major axis,} \tag{7.9}$$

$$n_0 = \sqrt{\frac{GM}{A^3}} \qquad \text{Computed mean motion,} \tag{7.10}$$

$$n = n_0 + \Delta n \quad \text{Corrected mean motion, and} \tag{7.11}$$

$$\overline{M}_k = \overline{M}_0 + n t_k \quad \text{Mean anomaly.} \tag{7.12}$$

Kepler's equation of the eccentric anomaly (3.53),

$$E_k = \overline{M}_k + e \sin E_k, \tag{7.13}$$

is solved by iteration. Because of the very small eccentricity of the GPS orbits ($e < 0.001$) two steps are usually sufficient:

$$E_0 = \overline{M}, \quad E_i = \overline{M} + e \sin E_{i-1}, \quad i = 1, 2, 3, \ldots \tag{7.14}$$

The satellite coordinates are then obtained, using equations (7.15) to (7.29):

$$\cos v_k = \frac{\cos E_k - e}{1 - e \cos E_k} \qquad \text{True anomaly,} \tag{7.15}$$

$$\sin v_k = \frac{\sqrt{1 - e^2} \sin E_k}{1 - e \cos E_k} \qquad \text{True anomaly,} \tag{7.16}$$

$$\Phi_k = v_k + \omega \qquad \text{Argument of latitude,} \tag{7.17}$$

$$\delta u_k = C_{uc} \cos 2\Phi_k + C_{us} \sin 2\Phi_k \qquad \text{Argument of latitude correction,} \tag{7.18}$$

$$\delta r_k = C_{rc} \cos 2\Phi_k + C_{rs} \sin 2\Phi_k \qquad \text{Radius correction,} \tag{7.19}$$

$$\delta i_k = C_{ic} \cos 2\Phi_k + C_{is} \sin 2\Phi_k \qquad \text{Inclination correction,} \tag{7.20}$$

$$u_k = \Phi_k + \delta u_k \qquad \text{Corrected argument of latitude,} \tag{7.21}$$

$$r_k = A(1 - e \cos E_k) + \delta r_k \qquad \text{Corrected radius,} \tag{7.22}$$

$$i_k = i_0 + \dot{i} t_k + \delta i_k \qquad \text{Corrected inclination,} \tag{7.23}$$

$$X'_k = r_k \cos u_k \qquad \text{Position in the orbital plane,} \tag{7.24}$$

$$Y'_k = r_k \sin u_k \qquad \text{Position in the orbital plane,} \tag{7.25}$$

$$\Omega_k = \Omega_0 + (\dot{\Omega} - \omega_e)t_k - \omega_e t_{0e} \qquad \text{Corrected longitude of ascending node,} \tag{7.26}$$

$$X_k = X'_k \cos \Omega_k - Y'_k \sin \Omega_k \cos i_k \qquad \text{Earth fixed geocentric satellite coordinates,} \tag{7.27}$$

$$Y_k = X'_k \sin \Omega_k + Y'_k \cos \Omega_k \cos i_k \qquad \text{Earth fixed geocentric satellite coordinates,} \tag{7.28}$$

$$Z_k = Y'_k \sin i_k \qquad \text{Earth fixed geocentric satellite coordinates.} \tag{7.29}$$

Equation (7.26) implicitly describes the relation between the vernal equinox, Υ, and the current position of the zero meridian, based on Earth's rotation ω_e.

7.1.5.4   Structure of the GPS Navigation Data

The GPS navigation data, the so-called *message*, is organized as in Fig. 7.13. The user has to decode the data signal [7.1.4] to access the navigation data. Decoding is executed in the internal receiver processor for on-line navigation purposes. Most

receiver manufacturers provide decoding software for postprocessing purposes, in many cases combined with the so-called *download software* for reading data from the receiver to a computer [7.3.3.2].

With a bit-rate of 50 bps, and a cycle time of 30 seconds, the total information content of a navigation data set is 1500 bits. The complete *data frame* is subdivided into five *subframes* of six seconds duration (corresponding to 300 bits each). Each subframe contains ten data words of 30 bits each, six of them being control bits. The first two words of each subframe are the *telemetry word* (TLM), and the C/A- to P-code *hand over word* (HOW). The TLM word contains a synchronization pattern which facilitates the access to the navigation data.
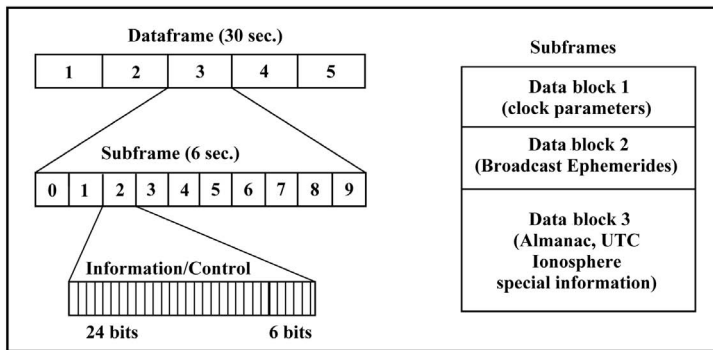
Figure 7.13. Structure of the GPS navigation data

The navigation data record is divided into three data blocks:

Data Block I appears in the first subframe and contains the GPS week number, the satellite clock correction terms, and the SV accuracy and health.

Data Block II appears in the second and third subframes and contains all necessary ephemeris parameters for computation of the satellite coordinates.

Data Block III appears in the fourth and fifth subframes and contains the almanac data [7.1.5.1] with clock and ephemeris parameters for all available satellites of the GPS system. The data block includes also ionospheric correction parameters [7.4.4.1], UTC data, and particular alphanumeric information for *authorized users*.

Unlike the first two data blocks, subframes four and five are not repeated every 30 seconds. The two subframes consist of 25 pages that appear subsequently, such that the total information content is available after 12.5 minutes. Each page covers the *almanac data* of one satellite from the total constellation. These are parameters representing the ephemeris of the particular space vehicle, corrections to the satellite clock, identification number, and satellite health status. The less accurate almanac data can be used for a computation of satellite predictions (alerts), and also for a faster lock-on to satellite navigation signals. Subframe 5 includes almanac and health status data for satellites numbered 1–24, and subframe 4 those for satellites numbered 25 and higher.

For a detailed description of the data format in the GPS navigation message see e.g. Van Dierendonck et al. (1980), Parkinson et al. (1996, chap. 4), and the current version of the official *Interface Control Document*, e.g. ICD-GPS-200C. These sources are essential for the development of decoding software and for a deeper understanding of GPS data signals. A very instructive text is also given by Tsui (2000). Note that the forthcoming L5 signal on the Block IIF satellites will have a different message format (Van Dierendonck, Hegarty, 2000).

7.1.6 Intentional Limitation of the System Accuracy

GPS is a military navigation system, a responsibility of the U.S. Department of Defense (DoD), and has hence to meet the national security interests of the United States. Accordingly, it has been stated from the beginning of the system's development that only limited access to the total system accuracy would be available to the national and international civil user community. The interests of the civil user community enter through the Department of Transportation (DOT) within the *Interagency GPS Executive Board* (IGEB). The IGEB is the coordinating body for GPS policy, and it is co-chaired by representatives of DoD and DOT.

The service available to the civil community is called *Standard Positioning Service* (SPS), while the service available to authorized (mainly military) users is called the *Precise Positioning Service* (PPS). Throughout the 1990s, the accuracy available to SPS users was 100 m 2D-RMS (cf. [7.4.2]). This figure means that a horizontal (two dimensional) position accuracy of 100 m or better can be expected by a stand-alone user 95% of the time. PPS provides the full system accuracy of 10 to 20 meters in three dimensions.

Two modes of limitation were introduced. These are *Anti-Spoofing* (AS) and *Selective Availability* (SA). *Anti-Spoofing* entails the *encryption* of the P-code, i.e. use of a protected code named *Y-code*. Only authorized users will have the means to get access to the P-code while AS is activated. *Selective Availability* means an intentional degradation of the GPS signals by adding controlled errors in the measurement data. SA uses two effects. These are:
  − ephemeris data manipulation (ε *technique*), and
  − dithering, or systematic destabilizing, of the satellite clock (δ *technique*).
Both effects corrupt the measured pseudoranges. Apparently mainly the dithering-technique was used, resulting in a roughly five-fold increase in positioning error (see Fig. 7.14).

SA was implemented for the first time on March 23, 1990, but was disabled again on August 2, 1990, due to the Gulf crisis. SA became effective again in July 1991 and was implemented to the Standard Positioning Service level in November 1991. PPS authorized users were able to remove SA. After a long discussion of the pros and cons, SA was permanently deactivated on May 2, 2000, based on a Presidential decision.

Anti-Spoofing (AS) has the aim to prevent an adversery from generating a copy of the GPS signal, and to "spoof", or mislead, a receiver. The encrypted P-code is referred to as a Y-code. AS has been active on all Block II satellites nearly continuously since
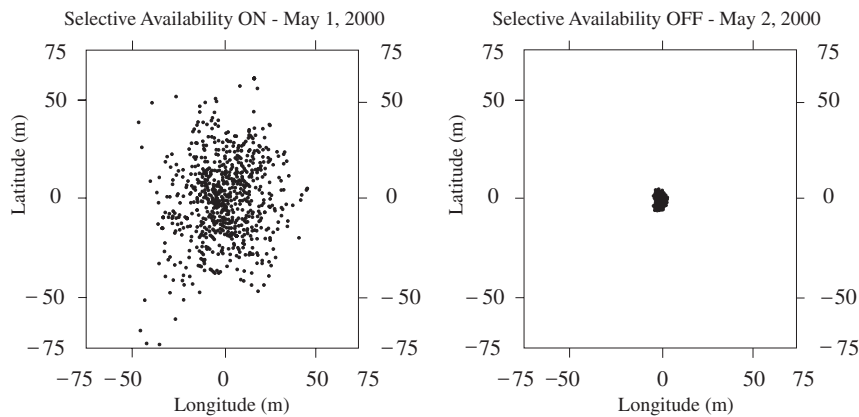
Figure 7.14. The effect of Selective Availability on positioning results with a single receiver; left SA active, right after deactivation

February 1994. As a consequence, SPS users only have clear access to the L1 carrier signal because L2 exclusively carries the encrypted P-code. Receiver manufacturers therefore have developed proprietary techniques to gain access to L2 signals under AS, however with decreased quality (see [7.2.3]).

With disabled SA the accuracy available to SPS users is similar to that for PPS users. The global average positioning accuracy is defined as $\leq$ 13 meters horizontal error and $\leq$ 22 meters vertical error (95%) (DOD, 2001). The main advantage of PPS over SPS is its robustness against jamming and spoofing, and the higher quality of the L2 signals. The situation will further improve under the "GPS Improvement Initiative" [7.1.7].

7.1.7   System Development

All satellites launched before the end of 1985 were prototype or development satellites (Block I) for test purposes. The spacecraft were launched into two orbital planes with an inclination of 63 degrees and semi-major axis of 26 600 km. The constellation was optimized for maximizing coverage in the vicinity of the Yuma Arizona Test Range but also provided good coverage for tests in other parts of the world. The ground track of the optimized constellation repeated every day and provided an identical configuration from day to day at a particular geographical location, only four minutes earlier with respect to Universal Time. Four of the ten successfully launched Block I satellites were still functioning in May 1993. The data of the prototype satellites were ideal for the development of software and observation concepts, because no signal encryption was existent.

Originally it was planned to carry all Block II satellites in the Space Shuttle and to complete the configuration by 1988. Due to the tragedy of the Shuttle *Challenger*

in January 1986 the launch schedule was considerably delayed and not resumed until February 1989 with the launch of the first Block II satellite on a Delta 2 booster.

A total of 28 Block II spacecraft have been built [7.1.2] and launched to support the so-called *baseline constellation* of 24 satellite positions with four satellites in each of the 55 degrees inclined equally-spaced orbit planes (cf. Fig. 7.4, page 214). The system was declared operational in April 1995 (*Full Operational Capability*, FOC).

The start of the Block IIR *replenishment satellites* began (after one launch failure) in July 1997. The current constellation (as of March 2003) is given in Table 7.5; the related satellite coverage for Washington D.C. is depicted in Fig. 7.15. In total, 20 Block IIR satellites will replace the current Block II/IIA satellites, and will be launched
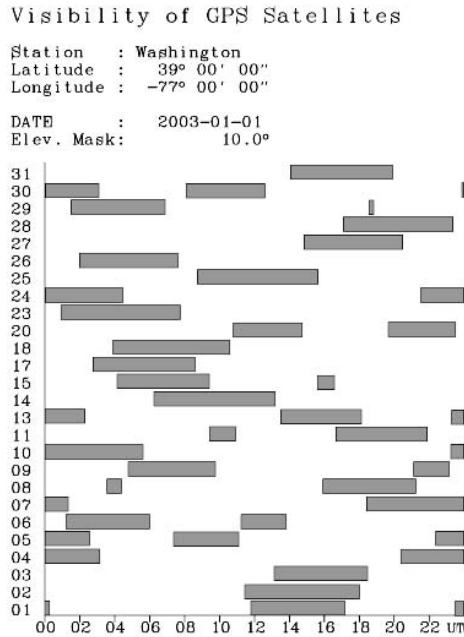
Figure 7.15. Satellite coverage for Washington, D.C., January 1, 2003

on need to maintain the constellation at least until 2005. For the time thereafter a new generation, the Block IIF satellites, is under construction (cf. [7.1.2]), and the next generation Block III is in design stages. Current plans for maintaining the constellation reach until 2030.

Two initiatives will improve and enhance the GPS, also for civil applications, within the next years. These are the
  − GPS Modernization Program, and the
  − Accuracy Improvement Initiative.
Within the *GPS Modernization Program*, besides new military capabilities (M-code), two new civil signals will be added to the forthcoming Block IIR and Block IIF

Table 7.5. Status of GPS satellites (March 2003)

Blk Seq	SVN	PRN Code	Orbit Position	Launch Date	Clock	Status Decomissioned
BLOCK II						
II-1	14	14		89-02-14		00-04-14
II-2	13	02	B–3	89-06-10	Cs	operable
II-3	16	16		89-08-18		00-10-13
II-4	19	19		89-10-21		01-09-11
II-5	17	17	D–3	89-12-11	Rb	operable
II-6	18	18		90-01-24		00-08-18
II-7	20	20		90-03-26		96-05-10
II-8	21	21		90-08-02	Cs	03-01-27
II-9	15	15	D–5	90-10-01	Cs	operable
BLOCK IIA						
II-10	23	23	E–5	90-11-26	Cs	operable
II-11	24	24	D–1	91-0704	Cs	operable
II-12	25	25	A–2	92-02-23	Cs	operable
II-13	28	28		92-04-10		92-04-25
II-14	26	26	F–2	92-07-07	Rb	operable
II-15	27	27	A–4	92-09-09	Cs	operable
II-16	32	01	F–4	92-11-22	Cs	operable
II-17	29	29	F–5	92-12-18	Rb	operable
II-18	22	22	B–1	93-02-03	Rb	operable
II-19	31	31	C–3	93-03-30	Cs	operable
II-20	37	07	C–4	93-05-13	Rb	operable
II-21	39	09	A–1	93-06-26	Cs	operable
II-22	35	05	B–4	93-08-30	Cs	operable
II-23	34	04	D–4	93-10-26	Rb	operable
II-24	36	06	C–1	94-03-10	Cs	operable
II-25	33	03	C–2	96-03-28	Cs	operable
II-26	40	10	E–3	96-07-16	Cs	operable
II-27	30	30	B–2	96-09-12	Rb	operable
II-28	38	08	A–3	97-11-06	Rb	operable
BLOCK IIR						
IIR-1	42	12		97-01-17		Launch failure
IIR-2	43	13	F–3	97-07-23	Rb	operable
IIR-3	46	11	D–2	99-10-07	Rb	operable
IIR-4	51	20	E–1	00-05-11	Rb	operable
IIR-5	44	28	B–5	00-07-16	Rb	operable
IIR-6	41	14	F–1	00-11-10	Rb	operable
IIR-7	54	18	E–4	01-01-30	Rb	operable
IIR-8	56	16	B–1	03-01-29	Rb	operable
IIR-9	45	21	D-3	03-03-31	Rb	operable
IIR-10						
IIR-11						

satellites. These are a civil signal, designated L2C, on L2, and another civil signal, L5, on a third frequency at 1176.45 MHz (McDonald, 1999; Van Dierendonck, Hegarty, 2000; Fontana et al., 2001), see Fig. 7.16.
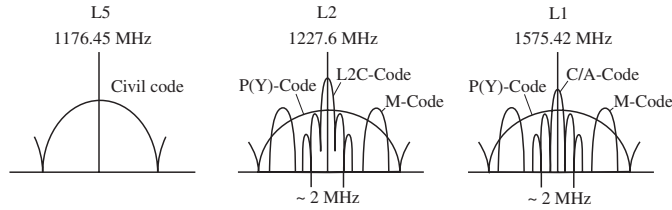
Figure 7.16. Future GPS signals

L2C will be included on modified (Block IIR-M) satellites from 2004 onwards, and L5 will be available on the Block IIF satellites, probably from 2005 onwards. L2C will carry two codes, one without data modulation. The signal will have a much better quality than it would have had as a simple addition of the C/A-code on L2 (as was previously planned). In particular, it will be possible to provide full wavelength on L2 with enhanced signal power [7.2.3]. The new L5 signal falls in a band which is protected for aeronautical radio-navigation and hence will not cause any interference to existing systems. L5 will have four times more power than L2C. The L2C signal will be available on 24 satellites by 2010 (Fontana et al., 2001). The availability of three civil signals with different capabilities will support real-time ionospheric corrections and facilitate the resolution of whole-cycle ambiguities in the carrier-phase measurements (Hatch et al., 2000) [7.3.2.3]. The high signal power will support indoor navigation and improve applications under difficult conditions, such as heavy foliage.

Several actions are planned within the *Accuracy Improvement Initiative* (AI I), cf. Hay (2000). A first step has already been realized with the use of ITRF coordinates for the monitor stations. A further step includes additional monitor stations for determination of the broadcast ephemerides. Data from 6 or more NIMA stations (see Fig. 7.8) will be transmitted via powerful datalinks to the MCS. The aging mainframe computer in the MCS will be replaced, and the Kalman filter in the orbit software will be improved, including the capability to process all satellites and ground stations simultaneously (Hay, 2000). Depending on necessity, more frequent uploads will provide submeter broadcast orbits.

Based on these improvements, the accuracy of a single pseudorange measurement, the *Signal-in-Space-Range-Error* (SISRE) will be below 1.5 m. With a PDOP of 2 [7.4.2] this corresponds to a position error of about 3 m. Further improvements can be expected with the inclusion of L5 and inter-satellite tracking capability.

Information on the system status can be obtained from different public and commercial sources. The U.S. Coast Guard (USCG) has the responsibility to provide GPS operational capability and status information to civil users. In Germany the GPS Information Service (GIBS), operated by the Bundesamt für Kartographie und

Geodäsie (BKG), delivers extensive information. Current lists of information services are published regularly in GPS periodicals (e.g. *GPS World*), see also [7.8.2].

7.2   GPS Receivers (User Segment)

Appropriate satellite receivers are required to use the GPS signals for navigation purposes and/or geodetic positioning. First- and second-generation user equipment has already disappeared from the market, and new models frequently appear. The number of manufacturers is growing fast which makes a complete treatment of makes and models impossible and meaningless within the scope of this book. Consequently only the basic aspects of GPS receivers will be discussed here. A general review is given, including some models for geodesy, surveying, and GIS/navigation currently available.

7.2.1   Receiver Concepts and Main Receiver Components

A GPS receiver detects the signals transmitted from a GPS satellite and converts the signals into useful measurements (observables). The GPS signals, when they arrive at the user antenna, are extremely weak. A particular technique, named *spread spectrum*, is used to transmit and detect the signal information. The name is due to the fact that the power of the signal to be transmitted is "spread" over a much larger bandwidth (e.g. 20 MHz for GPS) than that of the navigation message (50 bps). The *bandwidth* of a signal is the frequency domain in which about 99% of the signal power is transmitted.

For GPS the pseudorandom code sequence (P-code or C/A-code) is used as the spreading function. This technique is also named *binary phase shift keying* (BPSK). In the receiver the spreading function is known, so the signal can be de-spread by correlating the received signal with the locally generated signal. One advantage of the technique is that the signals are quite resistent against disturbances, and can be detected within a noisy environment. It is through this process that rather small antennas can provide the necessary signal-to-noise ratio (SNR) for the GPS receiver (Langley, 1991b). For details of the technique see e.g. Forsell (1991); Leick (1995); Kaplan (1996); Parkinson et al. (1996); Misra, Enge (2001). The SNR is the ratio of the power in the received signal S to the power in the noise level N. The SNR is a logarithmic measure and is given in *decibel* (dB), a dimensionless ratio between electric quantities. The SNR is 1 dB if

$$10 \log \frac{S}{N} = 1. \tag{7.30}$$

The basic components of a generic GPS receiver are (Fig. 7.17):

- antenna with (optional) preamplifier,
- radio-frequency (RF) and intermediate-frequency (IF) "front-end" section,
- signal tracker and correlator section,
- microprocessor for receiver control, data sampling, and data processing (navigation solution),
- oscillator,

– power supply,
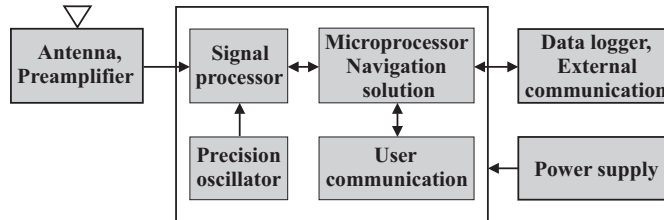– memory, data storage, and
– user interface.

Figure 7.17. Major components of a GPS receiver

The *antenna* detects the electromagnetic waves arriving from the satellites, converts the wave energy into an electric current, amplifies the signal strength and hands the signals over to the receiver electronics. The GPS signal structure requires that all GPS antennas must be right-handed circularly polarized. The antenna has to be very sensitive because of the rather weak satellite signal, and the gain pattern must allow signal reception from all elevations and azimuths of the visible hemisphere.

Further requirements for high precision geodetic applications are a high stability of the electrical phase center [7.4.5.1], and protection against multipath [7.4.4.3]. For applications in navigation (airplanes and ships) signal reception below the antenna's horizontal plane is required. The antenna is connected to the receiver by a coaxial cable through which a voltage is sent to a pre-amplifier at the antenna. The power of the received signal is increased and can then be sent into the receiver.

Several types of GPS antennas are available, e.g. (Fig. 7.18):
– monopole or dipole,
– quadrifilar helix (also named volute),
– spiral helix,
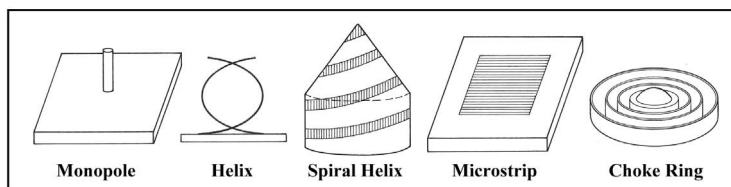– microstrip (also named patch), and
– choke ring.

Figure 7.18. Types of GPS antennas

One of the most frequently used types is the *microstrip*, because it is relatively easy to build. The antenna has a very low profile and is ideal for airborne application. It also meets the increasing demand for miniaturized GPS equipment, in particular when

the antenna is integrated to the receiver body. It is made up of one or more patches of metal and is therefore also named a *patch antenna*. The *quadrifilar helix* is often used for handheld receivers. The spiral helix has nearly disappeared.

Geodetic antennas are usually designed for the reception of both carrier frequencies L1 and L2. They can be protected against multipath by extra ground planes or the use of choke rings. A *choke ring* consists of strips of conductor which are concentric with the vertical axis of the antenna and connected to a ground plate. For more details about GPS antennas see e.g. Langley (1998a).

The incoming GPS signals are down-converted to a lower frequency in the RF/IF section (front end). RF stands for *radio frequency* and IF for *intermediate frequency*. This step is achieved by combining the incoming RF signal with a sinusoidal signal generated by a *local oscillator*. In general, a less expensive quartz oscillator is used because precise clock information is obtained from the GPS satellites, and user clock errors can be eliminated through double-differencing [7.3.2.1]. Some receiver types accept the input of an external high precision oscillator signal from atomic frequency standards with less clock noise [2.2.5]. A very precise oscillator can be used to replace one satellite in the navigation solution.

The IF signal contains all code and data signals from the original RF signal, however its carrier frequency is much lower. In some receivers several IF stages are used to reduce the carrier frequency in steps (Langley, 2000b); bandpass filters are applied to reduce and suppress interference with undesired signals. The IF signal then passes to the signal tracker or correlator. Here the signals coming from all visible satellites are isolated, identified by their codes [7.1.2] and assigned to a particular channel. The receiver channel can be considered to be the primary electronic unit of a GPS receiver. Signal processing within the channel is described in more detail in [7.2.2] and [7.2.3].

A receiver may have one or more channels. In the *parallel channel* concept each channel continuously tracks one particular satellite. A minimum of four parallel channels is required to determine three coordinates and time. With more channels additional satellites can be tracked. Modern receivers may contain up to 12 channels for each frequency and additional channels for multisystem processing.

In the *sequencing channel* concept the channel switches from satellite to satellite at regular intervals. A single-channel receiver must switch to at least four satellites to determine a three-dimensional position. The sequencing rate is asynchronous to the data rate; hence the full satellite message (the data signal) is complete only after several sequences. The receiver needs at least four times 30 seconds before the first "position fix" can be obtained. In most cases *fast sequencing channels* are used, i.e. the switching rate is about one second per satellite. The channels usually have no problems to recover the carrier phase when they return to the same satellite. Difficulties may, however, arise in kinematic applications in particular at high accelerations.

A further variation is the *multiplex technique*. A *multiplexing channel* sequences at a very high speed between different satellites and in some cases both frequencies. The switching rate is mostly synchronous with the navigation message, namely 50 bps or 20 milliseconds per bit.

The navigation message is obtained continuously from all satellites tracked, hence the first fix is achieved after about 30 seconds. Carrier phase measurements are continuous even at high accelerations. In essence, a single hardware channel is used to obtain quasi-simultaneous measurements to all satellites. One advantage of the multiplex technique, when compared with the parallel technique is that channel dependent systematic hardware delays (*interchannel biases*) do not play a role.

In the early years of GPS (until about 1990) it was cheaper to build receivers with single channels. Since then, prices for building channels have dropped rapidly. Current and future GPS receiver architecture will mainly be based on multichannel technology. Multiplex receivers have nearly completely disappeared from the civil market. The capability to track all visible satellites simultaneously is also called *all-in-view* tracking capability. An overview of the different channelization concepts is Fig. 7.19.
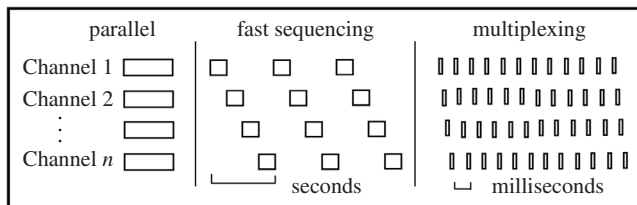

Figure 7.19. Different channelization concepts in GPS receivers

The *microprocessor* (CPU) is necessary to control the receiver's operation, including the acquisition of signals, signal processing, and decoding of the broadcast message. Further capabilities are the computation of on-line positions and velocities, conversion into a given local datum, or the inclusion of DGPS corrections [7.5]. More and more, user relevant software is included on integrated circuits. The microprocessor also controls the input of commands from the user, display of information, and the data flow through a communication port, if one is included.

The microprocessor, the signal tracker/correlator, and the memory, form the digital part of the receiver, whereas the RF/IF front end forms the analog part. At some point of the signal flow, signals are converted from analog to digital. In modern receiver development more and more functions of the receiver are performed by software rather than hardware. Receivers where signal correlation and data processing are integrated in one software controlled unit are also called *software receivers* (Pospelov, Botchkovki, 2000). They are still under development (see [7.2.5], Fig. 7.28).

Modern tendencies are to integrate the RF and IF functions of a GPS receiver on a single *application-specific integrated circuit* (ASIC) and the digital signal processing functions on another ASIC. This two-chip GPS receiver is called a *chipset* (Langley, 2000b).

The *power supply* was a rather critical issue for first-generation GPS receivers because of their very high power consumption. In some cases, the use of generators was necessary for field operation. Modern receivers use low voltage DC power and are designed to consume as little energy as possible. Most have an internal rechargeable nickel-cadmium or lithium battery in addition to an external power input. Depending on the observation rate, the internal battery may be sufficient for weeks, or more, of observation. The modern chipset has a power consumption of less than 1 Watt.

For post-processing purposes all data have to be stored on internal or external *memory* devices. Post-processing is essential for multi-station techniques (e.g. for some geodetic and surveying application) but also for off-line differential navigation. Pseudoranges, phase data, time and navigation message data have to be recorded. Depending on the sampling rate the amount of GPS data to be recorded may be very high. With six satellites and 1-second data a dual frequency receiver produces about 1.5 Mbyte of data per hour. Modern receivers have internal solid state (RAM) memories or removable memory cards. In general, data can also be recorded onto an external micro-computer (e.g. a laptop) connected to the receiver with a RS-232 or equivalent communication port, or they can be transmitted to a base station via an appropriate data link.

Most receivers have a keypad (often handheld) and a display for *communication* between the user and the receiver. The keypad is used to enter commands, external data like station number or antenna height, or to select a menu option. The display indicates computed coordinates, visible satellites, data quality indices and other suitable information. Current operation software packages are menu driven and very "user friendly". Developments in this respect are rapid.

GPS receivers can be divided into various groups according to different criteria. One early classification was into *code-dependent receiver technology* and *code-free receiver technology.* This kind of division is no longer meaningful because, usually, different types of techniques are implemented in each receiver. Hence it would be better to distinguish between

- − code-dependent signal processing,
- − codeless signal processing, and
- − semi-codeless signal processing.

These technologies are presented in the next two chapters [7.2.2], [7.2.3]. Another classification criterion is the available data-type, and differentiates receivers with

- − C/A-code,
- − C/A-code + L1 carrier phase,
- − C/A-code + L1 carrier phase + L2 carrier phase, and
- − C/A-code + P-code + L1, L2 carrier phase.

Note that the P-code under AS [7.1.6] is changed to the encrypted Y-code. Another distinction is related to the technical realization of the channels:

- − multi-channel receiver,
- − sequential receiver, and

    − multiplexing receiver.

Finally a classification is possible with respect to the user community, e.g.:
    − military receiver,
    − civilian receiver,
    − geodetic/surveying receiver,
    − navigation receiver,
    − timing receiver,
    − spaceborne receiver, and
    − handheld receiver

In addition, receiver units form parts of complex systems, e.g. in modern GIS applications or for machine control (see [7.6.2]). All the above classifications appear in the literature, and in technical discussions. For geodetic applications it is essential to use the carrier phase data as observables [7.3.1]. It is equally essential to use both frequencies (L1, L2), and to have access to the full wavelength on L2.

   For an introductory discussion of receiver technology see e.g. Langley (1991a, 2000b), Van Dierendonck (1995). More detailed information can be obtained from the excellent handbooks by Parkinson et al. (1996); Kaplan (1996); Misra, Enge (2001), and from the proceedings of GPS conferences like *ION GPS*.

7.2.2 Code Dependent Signal Processing

The pseudorange from code measurements is the fundamental observable in a code dependent receiver channel. The "phase position" of the received code sequence is compared with the phase of an identical code replica, generated by the receiver, via a correlation technique (cf. [7.3.1.2]). This is why the observable may be called the *code phase*. The user must have a priori knowledge of the code, i.e. the code must be generated within the receiver channel using the same algorithm that is utilized in the satellite. The received code sequence and the generated code sequence are correlated with each other, i.e. the two sequences are shifted stepwise in phase until maximum correlation is obtained.

   This process happens in one of the two tracking loops, namely the *delay lock loop*, or code tracking loop. The time shift that is necessary to align both code sequences with each other (*time delay*) corresponds to the signal travel time between the satellite and the receiver. Essentially, the code sequence is a unique function in time and hence provides us with a reading of the satellite clock at the moment a particular bit left the satellite. The time delay is converted into a range using the speed of light (cf. [7.3.1.2]). The pseudorange measurements are derived from either the P-code or the C/A-code.

   The second tracking loop is the *phase lock loop*, or carrier-tracking loop. Here the code and the carrier are separated to enable phase measurements to be made, and the bit information of the satellite message is extracted. This technique is also named *reconstruction of the carrier*. In most cases a *Costas loop*, that is specially designed for biphase modulated signals, is used (cf. Forsell (1991); Kaplan (1996)). Within the

loop, the demodulated satellite carrier phase signal is aligned with the phase signal of the receiver's oscillator. The observable is the *carrier beat phase*, the relative phase between the received carrier signal and the internal reference carrier signal derived from the local oscillator (cf. [7.3.1.2]). The total process of code-dependent signal processing is schematically illustrated in Fig. 7.20 (cf. Wells (ed.), 1986).
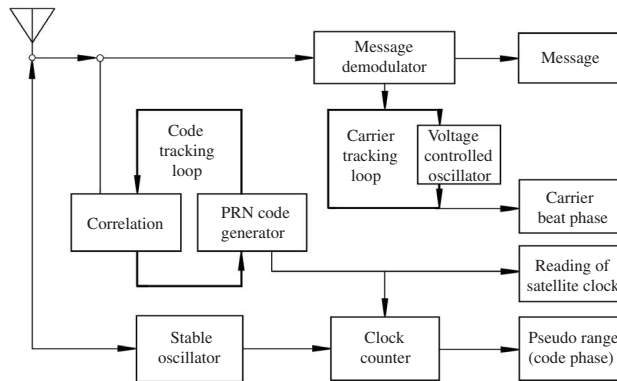

Figure 7.20. Simplified concept of a code correlation channel

A complete code-dependent correlation channel produces the following observables and information:
- code phase,
- carrier phase,
- change of carrier phase (Doppler frequency), and
- satellite message.

The code correlation and carrier reconstruction technique only works on L2 when the P-code is available (AS not activated), or for authorized users with access to the Y-code.

7.2.3   Codeless and Semicodeless Signal Processing

Codeless GPS channels exploit satellite signals without knowledge of the codes. The advantage of this concept is that the receiver systems are independent of possible restrictions on code access to civil users. The main disadvantage is that neither the broadcast ephemeris nor the almanac and precise time can be extracted from the signals; consequently alternative sources are required for mission planning and for data processing. A further disadvantage is that simultaneously-operating pure codeless receivers have to be synchronized before observation starts.

Today there is no doubt that at least the C/A-code will be freely available to all civil users. This is why no more totally code-free receivers are built. The technique

is, however, of much interest for L2 access in times of P-code denial (AS activated, cf. [7.1.6]).

A frequently used codeless method is the *squaring technique*. A *squaring channel* multiplies the incoming satellite signal by itself and generates a second harmonic of the original carrier; both the codes and the broadcast message are lost. Squaring the code signal portion in equation (7.2),

$$x = P(t) \sin(\omega t), \tag{7.31}$$

yields

$$x^2 = P(t)^2 \sin^2(\omega t) = P^2(1 - \cos 2\omega t)/2. \tag{7.32}$$

Since $P(t)$ is a sequence of $+1$ and -1, representing the code, it follows that $P(t)^2 = P^2$ is a sequence of $+1$, and thus disappears from equation (7.32).

With x^2 a pure carrier signal is obtained, that is related in phase to the original carrier, but with a frequency equal to twice that of the original. The signal-to-noise ratio (SNR) is also considerably decreased in the squaring process. Consequently, the squared phase measurements have the two disadvantages that the wavelength is cut in half and that the SNR of the phase observable is made much worse. The squaring technique was developed early on (Counselman, Steinbrecher, 1982) and used for the first time in the *Macrometer* [7.2.4.1]. Today, some commercial dual-frequency receivers use this type of approach for carrier phase measurements on L2. A schematic illustration of the squaring technique is given in Fig. 7.21.
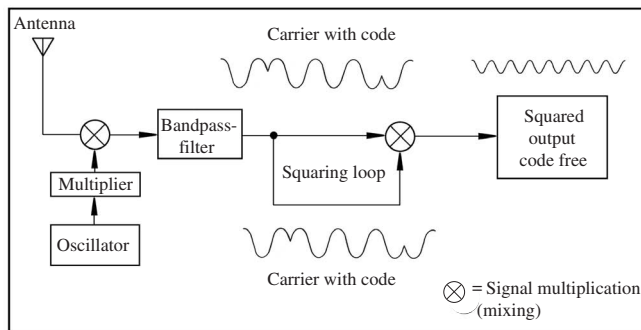
Figure 7.21. Concept of a squaring channel

Another concept uses the pure *interferometric principle*, known from Very Long Baseline Interferometry (VLBI) [11.1]. It was first proposed by Mac Doran (1979). In this method the GPS signals are recorded at two stations, together with precise timing signals from an external oscillator, but without knowing the code sequence. From a subsequent correlation process the time difference, τ, in the signal reception

at both stations is determined and transformed into a range difference (cf. [7.3.1]). The complexity of the instrumentation, necessary for signal reception and processing, means that the pure interferometric principle has never been applied in practice.

A variation of this technique uses certain periodic structures in the code sequence in order to identify the bit-rate of the satellite clock and to reconstruct the phase of the code modulations without having a knowledge of the code. This method is sometimes referred to as *ambiguous code phase observation* (Wells (ed.), 1986). The technique has been realized e.g. in the *SERIES* concept (*S*atellite *E*mission *R*adio *I*nterferometric *E*arth *S*urveying) (Mac Doran, 1983). One disadvantage of the last technique was that the observation noise was rather high.

Based on the above-mentioned experiences, powerful new techniques of using the L2 signals while the P-code is encrypted have been developed. These techniques are not completely codeless because they use the C/A-code on L1 for aiding the L2 tracking loops, taking advantage of the fact that both L1 and L2 have the same P-code modulation/encryption. They are hence called *semicodeless* techniques (Van Dierendonck, 1995). As a result, code and full carrier phase measurements are available for L1 and L2 even under activated AS. Modern receiver developments use and combine some of the above mentioned features to give as many usable and precise signals as possible [7.2.4.2]. The objective is to provide full wavelength for L2 with as low a noise level as possible. The semicodeless technique, however, always generates signals with increased noise; the ideal solution remains the availability of a non-encrypted code sequence on L2 and the use of the code correlation technique [7.2.2]. This will be realized with the L2C signals on the Block IIR-M satellites after 2003 [7.1.7].

Summarizing, we have the following possibilities for providing carrier phase observables on L2 (for details see e.g. Breuer et al. (1993); Van Dierendonck (1995); Hofmann-Wellenhof et al. (2001), p.81ff):

Squaring technique
Based on developments of Counselman, Steinbrecher (1982). Results in a signal of the half wavelength $\lambda_{L2}/2$; SNR reduced by 30 dB when compared to the ideal code correlation technique.

Cross correlation technique
Based on developments of Mac Doran (1983). Uses the fact that the encrypted Y-code is identical on both carriers. The cross correlation of both signals provides the difference in propagation time caused by the ionosphere [2.3.3.1] and results in the L2 code range and phase by

$$R_{L2} = R_{L1,C/A} + (R_{L2,Y} - R_{L1,Y}) \tag{7.33}$$
$$\Phi_{L2} = \Phi_{L1,C/A} + (\Phi_{L2} - \Phi_{L1}). \tag{7.34}$$

Compared to the ideal code correlation the reduction in SNR is 27 dB.

Code aided squaring:
This technique (Hatch et al., 1992) makes use of the fact that the Y-code is generated from the P-code by multiplication with the unknown W-code. The chipping rate of

the W-code is 20 times less the frequency of the Y-code, i.e. there are Y-code portions identical to the P-code. This property is used to despread portions of the received Y-code, and to enter with a much lower bandwidth into the squaring process. As a result, the output signal still has half the wavelength, however the SNR is 13 dB better than with the unaided squaring technique.

P-W tracking

The technique, also called *Z-tracking*, has been described by Ashjaee, Lorenz (1992) and uses the undisturbed portions of the P-code (about 2 μs length) in the Y-code signal on both carriers for cross-correlation with the P-code replica. Output signals are code ranges and full wavelength carrier signals on both L1 and L2. The SNR reduction compared with the ideal signal on L2 is just 14 dB.

All leading manufacturers of geodetic receivers have developed proprietary semi-codeless techniques for measurements under activated AS, using some of the above described or modified concepts.

7.2.4 Examples of GPS receivers

The receiver market is developing and growing with high speed. This is why only little space is given here to describe some typical receivers. The description can be regarded as a basis for the evaluation of current and future developments.

7.2.4.1 Classical Receivers

Two examples of receivers are given that have significantly influenced the development in geodetic GPS technology, the code-dependent *TI 4100* and the code-free *Macrometer*.

The Texas Instruments *TI 4100 GPS Navigator* (Fig. 7.22) came to the market in 1984. It was the first GPS receiver providing all observables of interest to the geodesist, surveyor and navigator. The TI 4100 is a dual frequency multiplexing receiver and can track up to four satellites simultaneously. The observables are:

- P-code pseudoranges on L1 and L2,
- C/A-code pseudoranges on L1, and
- carrier phases on L1 and L2

every three seconds. The data are recorded by an external tape recorder on digital cassettes, or are downloaded directly to an external microprocessor. Communication between observer and receiver is by a hand-held *control display unit* (CDU). For navigational purposes, the built-in microproces-

Figure 7.22. Texas Instruments TI 4100

sor provides positions and velocities in real time every three seconds. The equipment is rather bulky and can be packed into two transportation cases. The total weight is about 30 kg, and power consumption is about 100 Watts. The observation noise was found to be 0.6 to 1 m for P-code tracking and 2 to 3 mm for carrier phase observations.

The equipment has been widely used in numerous scientific and applied GPS projects. Many of the results published between 1985 and 1993 are based on observations with the TI 4100. This is also true for some of the examples within this book. Today, with activated anti-spoofing (AS), the TI 4100 can only be used as a single frequency C/A-code receiver. For applications and results see e.g. Seeber et al. (1985); Seeber (1989a); Gibbons, Maynard (1990); Jahn et al. (1991); Völksen (2000).

The *Macrometer V1000* (Fig. 7.23) was introduced in 1982 and was the first GPS receiver for geodetic applications. The exciting results obtained with this system have

Figure 7.23. Macrometer V1000

done much to demonstrate the potential of highly accurate GPS phase observations. The complete system consists of 3 units:
   − receiver/recorder with power supply,
   − antenna with large ground plane, and
   − the P 1000 processor.
The processor is essential for providing the almanac data (because the Macrometer cannot decode the satellite message) and to preprocess the data. The Macrometer V1000 is a single frequency receiver, and tracks up to 6 satellites on 6 parallel channels. At predetermined epochs the phase difference between the received (squared) carrier signal and a reference signal, taken from the receiver oscillator, is measured. Usually, 60 epochs are evenly distributed over the total observation period of several hours. A typical baseline accuracy over up to 100 km was found to be 1 to 2 ppm (Bock et al., 1986). A dual frequency version, the *Macrometer II*, was introduced in 1985 (Ladd et al., 1985). The architecture is comparable to the V1000, however the weight and the power consumption are much less.

Both Macrometer systems require the availability of external ephemerides. They were hence mainly operated by a restricted number of specialized companies. Another

disadvantage is that all instruments that participate in the same observation session have to be collocated, prior to observation, for clock synchronization. Because of these disadvantages, the dual-frequency macrometer was further miniaturized and combined with a single-frequency C/A-code receiver, resulting in the *MINIMAC* (Ladd et al., 1986).

The Macrometer V1000 and its successors were used extensively for a number of years, in particular for the establishment of geodetic control.

7.2.4.2 Examples of Currently Available Geodetic Receivers

The currently available GPS receivers that are used in geodesy, surveying, and precise navigation all contain several or all of the above-mentioned features. Nearly all models started as single frequency C/A-code receivers with the ability to track more than four satellites. In a second step access to L2 was added, using the squaring technique, and the number of satellites that could be tracked simultaneously was increased. In the third step, all leading manufacturers added the P-code on L2, and some on L1 and L2, mainly with the objective to increase the L2 data quality and to provide the full cycle length on L2. The next step was the inclusion of codeless, non-squaring L2 techniques in order to provide high quality L2 full wavelength signals under activated AS.

In the most recent development step all manufacturers further improved the data quality, designed rugged, light portable units with low power consumption, and integrated the GPS receiver into single-module multipurpose compact surveying equipment. In general, a high precision GPS system should fulfill the following requirements:
  − track all signals from each visible satellite at any given time (GPS-only system requires 12 dual frequency channels; GPS + GLONASS system needs 20 dual frequency channels),
  − full wavelength on L2, when AS activated,
  − low code and carrier phase noise,
  − high memory capacity for data storage,
  − high data rate ($\geq$ 10 Hz) for kinematic applications,
  − low power consumption (below 4 Watts); low weight (below 4 kg), small size,
  − track weak signals (under foliage and difficult environmental conditions),
  − multipath mitigation, interference suppression, stable antenna phase center,
  − modular hardware, easy to upgrade,
  − powerful on-board and office software, and
  − accept user commands and display results via Control Display Unit.

In addition some of the following features are helpful for flexible and unrestricted applications:
  − 1 pps timing output,
  − event marker,
  − ability to accept external frequencies,
  − fast data transfer to a computer,

- few or no cable connections,
- radio modem,
- DGPS and RTK capability [7.5],
- operate over large temperature range and in driving rain,
- easy interfacing to other systems, also from other manufacturers,
- ease and flexibility of use (multi-purpose application), and
- flexible setup (tripod, pole, pillar, vehicle).

To summarize, a modern GPS survey system should measure accurately and reliably anywhere under any condition, and it has to be capable of being used for almost any application (geodetic control, geodynamics, detailed GIS and topographic survey, stake out, engineering, hydrographic survey etc.), cf. [7.6.2].

It is impossible and not meaningful to review all available GPS receivers that are currently on the market. The annual "Receiver Survey" (*GPS World*, January) includes currently more than 500 makes and models. In the following only some of the leading products are summarized as examples. The order of discussion within this text does not reflect any priority in quality or performance.

Wild/Leitz (Heerbrugg, Switzerland) and *Magnavox* (Torrance, California) developed, in a joint venture, the *Wild/Magnavox WM 101* geodetic receiver, which appeared on the market in 1986 as a four channel L1 C/A-code receiver. The dual frequency *WM 102* followed at the end of 1988. One key feature of the WM 102 was a modified squaring technique for receiving L2 when P-code signals were encrypted.

In 1991 the company introduced a follow-up model, the *Wild GPS-System 200* (with the Magnavox SR299 dual-frequency GPS sensor) tracking up to 9 satellites on L1 and L2 using code-aided squaring for L2 with activated AS. In 1995 followed the Leica GPS System 300 with RTK capability. The last system released by the company (now named *Leica Geosystems*) was the *GPS System 500*, in 1998.

The System 500 family (Fig. 7.24) has a modular design and comprises three different receivers: SR510 (single frequency), SR520 (dual frequency), and SR530 (dual frequency, RTK). Various configurations are possible: tripod, pole mounted or backpack mounted. The dual frequency sensors have 24 parallel channels. A proprietary code-aided tracking technique provides full-wave L2 carrier phase measurements (AS on) and high accuracy L2 pseudoranges. Enhanced new multipath mitigation and interference rejection techniques are reported (Maenpa et al., 1997). Up to 85 MB data storage is possible with a PCMCIA card. The RTK sensor can be connected to a radio modem or to GSM phones [7.5.2]. The software platform for postprocessing is SKI Pro.

For precise navigation and machine guidance the MC1000 (Machine Control) receiver, also with 24 parallel channels, has been designed. Kinematic applications are supported with very little latency and 10 Hz data rate. A variation of the MC1000 is known as CRS1000 (continuous reference station) and is operated with a choke ring antenna and special control software. The product line of Leica Geosystems, as for other manufacturers, hence supports a broad variety of applications in geodesy and surveying.

Figure 7.24. Leica System 500; cour-
tesy Leica Geosystems

Figure 7.25. Trimble GPS Total Station
5700; courtesy Trimble Navigation

Trimble Navigation (Sunnyvale/California) has been producing its *Trimble 4000*
series since about 1985. The first generation was a L1 C/A-code receiver with five
parallel channels, capable of tracking up to five satellites simultaneously. Further
upgrades included increasing the number of channels to twelve, L2 squaring capability,
and P-code capability. The most advanced model of this series, for geodetic purposes
was the *Trimble Geodetic Surveyor 4000 SSi*, with 12 channels dual frequency and
proprietary codeless technique for full cycle carrier on L2 with activated AS. This
instrument has been used for many years worldwide in numerous geodetic projects for
a large number of applications.

In 2001 Trimble, now including Spectra Precision, Geodimeter and Zeiss Geodetic
Systems, launched the GPS 5700 family. The *5700 GPS receiver* has 24 channels L1
C/A code and L2 full cycle carrier phase. The advanced GPS chip provides very low
carrier phase noise and multipath mitigation. The *GPS Total Station 5700*, a modular,
kinematic real-time surveying system (Fig. 7.25) includes the receiver, a handheld
controller, GPS antenna, RTK radio and processing software. Its weight is 3.8 kg
and power consumption, including radio, is 3.8 Watts (the receiver alone weighs 1.4
kg and consumes 2.5 Watts). Based on a newly developed receiver concept, Trimble
offers *Continuously Operating Reference Stations* (CORS), *Virtual Reference Stations*
(VRS) and other solutions. Internal data storage with compact flash card is up to 128
MB. For continuous operation the 5700 GPS receiver can be linked to a choke ring
antenna or the geodetic *Zephir antenna*, designed for phase center stability, enhanced
multipath rejection, and low elevation satellite tracking.

Ashtech (initially Sunnyvale, California) was founded by Javad Ashjaee in the late
1980s. He was the first to announce a receiver with 12 parallel channels and hence
initiated the development of the current multi-channel technology. The *Ashtech XII
GPS receiver* entered the market in 1988 and was capable of measuring pseudorange,

carrier phase, and integrated Doppler of up to 12 satellites on L1. The L2 option added 12 physical L2 squaring type channels. In 1991 the *Ashtech P-12 GPS receiver* was launched, providing 12 dedicated channels of L1 P-code and carrier and 12 dedicated channels of L2 P-code and carrier. Together with the 12 L1 C/A-code and carrier channels and the 12 codeless L2 channels, the receiver contains, in total, up to 48 channels. All observables of interest to the geodesist and from all visible satellites are available; the reconstructed carrier phase data (full wavelength) on L2 uses the Z-tracking technique [7.2.3], Ashjaee (1989).

The latest version, based on the P-12 technology, is the *Ashtech Z-Surveyor*, a light weight, compact dual-frequency receiver with RTK capability. The receiver tracks up to 12 satellites and provides full wavelength on both carriers. Like the high-end products of other manufacturers the Z-Surveyor can be configured with dedicated software for a large variety of applications. In 1997 Ashtech Inc. was purchased by *Magellan Corporation*, a leading manufacturer of handheld GPS equipment and later on unified with *Thales Navigation*, a French manufacturer of navigation equipment. This group now offers every type of GPS receiver.

In 1996 Javad Ashjaee founded *Javad Positoning Corporation* (JPS), later (July 2000) bought by *Topcon Positioning Systems*. High-end products combined GPS-GLONASS receivers with up to 40 channels. One example is the *Odyssey-E GGD*, launched in 2001, with 20 channels L1, C/A- and P-code; 20 channels L2, P-code, either for GPS or GLONASS satellites. Its weight is about 2 kg and its power consumption 4 Watts. The internal memory capacity is up to 96 MB, and the data recording rate up to 20 Hz.

Another notable development is the *Rogue receiver*, from Allan Osborne Associates, originally a dedicated development for NASA's applications in geodynamics. The *TurboRogue SNR-8100*, launched in 1993, is still widely used in the scientific geodetic community. The instrument has 24 channels and can track up to 8 satellites on L1 C/A- and P-code and on L2 P-code. In the presence of encrypted P-code the codeless mode produces L1–L2 group and phase delay data (cf. [2.3.1.2], Meehan et al. (1992)). Many such instruments are operated at IGS stations [7.8.1] .

7.2.4.3   Navigation and Handheld Receivers

A very large market is rapidly developing for navigation with handheld receivers. In some cases, a single C/A-code sequencing or multiplexing channel is used, however, modules with 12 parallel channels are becoming increasingly popular. Positions and velocities are derived from L1 C/A-code pseudorange measurements, and are displayed, or can be downloaded, via a RS 232, or equivalent, port. Usually neither raw data nor carrier phase information is available. Differential navigation is possible with some advanced products, also carrier smoothed code phases are used. Power consumption is in the order of 1 Watt or less. Weight is significantly below 1 kg, and reaches in some cases only several hundred grams. Most models come with a map display. Some can be interfaced with map sources on CD-ROM. Several classes of products can be distinguished:

(a) accuracy class ≤ 10 m,

(b) accuracy class 2 to 5 m, and

(c) accuracy class below 1 m to several dm.

Class (a) receivers are the typical low-cost handheld receivers for recreation (sports, hiking, sailing), general navigation, surveillance, fleet management, and GIS applications with moderate accuracy requirements (Fig. 7.26). Since SA has been switched off, the accuracy range of up to 10 m or even better makes this instrument a powerful tool for many applications. With the ongoing *Accuracy Improvement Initiative* (AI I) [7.1.7] the accuracy level of class (b) may even be reached. For this receiver type a mass market with very low unit prices is developing. Receivers are built on one or two chips (chipset). With the reduced size and very low-power consumption they can easily be integrated into mobile phones, palmtop computers or even wrist watches, and they are well suited for *location-based services* [7.6.2.5].

Figure 7.26. Handheld receiver Garmin eTrex; © GARMIN Corp. 2003

Figure 7.27. Trimble Pathfinder ProXR; courtesy Trimble Navigation

Class (b) receivers use L1 C/A code and accept DGPS data [7.5.1]. They are ideal for many real-time GIS applications (forestry, farming, environmental monitoring) and may be linked to GIS and CAD software packages. They are also suited to precise car navigation, fleet management and traffic control.

Class (c) receivers use, in addition, the L1 carrier measurements to filter the code phases without solving ambiguities [7.3.6]. Depending on the observation time, the DGPS *carrier smoothed pseudo ranges* provide submeter accuracy on a second-by-second basis, about 30 cm after 5 minutes, and about 10 cm after 20 minutes. Some models offer a data logging mode. A typical example is the *Trimble Pathfinder Pro XR* (Fig. 7.27). Together with dedicated software the system is, among others, suited for GIS, precision farming, fleet management and mining control.

Several manufacturers offer GPS cards or chips for integration into other systems. The developer needs to provide a power supply, antenna, preamplifier, an application processor and user software. Most of such modules deliver position and time; some developments also include raw carrier phase data, and hence provide a powerful basis for the configuration of a broad variety of application-orientated systems.

Latest developments in the fast growing market of handheld receivers are announced in GPS related periodicals such as *GPS World*, and also in the GIS literature. A very profound source is the World Wide Web. Current web-addresses are listed in the annual *GPS World* Receiver Survey.

7.2.5 Future Developments and Trends

The GPS market will grow further. More than 1 million GPS receivers are produced each year, and GPS applications are developing fast. The worldwide market for GPS applications and services was estimated to have reached nearly 20 billion U.S. Dollars by 2001, and is projected to grow to about 60 billion U.S. Dollars by 2005 (Groten et al., 2001). This number will certainly further increase with the forthcoming European system GALILEO. Geodesy and surveying, although the most challenging part and driving force, will only cover a small section of the global market. The most important portions will certainly come from car navigation and traffic control, together with communication services.

Selling and buying of companies will continue, and hence the names of receiver makes and models will change rapidly. Probably there will be only few manufacturers of GPS or GNSS hardware chips; the great variety in makes and models comes from the configuration of sensors and the application software.

With respect to electronics and *signal processing* the trend is toward the *chipset*, i.e. a complete GPS receiver on one or two chips, and to *software receivers*, i.e. as much digital signal processing as possible. In a final step the digital data processing could start directly behind the antenna (*digital radio*). Software receivers are much more flexible and economic than the current hardware-based data processing components (Fig. 7.28).
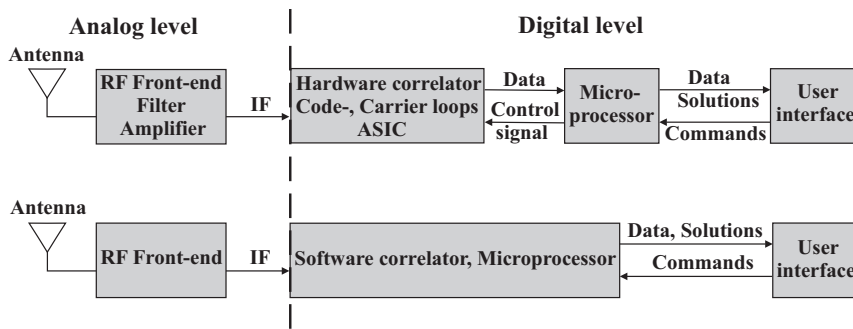

Figure 7.28. Modern GPS receiver architecture; top: current situation; bottom: software receiver

Future satellites will broadcast additional signals. Hence the demand for multi-frequency receivers will increase. Receivers will simultaneously track signals from GPS, GLONASS and GALILEO. Suitable signals for specific applications will be selected automatically by the receiver. The new signals equally provide enhanced cross correlation properties and carrier tracking with much lower SNR. Operation under foliage and other suboptimal conditions, as well as indoor operation, will be facilitated. The semicodeless techniques, however, will remain an important feature as long as not enough satellites with the new signals are in orbit, i.e. until about the year 2008.

With respect to the user market we can distinguish between two different development lines. For the *navigation receiver* we will see further miniaturization and fall in prices. The one-chip receiver can be integrated into a wearable computer built into clothes (Langley, 2000b). The limiting factor could be the size of the antenna. However, with improved signal processing even the antenna could be as small as 1 cm $\times$ 1 cm. Hence, a GPS receiver could become small enough to be implemented under the skin and, with a suitable data link, be used to track people and transmit critical health data. This concept could also be a powerful protection against criminal attacks. Issues of morality and personal freedom are certainly touched by such developments.

For *surveying receivers* the development lines are different. They are also likely to decrease in size; however the price level most probably will remain rather high compared to the price level of low-cost navigation receivers. The main reasons are:

— the market segment is rather small,
— the software development costs for a surveying receiver are much higher than for navigation receivers, and
— the capacity and complexity of surveying receivers is continuously growing.

Receivers will become more "intelligent". Preprocessing of the data within the receiver already includes automatic and remotely controlled continuous operation, cycle slip editing, and data compression. The RTK option will be a standard feature. Antennas will be of improved design and increased flexibility, so that one generic antenna serves for all kinds of applications.

The tendency of manufacturers is to provide complete and integrated sets of surveying tools, with the GPS receiver as one sensor among others, and "plug-in solutions" with no interfacing problems. The *surveying toolbox* includes for example (cf. [7.6.2]):

— GPS based products (GPS reference stations, GPS total station),
— conventional instruments,
— digital levels,
— robotic total station,
— data collector, field computer, and
— office and field software.

7.3   GPS Observables and Data Processing

7.3.1   Observables

7.3.1.1   Classical View

Four basic observables can be identified:
  - pseudoranges from code measurements,
  - pseudorange differences from integrated Doppler counts,
  - carrier phases or carrier phase differences, and
  - differences in signal travel time from interferometric measurements.

A *pseudorange from code measurements* equals the time shift that is necessary to correlate the incoming code sequence with a code sequence generated in the GPS receiver, multiplied by the velocity of light [7.1.4], [7.3.1.2]. The fundamental observation equation for a single pseudorange is

$$PR_i = |X_i - X_B| + cdt_u = c\tau_i$$
$$= ((X_i - X_B)^2 + (Y_i - Y_B)^2 + (Z_i - Z_B)^2)^{\frac{1}{2}} + cdt_u,$$

(7.35)

with the notations from Fig. 7.29:

R_i geometrical distance (slant range) between satellite antenna S_i and receiver antenna B,

X_i satellite position vector in the geocentric CTS [2.1.2] with the components X_i, Y_i, Z_i,

X_B position vector of the receiver antenna B in the CTS with the components X_B, Y_B, Z_B,

τ_i observed signal propagation time between satellite antenna S_i and observer antenna B,

dt_u clock synchronization error between GPS system time and receiver clock, and
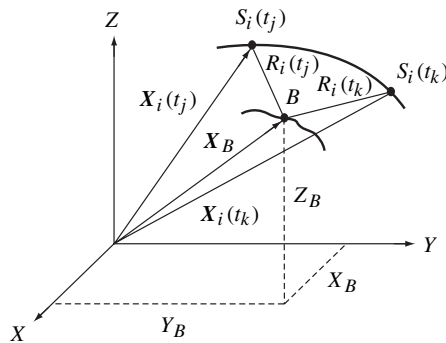
c signal propagation velocity.

Figure 7.29. Geometric relations in satellite positioning

The coordinates of the observer antenna X_B can be derived from simultaneous range measurements to four satellites (Fig. 7.30(a)).
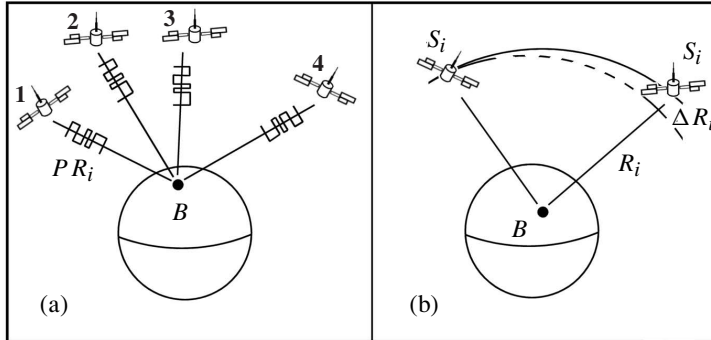

Figure 7.30. Observation concept with pseudoranges (a) and pseudorange differences (b)

Pseudorange differences can be derived from observations of the Doppler shift of the received carrier frequency, in a similar way to that done with the TRANSIT NNSS system (Fig. 7.30(b), cf. [4.2.3], [6.2]). The frequency shift of the received carrier, f_r, is measured with respect to a reference frequency, f_g, within the receiver, and yields the integrated Doppler count:

$$N_{jk} = \int\limits_{t_j}^{t_k} (f_g - f_r)\, dt. \tag{7.36}$$

N_{jk} is a measure of the range difference between the receiver antenna, B, and two consecutive orbital positions of the same satellite, S_i, at two different epochs, t_j, and t_k, (Fig. 7.29). The related observation equation is, see (6.9):

$$\Delta R_i = |X_i(t_k) - X_B| - |X_i(t_j) - X_B| = \frac{c}{f_0}(N_{jk} - (f_g - f_s)(t_k - t_j)), \tag{7.37}$$

with

  f_g a reference frequency generated in the receiver,
  f_s the frequency of the signal emitted by the satellite antenna, and
  f_r the frequency of the signal received at the observer antenna.

It is also possible to measure the Doppler shift of the codes. The resolution is, however, very poor because of the low code frequency when compared with the carrier frequency. The integrated Doppler measurement (7.37) should not be confused with the instantaneous Doppler measurement, used in velocity determination with navigational receivers [7.6.2.7].

The *carrier phase* is derived from a phase comparison between the received Doppler-shifted carrier signal, f_{CR}, and the (nominally constant) receiver-generated reference frequency, f_0. The proper observable is then the measured phase difference

$$\Phi_B = \Phi_{CR} - \Phi_0. \tag{7.38}$$

With

λ the carrier wavelength,

N_{B_i} the integer number of complete carrier cycles within the range R_i, and

dt_u the clock synchronization error,

the fundamental observation equation of carrier phase measurements (cf. Fig. 7.31(a)) follows thus:

$$\Phi_{B_i} = \frac{2\pi}{\lambda}(|X_i - X_B| - N_{B_i}\lambda + cdt_u). \tag{7.39}$$

The main difficulty related to this method is the determination of the cycle ambiguity, N_{B_i}, because the observable only determines the phase within one wavelength. The *ambiguity term*, N_{B_i}, has to be determined with appropriate methods [7.3.2.3], or the range must be known a priori with an accuracy corresponding to the half cycle length ($\sim$10 cm).
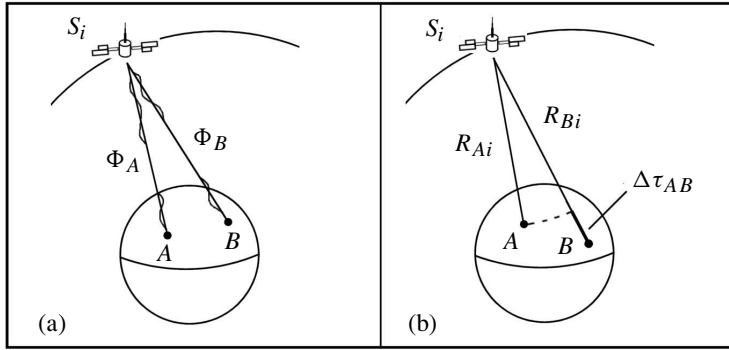

Figure 7.31. Observation concept with phase differences (a) and interferometric time differences (b)

Frequently, the phase difference of the same satellite signal, observed at two stations A and B, is considered as the basic observable. The observation equation of the *single phase difference* (or in short, *single difference*) is:

$$\Delta\Phi_{AB_i} = \Phi_{B_i} - \Phi_{A_i} = \frac{2\pi}{\lambda}(|X_i - X_B| - |X_i - X_A|) \\ - (N_{B_i} - N_{A_i})\lambda + c(dt_{u_B} - dt_{u_A}). \tag{7.40}$$

The original observable in (7.39) is, for clarity, sometimes referred to as *undifferenced phase observable* or *zero phase measurement*.

In the case of pure *interferometric* observations (Fig. 7.31(b)) the GPS signals are used without knowledge or use of the signal structure (cf. [7.2.3]). The signals are recorded, together with precise time marks, at least at two stations, A and B, and are then correlated. The fundamental observable is the difference, $\Delta\tau_{A,B_i}$, between the signal arrival times at both stations with respect to a particular satellite S_i. The observable can be scaled into a range difference, $R_{B_i} - R_{A_i}$. The observation equation is:

$$\Delta\tau_{A,B_i} = \frac{(R_{B_i} - R_{A_i})}{c} = \frac{(|X_i - X_B| - |X_i - X_A|)}{c} + (dt_{u_B} - dt_{u_A}). \quad (7.41)$$

The technique is very similar to *Very Long Baseline Interferometry* (VLBI) which uses radio signals from Quasars [11.1].

Two of the four observables mentioned above are never or only rarely used in applied geodesy. The integrated Doppler count requires a rather long observation time (several hours) to allow the satellite configuration to change sufficiently, and it requires very stable oscillators in the user segment. The method is, however, implicitly used for the determination of ambiguities (cf. [7.3.2.3]). The genuine interferometric technique requires highly sophisticated instrumentation and high data processing expenditure.

In practice, therefore, only two fundamental observables are used that can be regarded as measurements of pseudoranges:
 – code phases (pseudoranges from code observations), and
 – carrier phases (pseudoranges from carrier observations).

7.3.1.2   Code and Carrier Phases

The main characteristics and differences between the two observables are summarized in Table 7.6. Signal propagation and the observation procedure are discussed below in more detail. Fig. 7.32 illustrates the signal propagation of the code and carrier phases, in which

T	is the satellite time (time system of the individual space vehicle),
t	is the receiver time (subscript RCV),
subscript t	refers to the transmitted signal,
subscript r	refers to the received signal,
f_{CD}	is the code frequency, and
f_{CR}	is the carrier frequency.

The signal states (phases) at the epoch of transmission are:

$$\begin{aligned} \text{Code} \quad &\Phi_{CD}(T_t) = T_{t_{SV}} f_{CD}, \\ \text{Carrier} \quad &\Phi_{CR}(T_t) = T_{t_{SV}} f_{CR}, \end{aligned} \quad (7.42)$$

according to the fundamental relation (2.80) between phase, frequency and time.

Table 7.6. Main characteristics of code- and carrier phases

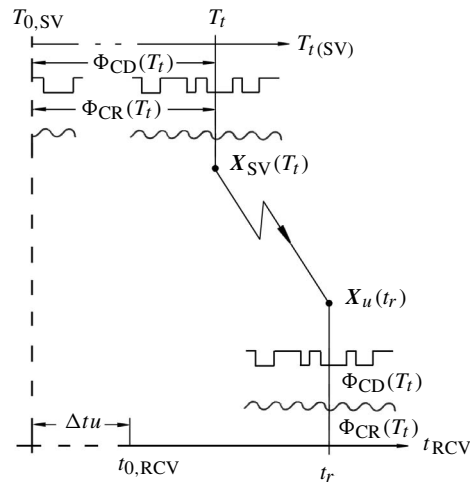|	Code	Carrier
wavelength	P-code 29.3 m	L1 19.0 cm
	C/A-code 293 m	L2 24.4 cm
observation noise	P-code 0.3–1 m	
	C/A-code 3–10 m	1–3 mm
new development	P-code cm–dm	
	C/A-code dm–m	< 0.2 mm
propagation effects	ionospheric delay $+\Delta T_{\text{ION}}$	ionospheric advance $-\Delta T_{\text{ION}}$
ambiguity	non-ambiguous	ambiguous

Figure 7.32. Signal propagation of code and carrier phases

The carrier phase, $\Phi_{\text{CD}}(T_t)$, leaves the satellite antenna at the epoch, T_t, measured in the satellite time frame. The signal state propagates with approximately the velocity of light, and reaches the receiver antenna at the epoch, t_r, measured in the receiver time frame. The same is valid for the code signals. Note that the phase states of the received signals are identical to the phase states of the transmitted signals. In other words, measuring the phase state at the receiver means measuring the signal's emission epoch at the satellite.

The process of *code phase observation* is illustrated in Fig. 7.33. The code sequence, generated in the receiver, is shifted stepwise against the code sequence received
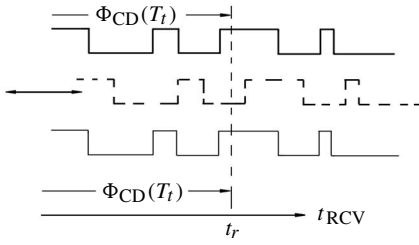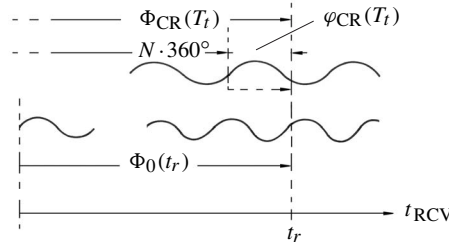
Figure 7.33. Code phase measurement

Figure 7.34. Carrier phase measurements

from the satellite, until maximum correlation is achieved [7.2.2]. At the moment of maximum correlation the code phase, $\Phi_{CD}(T_t)$, of the internal code sequence is measured in the receiver time frame, giving t_r. This code phase is identical to the code phase transmitted from the satellite, if we neglect propagation delays within the receiver. We hence obtain the epoch of transmission, T_t, of the code state in the satellite time frame. The difference between both clock readings yields the pseudorange, i.e.

$$PR = c(t_r - T_t). \tag{7.43}$$

Defining:

dt_s satellite clock error with respect to GPS system time,
dt_u clock synchronization error,
dt_a atmospheric propagation delay,
ε_R observation noise, and
R slant range

we obtain the more developed observation equation for *code measurements* (cf. (7.35)):

$$PR_{CD} = c(t_r - T_t) = R + cdt_u + cdt_a + cdt_s + \varepsilon_R. \tag{7.44}$$

In addition, propagation delays in the satellite or the receiver hardware may be included; they cannot be separated from clock errors, and hence are included in the respective clock model. The signs of the different terms come from convention and they differ in the literature.

Note the iterative character of equation (7.44) because the slant range, R, between receiver, B, and satellite, S, at the epochs of transmission and reception is given by

$$R^2 = (X_S(T_t) - X_B(t_r))^2 + (Y_S(T_t) - Y_B(t_r))^2 + (Z_S(T_t) - (Z_B(t_r))^2, \tag{7.45}$$

with

$$T_t = t_r - \frac{R}{c}.$$

The process of *carrier phase observation* is illustrated in Fig. 7.34. The observable is the difference between the transmitted and Doppler-shifted carrier phase, $\varphi_{CR}(T_t)$,

defined in the satellite time frame, and the phase of the reference signal, $\Phi_0(t_r)$, defined in the receiver time frame. The "observed" relative phase is:

$$\varphi_m(t_r) = \varphi_{CR}(T_t) - \Phi_0(t_r). \tag{7.46}$$

The carrier phase can be written, using the relation (7.42), as

$$\varphi_{CR}(T_t) = \varphi_m(t_r) + \Phi_0(t_r) = \varphi_m(t_r) + t_r f_0. \tag{7.47}$$

With the substitution (see Fig. 7.34),

$$\Phi_{CR}(T_t) = N \cdot 360° + \varphi_{CR}(T_t), \tag{7.48}$$

and with N as the integer ambiguity, we obtain one expression for the epoch of transmission, T_t, of the carrier phase signal, referred to the satellite time frame:

$$T_t = \frac{\Phi_{CR}(T_t)}{f_{CR}} = \frac{\varphi_{CR}(T_t) + N}{f_{CR}} = \overline{T}_t + \frac{N}{f_{CR}}. \tag{7.49}$$

The pseudorange from carrier phase measurements is then

$$PR_{CR} = c(t_r - \overline{T}_t). \tag{7.50}$$

With the ambiguity term,

$$c \cdot \frac{N}{f_{CR}} = N \cdot \lambda_{CR},$$

the observation equation for *carrier phase measurements*, corresponding to (7.44), becomes

$$PR_{CR} = R + cdt_u + cdt_a + cdt_s + c\left(\frac{N}{f_{CR}}\right) + \varepsilon_R. \tag{7.51}$$

Note that the clock parameters, dt_u, dt_s, the ambiguity term, N, and the hardware signal delays are linear dependent. Hence, ambiguity fixing is not a trivial problem (Wübbena et al., 2001b). Either the parameters are eliminated by forming differences, or the singularity is carefully treated in the parameter estimation process [7.3.2.2].

7.3.2 Parameter Estimation

7.3.2.1 Linear Combinations and Derived Observables

We can see that both observables, carrier phases and code phases on both frequencies, lead to pseudoranges. It may therefore be of advantage to use all observables, or linear combinations thereof, in the parameter estimation process. In principle, an unlimited number of possibilities exists, to combine the different observables, and to form derived observables, but only some combinations are meaningful in the context of positioning. We distinguish combinations
   − between observations at different stations,

  − between observations of different satellites,
  − between observations at different epochs,
  − between observations of the same type, and
  − between observations of different type.

One advantage of the use of derived observations is that errors that are present in the original observations are eliminated or reduced when differences are formed between observables. In some cases the ambiguities of derived observations are easier to solve than for those of the original observations. On the other hand, the noise level may be considerably increased on combination. The use of linear combinations in the parameter estimation process must hence be thoroughly studied. Common linear combinations are between stations and satellites. Following Fig. 7.35 we introduce

  2 receivers i, j,
  2 satellites p, q,
    epoch t_1: position 1 of the satellites p, q,
    epoch t_2: position 2 of the satellites p, q, and
  8 pseudorange measurements:

$$PR_{1i}^p, \ PR_{2i}^p, \ PR_{1i}^q, \ PR_{2i}^q; \ PR_{1j}^p, \ PR_{2j}^p, \ PR_{1j}^q, \ PR_{2j}^q \ .$$

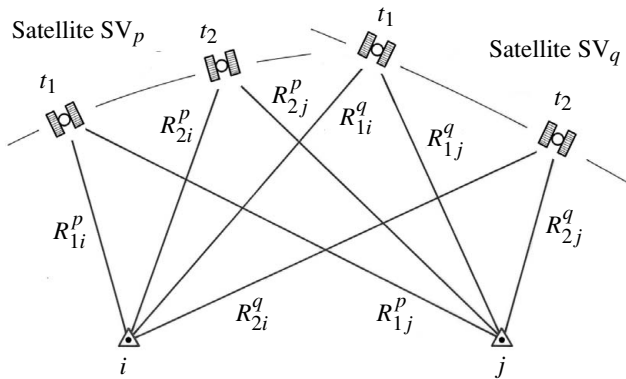These may be either code or carrier phase observations.

Figure 7.35. Differencing between receivers and satellites

Single differences can be formed between two receivers, between two satellites, or between two epochs. Following the notation of Wells (ed.) (1986) and Fig. 7.35 we introduce the differencing operators

for between-receiver single differences:

$$\Delta(\bullet) = (\bullet)_{\text{receiver } j} - (\bullet)_{\text{receiver } i}, \tag{7.52}$$

for between-satellite single differences:

$$\nabla(\bullet) = (\bullet)^{\text{satellite } q} - (\bullet)^{\text{satellite } p}, \tag{7.53}$$

for between-epoch single differences:

$$\delta(\bullet) = (\bullet)_{\text{epoch 2}} - (\bullet)_{\text{epoch 1}}. \tag{7.54}$$

Between-epoch single differences (7.54) correspond to the well-known *Doppler solution* [6.1].

In GPS geodesy single differences between two receivers are usually formed, i.e. pseudorange measurements at one station are subtracted from simultaneous pseudorange measurements to the same satellite at a second station. The observables are

$$((PR^p_{1i} - PR^p_{1j}), (PR^p_{2i} - PR^p_{2j}), (PR^q_{1i} - PR^q_{1j}), (PR^q_{2i} - PR^q_{2j})). \tag{7.55}$$

For code phases, equation (7.52) reads with (7.44) (the unchanging indices are omitted for simplicity) as

$$\Delta PR_{\text{CD}_{ij}} = \Delta R_{ij} + c(dt_{u_j} - dt_{u_i}) + c(dt_{a_j} - dt_{a_i}) + c(dt_s - dt_s) + \varepsilon_{\Delta_{\text{CD}}}, \tag{7.56}$$

or, in a simplified notation (e.g. Wells (ed.), 1986), as

$$\Delta PR = \Delta R + c\Delta dt_u + \Delta d_a + \varepsilon_{\text{CD}}. \tag{7.57}$$

The atmospheric delay, d_a, is often separated into the ionospheric and the tropospheric parts $d_{\text{ion}}, d_{\text{trop}}$, so that

$$\Delta PR = \Delta R + c\Delta dt_u + \Delta d_{\text{ion}} + \Delta d_{\text{trop}} + \varepsilon_{\text{CD}}. \tag{7.58}$$

For carrier phases we obtain, with (7.51):

$$\begin{aligned} \Delta PR_{\text{CR}_{ij}} &= \Delta R_{ij} + c(dt_{u_j} - dt_{u_i}) + c(dt_{a_j} - dt_{a_i}) \\ &\quad + c(dt_s - dt_s) + \lambda_{\text{CR}}(N_i - N_j) + \varepsilon_{\Delta\text{CR}}, \end{aligned} \tag{7.59}$$

or, in its abbreviated form:

$$\Delta PR_{\text{CR}} = \Delta R_{ij} + c\Delta dt_{u_{ij}} + c\Delta dt_{a_{ij}} + \lambda_{\text{CR}}\Delta N_{ij} + \varepsilon_{\Delta\text{CR}}. \tag{7.60}$$

In a simplified notation (e.g. again Wells (ed.), 1986) the between-receiver single differences read:

$$\Delta\Phi = \Delta R + c\Delta dt_u - \Delta d_{\text{ion}} + \Delta d_{\text{trop}} + \lambda\Delta N + \varepsilon_{\Phi}. \tag{7.61}$$

It becomes evident that the satellite clock error, dt_s, has disappeared, and that the errors in the propagation delay, dt_a, only affect the range measurements with the remaining differential effect. For stations close together, dt_{a_j} and dt_{a_i} may be regarded as equal, and they consequently vanish from equations (7.56) and (7.59). The same is true for the effect of orbital errors (cf. [7.4.3]).

If single differences between satellites are formed, i.e. the observations of two satellites simultaneously recorded at a single station are differenced, the receiver clock term, dt_u, in (7.58) and (7.60) cancels.

With single differences between two epochs for the same satellite the ambiguity term, N, from equation (7.51) cancels because the initial phase ambiguity does not change with time (as long as no cycle slips occur (cf. [7.3.3.1]).

Double differences (DD) are usually formed between receivers and satellites. They are constructed by taking two between-receiver single differences (7.57), (7.61) and differencing these between two satellites, SV_p and SV_q. The derived observables are (Fig. 7.35):

$$((PR_{1i}^p - PR_{1j}^p) - (PR_{1i}^q - PR_{1j}^q)), ((PR_{2i}^p - PR_{2j}^p) - (PR_{2i}^q - PR_{2j}^q)). \quad (7.62)$$

The observation equation for carrier phases develops, from (7.60), as

$$\begin{aligned}
\nabla\Delta PR_{CR} &= (\Delta R_{ij}^p - \Delta R_{ij}^q) + c(\Delta t_{u_{ij}} - \Delta t_{u_{ij}}) + c(\Delta t_{a_{ij}^p} - \Delta t_{a_{ij}^q}) \\
&\quad + \lambda_{CR}(\Delta N_{ij}^p - \Delta N_{ij}^q) + \varepsilon_{\nabla\Delta},
\end{aligned} \quad (7.63)$$

or in a simplified notation, as

$$\nabla\Delta\Phi = \nabla\Delta R - \nabla\Delta d_{ion} + \nabla\Delta d_{trop} + \lambda\nabla\Delta N + \varepsilon_\Phi. \quad (7.64)$$

The equivalent equation for code measurements is:

$$\nabla\Delta PR = \nabla\Delta R + \nabla\Delta d_{ion} + \nabla\Delta d_{trop} + \varepsilon_{CD}. \quad (7.65)$$

Note that the receiver clock term, dt_u, vanishes. The double difference observables are free from satellite and receiver clock errors and include only reduced propagation and orbit errors. The double difference observable is the basic observable in many adjustment models for GPS observations (cf. [7.3.4]). Double differences are also the basic observables in many techniques used for the resolution of ambiguities [7.3.2.3].

Triple differences between receivers, satellites, and time, are constructed by taking two epochs, t_1 and t_2. The derived observables are (Fig. 7.35):

$$((PR_{1i}^p - PR_{1j}^p) - (PR_{1i}^q - PR_{1j}^q)) - ((PR_{2i}^p - PR_{2j}^p) - (PR_{2i}^q - PR_{2j}^q)). \quad (7.66)$$

As discussed earlier, the initial cycle ambiguity, N, cancels from the observation equation. What remains is a linear combination of all pseudoranges and the residual propagation biases as well as unmodeled orbital biases. The observation equation for triple differences is identical for code and carrier phases, except for the sign of the ionospheric delay. It is, however, usually not established for code observables. In the simplified notation we obtain directly that

$$\delta\nabla\Delta\Phi = \delta\nabla\Delta R - \delta\nabla\Delta d_{ion} + \delta\nabla\Delta d_{trop} + \varepsilon_{res}. \quad (7.67)$$

The last three terms contain the residual propagation biases and unmodeled effects. Triple differences are often used to provide approximate (Doppler) solutions. They are also very useful to aid the removal of cycle slips [7.3.3.1] in an automatic editing process.

For a more complete presentation and for complete expressions see e.g. Wells (ed.) (1986); Wübbena (1991); Leick (1995); Hofmann-Wellenhof et al. (2001).

Linear combinations between observations of the same type can be formed between carrier phases and between code phases. The main uses of such linear combinations are for the elimination of the ionospheric delay [2.3.3.1], [7.4.4.1] and for the resolution of carrier phase ambiguities.

An arbitrary linear combination of the carrier phases on L1 and L2 is formed with integer coefficients, n, m (Wübbena, 1989):

$$\Phi_{n,m}(t) = n\Phi_1(t) + m\Phi_2(t). \tag{7.68}$$

With (2.80) the linear combination fulfils (for a given clock i) the equation

$$t^i(t) = \frac{\Phi_{n,m}^i(t)}{f_{n,m}}. \tag{7.69}$$

Here,

$$f_{n,m} = nf_1 + mf_2, \tag{7.70}$$

is the frequency of the derived signal. Scaling with the velocity of light yields the related wavelength:

$$\lambda_{n,m} = \frac{c}{f_{n,m}} = \frac{c}{nf_1 + mf_2}. \tag{7.71}$$

The ambiguity of the linear combination is then

$$N_{n,m} = nN_1 + mN_2, \tag{7.72}$$

i.e. $N_{n,m}$ is an integer if n and m are integers. Based on these formulas, the first order ionospheric effect and the resulting influence on the combined phase can be computed as (Wübbena, 1989):

$$\delta\Phi_{n,m,I} = -\frac{C_I}{f_1 f_2}(nf_2 + mf_1), \tag{7.73}$$

where C_i is a function of the total electron content (cf. [2.3.3.1], [7.4.4.1]).

The corresponding ionospheric influence on signal propagation time can be characterized with an amplification factor, V_I, as

$$\delta T_{n,m,I} = \frac{\delta\Phi_{n,m_I}}{f_{n,m}} = -\frac{C_I}{f_1 f_2}\frac{nf_2 + mf_1}{nf_1 + mf_2} = -\frac{C_I}{f_1 f_2}V_I. \tag{7.74}$$

The standard deviation of the original phase observation, σ_Φ, can be propagated into the standard deviation of the combined phase observation, with

$$\sigma_{\varphi_{n,m}} = \sqrt{n^2 + m^2}\sigma_\varphi, \tag{7.75}$$

or scaled into a range:

$$\sigma_{n,m} = \lambda_{n,m}\sigma_{\varphi_{n,m}}. \tag{7.76}$$

Among the unlimited number of linear combinations, only those that fulfill some important criteria for the combined signals are of interest:

– integer coefficients to produce *integer ambiguities*,
– reasonably *large wavelength* to help ambiguity fixing,
– *low ionospheric* influence, and
– limited observation noise.

Table 7.7. Selected linear combinations of carrier phases

Signal	n	m	λ	$\lambda_{1/2}$	V_I	σ
			cm	cm		mm
L_1	1	0	19.0	19.0	0.779	3.0
L_2	0	1	24.4	12.2	1.283	3.9
L_Δ	1	-1	86.2	43.1	-1.000	19.4
L_Σ	1	1	10.7	5.4	1.000	2.1
L_{-12}	-1	2	34.1	34.1	2.168	12.1
L_{32}	3	-2	13.2	13.2	0.234	7.6
L_{43}	4	-3	11.4	5.7	0.070	9.1
L_{97}	9	-7	5.4	2.7	0.004	9.7
L_{54}	5	-4	10.1	10.1	-0.055	10.3
L_{65}	6	-5	9.0	4.5	-0.154	11.2
L_0	-	-	≈ 5.4	≈ 2.7	0.000	10.0
L_I	-	-	≈ 10.7	≈ 5.4	2.000	20.0

Table 7.7 (Wübbena, 1989; Wanninger, 1994) summarizes the leading properties of some selected linear combinations. The column, $\lambda_{1/2}$, contains the effective wavelength for a receiver with squaring technique on L2. The observation noise of the original observation is taken from the specification in the "Interface Control Document" as 0.1 rad, corresponding to 3 mm (ICD, 1993). Well known linear combinations are the *wide lane*:

$$L_\Delta = L_1 - L_2; \quad \lambda_\Delta = 86.2 \text{ cm}, \tag{7.77}$$

and the *narrow lane*:

$$L_\Sigma = L_1 + L_2; \quad \lambda_\Sigma = 10.7 \text{ cm}. \tag{7.78}$$

The advantage of the wide lane observable, compared with the original observation, is that the ambiguity has to be resolved for a signal with a wavelength four times larger; the related observation noise is, however, six times greater.

The narrow lane has the lowest noise level of all linear combinations and hence yields the best results. Its ambiguity is, however, difficult to resolve (cf. [7.3.2.3]). The narrow lane is mainly used over short interstation distances.

The magnitude of the ionospheric effect in the wide lane and in the narrow lane is equal, but it has an opposite sign. Hence the mean of the wide and the narrow lane

yields the *ionospheric free signal*:

$$L_0 = \frac{L_\Delta + L_\Sigma}{2}. \qquad (7.79)$$

L_0 is not related to integer ambiguities (cf. [7.4.4.1]) and is therefore not a suitable signal for very precise solutions. The associated observation noise is rather high.

Finally, the *ionospheric signal*,

$$L_I = L_\Sigma - L_\Delta, \qquad (7.80)$$

formed from the difference between wide and narrow lane, is of interest because it contains the complete ionospheric effect. The signal allows a detailed analysis of the ionospheric behavior and is helpful in ambiguity resolution strategies (cf. [7.3.2.3]). The ambiguities of the wide and narrow lanes are:

$$N_\Delta = N_1 - N_2; \quad N_\Sigma = N_1 + N_2 \qquad (7.81)$$

i.e. they are not independent. When N_Δ is even, N_Σ has to be even, and when N_Δ is odd, N_Σ has to be odd:

$$N_\Delta \bmod 2 = N_\Sigma \bmod 2. \qquad (7.82)$$

This *even-odd condition* implies that when the ambiguity is resolved for one of the two combinations the effective wavelength of the other combined signals is increased by a factor of two. For example, when the ambiguity of the wide-lane, N_Δ, is resolved then the effective wavelength of the narrow lane is 21.4 cm and N_Σ can be resolved much more easily. For a table of effective wavelength factors see Table 7.10. The condition (7.82) is used in the *extra wide laning* technique [7.3.2.3], and generates, under the assumption of vanishing ionospheric differences, an effective wavelength of

$$2\lambda_\Delta = 1.72 \text{ m}.$$

Some linear combinations of Table 7.7 are very close to the ionospheric free signal, but with integer ambiguities, e.g. L_{54}, L_{43} and L_{97}. The related wavelengths of L_{54} and L_{43} are larger than λ_0, hence the ambiguity may be estimated more easily. This is in particular true for L_{54}, because the effective wavelength remains unchanged for squaring receivers.

Linear combinations can also be formed for code phases (Wübbena, 1988). Table 7.8 gives an overview. Only P-code observations can be used because the C/A-code is not available on L2.

Because of the dispersive ionospheric effect [2.3.3] the propagation time of signals and signal combinations is different. If signals are received at exactly the same epoch the related transmission epochs are different. Fig. 7.36 illustrates the situation for selected code and carrier signals with respect to the satellite time frame. Taking the first order ionospheric effect into account, it can be shown (Wübbena, 1988, 1991) that the apparent transmission epochs are identical for L_Δ and C_Σ, as well as for L_Σ

Table 7.8. Code phase linear combinations

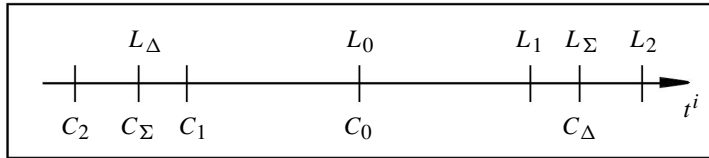| Signal	n	m	$-V_I$	$\bar{\sigma}$ [m]
C_1	1	0	-.779	0.47
C_2	0	1	-1.283	0.47
C_Δ	1	-1	1.000	2.68
C_Σ	1	1	-1.000	0.33

Figure 7.36. Apparent transmission epochs of linear signal combinations for an identical time of arrival

and C_Δ. This opens the possibility to combine code and carrier observations. The unambiguous code signals can be used in the estimation of ambiguities in the carrier signals (cf. [7.3.2.3]).

As part of the GPS modernization program [7.1.7] a third civil frequency, L5, located at 1176.45 MHz, will be available from about 2005. This third frequency can be used to form additional linear combinations, for instance:

$L_1 - L_5$, with $\lambda = 0.75$ m, and

$L_2 - L_5$, with $\lambda = 5.86$ m

These signals will certainly contribute to GPS data processing, although the noise level and ionospheric influences are rather high. Favorable applications are expected for processing short baselines (Hatch et al., 2000).

7.3.2.2 Concepts of Parametrization

As has been stated above, both primary observables, the code phases and the carrier phases, yield pseudoranges between the receiver antenna and a certain number of satellites. Instead of pseudorange the expression *biased range* (range with systematic error effects) is often used. The fundamental difference between both observables is the ambiguity term in equations (7.44) and (7.51). The handling of the ambiguity and bias problem leads to different evaluation concepts. Two main approaches may be distinguished (cf. [4.1]):

  (a) parameter estimation, and

  (b) parameter elimination.

(a) All systematic influences (biases) that have a stable and well-described structure are *estimated* together with the station parameters as so-called *nuisance* or *bias parameters*. Bias parameters may be, for instance, corrections to the satellite orbit, clock parameters, ambiguity terms, and tropospheric scale factors. These biases can be either directly measured by additional observations (e.g. the ionospheric delay) or they are included in an extended adjustment model (e.g. the tropospheric delay). The *undifferenced phase measurements* (7.39) are used as the basic observables.

(b) Most of the biases are *eliminated* by taking the difference between observables. It is assumed that the disturbing terms are linearly dependent with one another in the various data sets. Up to a certain degree this is correct, e.g. for clock biases, orbital biases and ambiguities. *Single, double* and *triple differences* of the carrier phase measurements are used as derived observables [7.3.2.1]. The concept is primarily applied for *baselines* between two stations.

Both procedures have advantages and disadvantages. The main advantage of the *parameter estimation* approach (a) is its flexibility and independence of the requirement for simultaneous observations at all participating stations. Coordinates of a point field (network) are determined, not derived quantities such as baselines. The behavior of certain parameters, like the comportment of clocks, can be controlled. Highly stable satellite oscillators can be used to improve the stability of the adjustment.

Additional biases like antenna phase center variations, multipath effects or hardware propagation delays can be modeled in a rigorous way and can be directly integrated into the observation equation. Modern approaches like *Precise Point Positioning* (PPP) [7.3.4] are only possible with undifferenced data. In total, it can be stated that carrier phase observations are physically much better represented by undifferenced observables than by double differences. On the other hand, the undifferenced approach requires that all bias and nuisance parameters have to be included explicitly in the observation equation.

The *baseline concept* (b) was introduced from experiences with Very Long Baseline Interferometry (VLBI) (Counselman, Steinbrecher, 1982). Its main advantage is that many common error effects are eliminated from the observations by differencing, thus simplifying the parameter estimation. This is in particular true for the satellite and receiver clock errors, but also to some extent for orbit and signal propagation errors. For larger station separations, however, the elimination process no longer works out in a rigorous manner; it becomes more difficult or even impossible to fix the ambiguities to whole numbers (integers). Also, the number of independent observations is considerably reduced by the differencing, and information is lost.

Through the differencing process all absolute biases are eliminated and only the differences between the biases remain in the data. Differenced observables are no longer single station related but vector related. The error modeling of these differences, however, is much more difficult than is the error modeling for the undifferenced original bias effects (see e.g. Wübbena et al., 2001b). The double difference approach introduces mathematical correlations into the resultant observables. These correlations have to be modeled in the variance–covariance matrix.

The principle of positioning with GPS in the *parameter estimation* approach follows equations (7.44) and (7.51). It is illustrated in Fig. 7.37. Here T_{t_i} is the epoch of signal transmission in the time system of the individual satellite, S_i. The relation to the GPS system time (satellite time frame) is written as

$$T_t = T_{t_i} - dt_{s_i}. \qquad (7.83)$$

Observations are:

  t_r the epoch of signal reception in the receiver time frame, and

  T_{t_i} the epoch of signal transmission in the individual time frame of satellite, S_i.

Figure 7.37. Main solution parameters in positioning with pseudoranges

Known quantities are the coordinates, $X_i(T_t)$ of the satellite, S_i, at the epoch, T_t, in the CTS coordinate system. In their fundamental simplified formulation the observation equations (7.44) and (7.51) are solved for four unknowns. These are:

  X_u three coordinates (X_u, Y_u, Z_u) of the user antenna in the CTS coordinate system, and

  dt_u clock synchronization error between the user's clock and GPS system time.

The parameter vector of the linearized observation equation is then

$$X = (X_u, Y_u, Z_u, dt_u). \qquad (7.84)$$

For a more developed model of parameter estimation, additional parameters can be introduced, for example:

  3 clock biases per station,
  3 clock biases per satellite,
  6 orbital biases per satellite,
  1 parameter for solar radiation pressure,
  1 tropospheric parameter per station and satellite, and
  1 ambiguity parameter per station and satellite.

The parameter for tropospheric propagation delay, dt_{a_i}, is already contained in Fig. 7.37 and in the observation equations (7.44) and (7.51). The clock behavior can be described by a polynomial model with terms for clock bias, drift and ageing (cf. (7.4), [2.2.5]):

$$dt_u = a_{0u} + a_{1u}(t - t_0) + a_{2u}(t - t_0)^2,$$
$$dt_s = a_{0s} + a_{1s}(t - t_0) + a_{2s}(t - t_0)^2. \qquad (7.85)$$

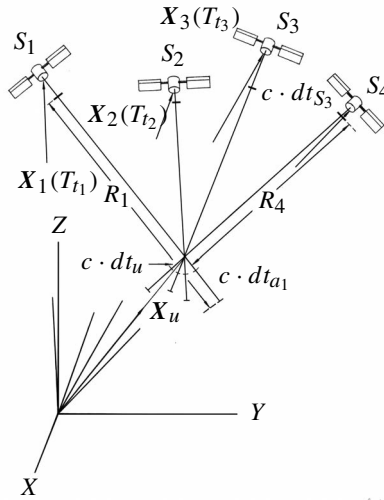The observation equation must be extended accordingly. More parameters can be included if necessary, for example antenna phase center variations or receiver hardware delays.

The singularity in equation (7.51), namely the linear dependency between clock parameters, signal propagation delay, and ambiguity term, has to be treated carefully by fixing certain ambiguity parameters. It is important to choose the correct number of parameters so that the singularity can just be removed, and the ambiguities remain integers (Wübbena et al., 2001b).

The unknown parameters are combined into the vector of unknowns, X, and the observations into the vector of observations, PR. The linearized observation equation is then written as

$$l = PR - PR(X_0). \tag{7.86}$$

X_0 is the approximation vector of the unknown parameters and l the vector of reduced observations. With

$$x = X - X_0, \tag{7.87}$$

it follows that

$$l = Ax. \tag{7.88}$$

The design matrix A contains the partial derivatives of the observations with respect to the unknowns:

$$A = (\partial PR/\partial X). \tag{7.89}$$

The solution vector of the system is

$$x = A^{-1}l. \tag{7.90}$$

With ε_R as the measurement noise of the pseudoranges (cf. (7.44)), the error ε_x of the adjusted parameters is found to be

$$\varepsilon_x = A^{-1}\varepsilon_R. \tag{7.91}$$

For more detailed information on adjustment techniques see e.g. Vaníček, Krakiwsky (1986, chap. 12), Leick (1995, chap. 4), Strang, Borre (1997, chap. 1) or Niemeier (2002).

In the *parameter elimination* process (b) various combinations of differences are formed, as described in [7.3.2.1]. The formulation of differences implies some consequences that have to be considered for data processing. The most important aspects are briefly discussed.

Algebraic correlation, introduced into the differences, has to be taken into account in a rigorous adjustment. The covariance matrix of observations has to be updated with the increasing number of receivers and receiver types. The problem is discussed e.g. by Beutler et al. (1987); Goad, Müller (1988); Goad (1998) and in most GPS textbooks like Leick (1995); Hofmann-Wellenhof et al. (2001).

A *pre-selection of differences* is necessary if several stations and satellites are involved in the same session, because only one part of the possible differences is

independent. One possible concept is the definition of a *reference satellite* and a *reference station* (e.g. Goad, 1985, 1998; Hofmann-Wellenhof et al., 2001) that have to be introduced into all differences. Difficulties arise with data gaps for the reference satellite or the reference station.

Differencing of observations eliminates the *absolute information* that is contained in the undifferenced observations. A reliable absolute datum has to be introduced into the solution in order to avoid mismodeling [7.6.1.3].

Identical observation epochs are required for the use of differenced observations. For receivers observing at different epochs within the same session, the observations have to be reduced to identical epochs with appropriate interpolation models, e.g. stochastic clock modeling (e.g. Wübbena, 1988).

Because of the simple basic model and the good results in processing short baselines, most commercial software packages use the baseline approach (b) (parameter elimination) with double differences as primary observables. The parameter estimation technique (a) with undifferenced observables is preferred in scientific software packages such as GEONAP and GIPSY-OASIS II. The well-known scientific package BERNESE, however, is also based on the double difference concept [7.3.4].

7.3.2.3 Resolution of Ambiguities

Carrier phase measurements are affected by the ambiguity term N (7.51), that is, by an unknown number of complete wavelengths between the satellite and the receiver antenna. This *initial ambiguity* has to be determined with appropriate techniques to exploit the full accuracy potential of the GPS carrier phase measurements. Ambiguity determination is one of the most demanding problems in the geodetic technique of evaluating GPS observations. On the other hand, it is the integer nature of the phase ambiguities that guarantees the high accuracy of relative positioning with GPS, in particular when the observation time is short.

The best and simplest possibility for determining the ambiguity would be the use of additional frequencies or signals, as is the case for terrestrial electronic distance measurements (e.g. Kahmen, Feig, 1988). Unfortunately, for the time being, GPS does not provide more than two frequencies, hence other strategies were developed to solve the ambiguity problem. The main approaches are (see e.g. Han, Rizos, 1997):
 (a) the geometric method (coordinate domain search),
 (b) code and carrier phase combinations (observation domain search),
 (c) ambiguity search methods (ambiguity domain search), and
 (d) combined methods.

Currently the approach (c) is considered to be the most effective and powerful search method, in particular for fast solutions, and is discussed widely in the literature.

(a) *Coordinate domain search*
The *geometric method* makes use of the time-dependent variation in the geometric relation between receiver and satellites. In general, continuous phase measurements are used, and the ambiguities are estimated as real-number parameters. A simple

explanation of the procedure is as follows. Once the satellite signals are identified by the receiver, the whole number of incoming cycles is measured and counted. The unknown initial ambiguity N is maintained throughout the observation process and can be represented by a single parameter (bias). Continuous tracking of the carrier phases results in the determination of ambiguity-free range differences. These are used in a Doppler solution [6.1], to determine the coordinates of the user antenna. Ambiguity-free pseudoranges R_i between the user antenna and the satellites can be derived from the Doppler solution, and compared with the ambiguous range observations. The ambiguities are directly derived from this comparison.

The technique works if the change of geometry is sufficiently large, i.e. with a rather long observation time. If the receiver loses lock on one or more satellites and the remaining number of simultaneously-observed satellites is less than four, a new initial ambiguity has to be introduced. This is the *cycle slip problem* [7.3.3.1]. With more than four satellites, the ambiguities are not independent. A geometrical interpretation of the technique is given with Fig. 7.38.

The estimated ambiguities are real numbers (*ambiguity float* or *ambiguity free solution*). They can be fixed to integer numbers (*ambiguity fixed solution*) if the estimated values are very close to an integer number, or in other words if the relative position error in the direction of the satellite is smaller than half a cycle length. This requirement leads to the rather long observation time in the pure geometric method. A reduction of

Figure 7.38. Geometric method of ambiguity resolution

the observation time is possible with additional satellites, better geometry, or when signals with a larger wavelength are used (e.g. the *wide lane* L_Δ (7.77)).
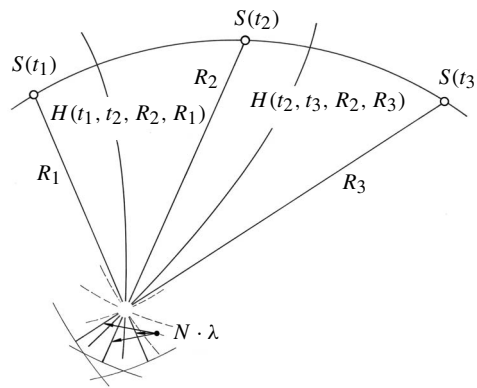
One advantage of the geometric method is the clear and simple modeling. For short interstation distances the method usually works out properly and provides reliable results. Problems arise when unmodeled systematic effects remain in the observations, e.g. due to troposphere, ionosphere, and the satellite orbit. The influences of these effects increase with station separations, and make reliable ambiguity fixing difficult or impossible. A wrong determination of the integer ambiguity introduces systematic errors into the coordinates.

The geometric method of ambiguity fixing is one of the earliest methods and has been widely discussed in literature (e.g. Gurtner et al. (1985); Frei, Beutler (1989); Wübbena (1991)). One specific disadvantage of the method is the long observation time necessary to obtain real-value results good enough to resolve the ambiguities. To overcome this problem, at least for short baselines, special observational techniques

have been developed such as "antenna swapping" or "re-occupation" (see [7.3.5.3]).

One particular development, the *ambiguity function* method, was described for the first time by Remondi (1984). The basic idea is as follows: Observables are the *single differences* [7.3.2.1] between two stations where the coordinates of one station are taken as known. Unknown parameters are only the coordinates of the second station and the difference in the receiver clock errors. A search algorithm is defined that varies the baseline vector until the related computed single differences correspond best with the observed single differences. Further developments of this method are described by Mader (1990); Remondi (1990); Han, Rizos (1997) and Hofmann-Wellenhof et al. (2001, p. 229ff). The technique is, from today's understanding, of rather poor computational efficiency and consequently it is of only historical interest (Kim, Langley, 2000). Advantages and disadvantages of the geometric methods are summarized as follows:

Advantages
- basically simple and clear modeling,
- works with few satellites,
- usable for short, long and very long distances, and
- the ambiguity float solution rapidly provides approximate results.

Disadvantages
- long observation time necessary for sufficient geometric rigor,
- influenced by unmodeled effects like ionosphere, orbits, etc.,
- no a priori use of the integer nature of ambiguities, and
- sensitive to unrecovered cycle slips.

(b) *Observation domain search*

In the second approach to ambiguity solution the *combination of code* and *carrier phase* observations is applied. The non-ambiguous code phase measurements are used as an additional wavelength to resolve the carrier phase ambiguity:

$$PR_{CR} - PR_{CD} = \lambda N + dT_A + d\varepsilon. \tag{7.92}$$

This method is independent of the geometry and it is sometimes referred to as the *geometry-free* technique (Hatch et al., 2000). The difference between both observables contains, however, the residual errors dT_A (see below). The basic idea is to make code measurements until the noise level of the code solution is less than half the wavelength of the carrier wave. Because of the much larger (mainly multipath induced) noise in the code measurements, determination of the cycle ambiguity requires that the code observations be smoothed over multiple epochs.

The idea behind this approach was discussed early by Bossler et al. (1980) and Hatch (1982b), but it was not considered to be operational. First results were presented independently by Melbourne et al. (1985) and Wübbena (1985). Today the technique is widely applied, in particular for kinematic applications over short distances (Hatch et al., 2000; Hofmann-Wellenhof et al., 2001). It also promises good results with the forthcoming availability of three frequencies (Vollath et al., 1999).

It is obvious that the method requires a receiver with a low noise level for code-measurements. Even then it is difficult to resolve the ambiguities of the original signals L_1 and L_2, because of their short wavelengths. Instead, the *wide lane* L_Δ (7.77) with 86.2 cm is used. The situation will improve for receivers with a code noise as low as a few centimeters [7.2.4.2].

For proper modeling the different propagation properties of codes and carriers in the ionosphere have to be considered. The signals L_Δ and C_Σ are combined because they have identical propagation times (cf. [7.3.2.1], Fig. 7.36, Wübbena (1988)).

The term dT_A in (7.92) contains the different propagation delays of both signal types due to the satellite and receiver hardware, and in particular due to *multipath effects* [7.4.4.3]. Note that the method is completely independent of the observation geometry, of the satellite and receiver clocks, and of the atmospheric delays. The method hence also works for longer baselines and in kinematic mode.

Once the wide lane ambiguity is correctly resolved, the ambiguities of further signals can be determined with the geometrical method if the ionosphere can be properly modeled; these are for example the ambiguity of the ionospheric free signal or the ambiguity of the narrow lane. The narrow lane is of particular interest to precise applications because of its very low noise level (cf. Table 7.7). It is, however, very difficult to solve for larger distances, because of its very small wavelength, and it is mainly applied in short range applications.

For short interstation distances, where the ionospheric effect can be considered as equal on both stations, the *extra wide laning* technique can be applied (Wübbena, 1988, 1989). Because of the even-odd condition (7.82), the effective wavelength of the wide lane increases to 1.72 m.

To give an example, for short distances the ionospheric signal in the single or double difference must be zero, that is

$$L_I = L_\Sigma - L_\Delta \overset{!}{=} 0. \tag{7.93}$$

This condition is only fulfilled if the wide lane ambiguity has been resolved correctly. Let the wide lane ambiguity be estimated as 0.6 cycles, and fixed to 1 cycle. Following the condition (7.93) the narrow lane ambiguity will then be estimated to 8 cycles because

$$\lambda_\Delta \approx 8\lambda_\Sigma. \tag{7.94}$$

This disagrees with the even-odd condition, consequently the wide lane ambiguity was erroneous and will be fixed to 0.

The application of the extra-wide laning technique leads to dramatic savings in observation time, and it is also applicable for ambiguity resolutions "*on the way*" (or "*on the fly*") in kinematic surveying ([7.3.5.4], [7.5.2]). Advantages and disadvantages of the code/carrier methods are summarized as follows:

Advantages:
   − independence of geometry,
   − kinematic application, and

  − long and very long baselines possible.

Disadvantages:
  − dual frequency P-code receiver necessary,
  − sensitive to multipath, and
  − only wide lane ambiguities are resolved.

(c) *Ambiguity domain search*

Ambiguity search methods have been developed with the objective of cutting down the necessary observation time for an individual observation station. The more satellites that are available, the better this method works. The basic idea is to search for the optimum ambiguity combination of L_1, L_2 or derived signals. The search algorithm usually starts with an initial ambiguity float solution and then restricts the solution vector to discrete integer values applying some optimization techniques. The possible combinations within a pre-defined "ambiguity-space" are examined. The procedure is illustrated in Fig. 7.39 for the case of two and three satellites. Each additional satellite restricts the number of possible solutions.

○   possible solutions (selection)
◎   final solution
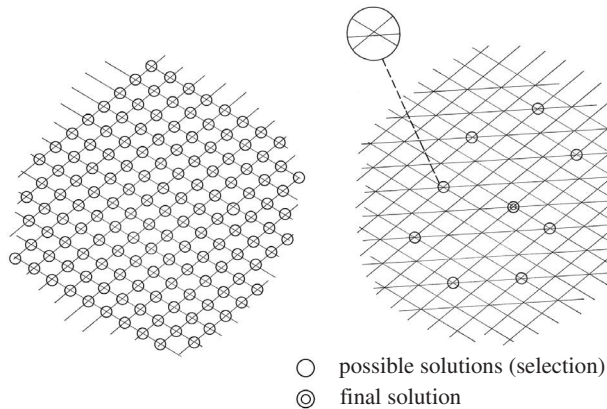
Figure 7.39. Possible solutions for the ambiguities are selected; situation for two satellites (left) and three satellites (right)

   The basic problem with this is that the number of necessary mathematical operations increases rapidly beyond all limits. If n is the number of cycles within the search interval, and m the number of ambiguities to be determined, then the number of necessary operations is as given in Table 7.9. It becomes evident that the problem is not solvable by the examination of all possible combinations, but that appropriate selection strategies have to be applied.

   Various proposals for selection strategies exist in the literature, and have been introduced in software packages (see e.g. Hatch, 1991; Leick, 1995; Teunissen, 1998; Hofmann-Wellenhof et al., 2001). Early examples are the treatment as a *neural network*

Table 7.9. Number of necessary operations for fixing ambiguities in all possible combinations (Wübbena, 1991)

$n \backslash m$	1	3	10	20
4	1 620	48 020	$3.9 \cdot 10^6$	$5.6 \cdot 10^7$
8	$4.7 \cdot 10^5$	$4.1 \cdot 10^8$	$2.7 \cdot 10^{12}$	$5.7 \cdot 10^{14}$
20	$1.5 \cdot 10^{12}$	$3.3 \cdot 10^{19}$	$1.1 \cdot 10^{29}$	$7.5 \cdot 10^{34}$

(Landau, 1990), or the *Fast Ambiguity Resolution Approach* (FARA) (Frei, Beutler, 1990). A very powerful modern technique is the *LAMBDA method*, developed at the Delft University of Technology (Teunissen et al., 1995; Teunissen, 1998; Joosten, Tiberius, 2000). LAMBDA stands for "Least Squares Ambiguity Decorrelation Adjustment". The basic idea is to transform the original real-valued double difference ambiguities, which are highly correlated, into decorrelated real-valued ambiguities. As such the number of solution candidates is considerably reduced. By this procedure the original highly elongated search space is transformed into a sphere-like search space with the same volume, which allows a much more efficient identification of the integer ambiguities (see Fig. 7.40).
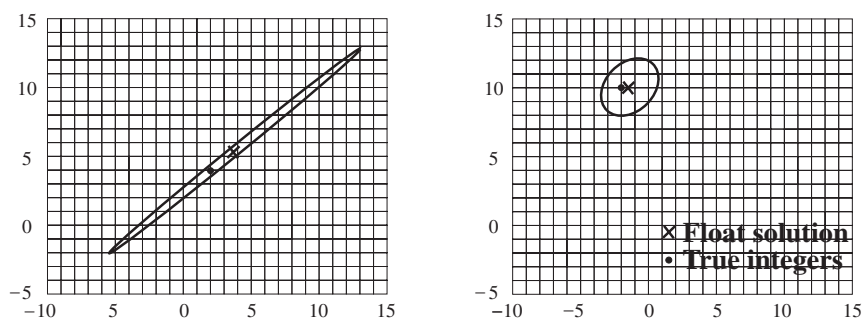
Figure 7.40. Lambda method: ambiguity search space before (left) and after (right) decorrelation (adapted from Joosten, Tiberius 2000)

Other representative techniques are, for example, FAST (Fast Ambiguity Search Filter, Chen, Lachapelle, 1995) and OMEGA (Optimal Method for Estimating Ambiguities, Kim, Langley, 1999). Further algorithms are still under development and will improve the methods of fast ambiguity resolution, in particular for surveying applications over short distances (cf. [7.3.5.2]) but also for long-range real-time kinematic positioning. For a short review of status and trends see e.g. Kim, Langley (2000). Advantages and disadvantages of the ambiguity search methods are summarized as follows:

Advantages:
  - allows fast ambiguity resolution (e.g. in rapid static applications),
  - true kinematic applications possible, and
  - uses the integer nature of ambiguities.

Disadvantages:
  - sensitive to systematic errors, and
  - requires the observation of as many satellites as possible.

(d) *Combined methods*

These methods include a combination of all the possibilities mentioned above, and should yield the best results. Other techniques may be added. The basic idea is that each fixed ambiguity improves and stabilizes the solution in the next iteration step. This is not only true for the resolution of individual ambiguities but also for fixing of arbitrary integer linear combinations of different ambiguities.

We have seen that the effective wavelength of the wide or narrow lane increases by a factor of two if the ambiguity of one linear combination has been already resolved (7.93), (7.94). Table 7.10 shows the effective wavelength factors for some frequently used linear combinations. If the ambiguity of the "fixed" signal is known, the effective wavelength of the "free" (ambiguity not yet fixed) signal increases by the given factor.

Table 7.10. Effective wavelength factors

Fixed Signal$\Rightarrow$	L_1	L_2	L_Δ	L_Σ	L_{54}	L_{43}	L_I	L_0
$\Downarrow$ Free Signal								
L_1	–	1	1	1	4	3	9	7
L_2	1	–	1	1	5	4	7	9
L_Δ	1	1	–	2	1	1	2	2
L_Σ	1	1	2	–	9	7	16	16
L_{54}	4	5	1	9	–	1	17	1
L_{43}	3	4	1	7	1	–	15	1
L_I	9	7	2	16	17	15	–	32
L_0	7	9	2	16	1	1	32	–

Options to use these signals are included in some scientific software packages. The adjustment of a complete network can be improved if the ambiguities for some of the baselines are known a priori. Ambiguities of shorter baselines are in most cases easier to resolve than ambiguities in long baselines. Techniques have been developed for including very short baselines within large scale networks, and then starting with the ambiguity resolution for the short baselines as a first step (Blewitt et al., 1988).

Considerable benefits will be derived from the addition of a third frequency, L5. The new signals $L_1 - L_5$ and $L_2 - L_5$ can be used to resolve ambiguities in a step-wise approach, starting with the longest wavelength. This procedure is also referred to as

cascade ambiguity resolution. For the promises and limitations see e.g. Hatch et al. (2000).

Particular methods have been developed to rapidly determine the initial phase ambiguities for *kinematic surveys* [7.3.5]. One powerful early procedure is the *antenna swapping* technique, i.e. the exchange of two antennas over a very short baseline (several meters) before the start of the kinematic survey (Remondi, 1985). This procedure is described in more detail in section [7.3.5.2]. Today, with the availability of a large number of visible satellites (sometimes eight and more) the ambiguity search methods are in particular a powerful means for the rapid resolution of ambiguities, often based on a single epoch observation.

The resolution of ambiguities is a key factor for precise GPS surveying. In many cases, in particular if the interstation distances are small, and if the data quality is good, the ambiguity resolution works out satisfactorily with the routine options in the software supplied by the manufacturer. In all those cases where

- the interstation distances are large (> 10 km) and highest accuracy (≤ 1 cm) is required,
- the data quality is poor (e.g. multipath, cycle slips),
- only a few satellites are visible,
- the ionosphere is disturbed, and/or
- the observation time is short,

problems may arise in solving for ambiguities. In such cases, a careful and interactive data processing operation with multipurpose GPS adjustment software may be necessary. The proper use of the different possibilities, as they have been discussed in this chapter, usually requires trained and experienced personnel.

For precise differential GPS (PDGPS) in reference station networks [7.5.3] a modeling of the error state in real-time helps to reduce the *Time To Fix Ambiguities* (TTFA) and to improve the ambiguity success rate.

One key question is whether the ambiguities have been fixed correctly (*ambiguity validation*). This question can be formalized in a probabilistic measure, the *ambiguity success rate* (Joosten, Tiberius, 2000). Since ambiguities are determined from noisy data, the estimated integer ambiguities can be treated as stochastic variates, similar to standard adjustment practice. If the success rate is sufficiently high, for example 99%, it is likely that the correct integer has been found. For details see Teunissen (1998); Joosten, Tiberius (2000). Other procedures are to compute *contrast* or *ratio* values between the best and the second best solution, namely the two smallest values of the square root sum of residuals. Only when the ratio of these two values exceeds a selected threshold, the solution with the smallest value is chosen as the correct solution (Hatch et al., 2000).

With today's satellite coverage it is possible to extend the observation time at a site up to several hours or even days. In such cases the *ambiguity float solution* already provides excellent results, also for large interstation distances, without the necessity to resolve the integer ambiguities. For applications like the establishment of fundamental geodetic control or the monitoring of crustal deformation the problem of ambiguity fixing therefore is of minor relevance.

7.3.3   Data Handling

7.3.3.1   Cycle Slips

Cycle slips occur if the receiver loses phase lock of the satellite signal. The reasons for cycle slips may be
observation dependent, e.g.:

   − obstructions, in particular for kinematic observations,
   − signal noise, in particular caused by multipath and ionospheric scintillation,
   − low satellite elevation, causing low signal strength,

or *receiver dependent*, e.g.:

   − weak signals, partly caused by signal interference,
   − antenna inclination in kinematic application (airplane, ship),
   − caused by signal processing.

In a cycle slip, the carrier phase shows a sudden jump by an integer number of cycles; the fractional part of the phase observable remains unchanged (Fig. 7.41). The cycle slip may be as small as one or a few cycles, or contain millions of cycles.

Cycle slips have either to be removed from the data at the preprocessing level, or a new ambiguity has to be determined for the particular pseudorange.

A cycle slip can easily be detected if double and triple differences are formed. This is demonstrated in Table 7.11. The notation corresponds to Fig. 7.35 and to chapter [7.3.2.1]. A cycle slip SL is introduced into the phase observation between station j and satellite p at epoch t. All single and double differences, starting with epoch t, are corrupted by the cycle slip whereas only one of the triple differences is affected. It is evident that the triple difference technique of fixing cycle slips belongs to the very early and classical methods (e.g. Remondi, 1985).

Figure 7.41. Representation of a carrier phase cycle slip

Two aspects have to be distinguished: *cycle slip detection* and the elimination of cycle slips from the data, the *cycle slip repair*, also denoted as *cycle slip fixing*. Most modern receivers have built-in algorithms that identify all or most of the cycle slips, and indicate (*flag*) the slips in the data set. These indications are very helpful for data preprocessing. Advanced receivers also have sophisticated implementations of the phase lock loops with fewer occurrences of cycle slips (Misra, Enge, 2001). Cycle slip repair belongs to the process of data editing, either automatic or interactive.

Several methods are in use and have been widely discussed in the literature. For a review see Gao, Zuofa (1999); Hofmann-Wellenhof et al. (2001) and Bisnath et al. (2001). The main differences come from

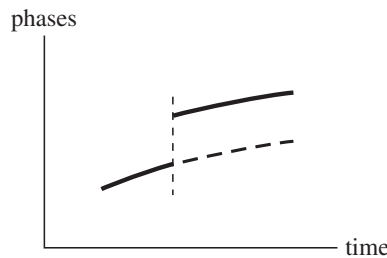   − the available data sets (single or dual frequency, codes),

Table 7.11. Effect of a cycle slip SL on single, double and triple differences

Carrier Phases			
$\Phi_i^p(t-2)$	$\Phi_j^p(t-2)$	$\Phi_i^q(t-2)$	$\Phi_j^q(t-2)$
$\Phi_i^p(t-1)$	$\Phi_j^p(t-1)$	$\Phi_i^q(t-1)$	$\Phi_j^q(t-1)$
$\Phi_i^p(t)$	$\Phi_j^p(t)+SL$	$\Phi_i^q(t)$	$\Phi_j^q(t)$
$\Phi_i^p(t+1)$	$\Phi_j^p(t+1)+SL$	$\Phi_i^q(t+1)$	$\Phi_j^q(t+1)$
$\Phi_i^p(t+2)$	$\Phi_j^p(t+2)+SL$	$\Phi_i^q(t+2)$	$\Phi_j^q(t+2)$
Single Differences			
$\Delta\Phi_{ij}^p(t-2)$		$\Delta\Phi_{ij}^q(t-2)$	
$\Delta\Phi_{ij}^p(t-1)$		$\Delta\Phi_{ij}^q(t-1)$	
$\Delta\Phi_{ij}^p(t)+SL$		$\Delta\Phi_{ij}^q(t)$	
$\Delta\Phi_{ij}^p(t+1)+SL$		$\Delta\Phi_{ij}^q(t+1)$	
$\Delta\Phi_{ij}^p(t+2)+SL$		$\Delta\Phi_{ij}^q(t+2)$	
Double Differences		**Triple Differences**	
$\nabla\Delta\Phi_{ij}^{pq}(t-2)$			
$\nabla\Delta\Phi_{ij}^{pq}(t-1)$		$\delta\nabla\Phi_{ij}^{pq}(t-1,t-2)$	
$\nabla\Delta\Phi_{ij}^{pq}(t)-SL$		$\delta\nabla\Phi_{ij}^{pq}(t,t-1)-SL$	
$\nabla\Delta\Phi_{ij}^{pq}(t+1)-SL$		$\delta\nabla\Phi_{ij}^{pq}(t+1,t)$	
$\nabla\Delta\Phi_{ij}^{pq}(t+2)-SL$		$\delta\nabla\Phi_{ij}^{pq}(t+2,t+1)$	

 – the available a priori knowledge (station coordinates, satellite coordinates), and
 – the kind of observations (static or kinematic).

Of particular interest are methods that can be applied for a single receiver and hence can be used for single station preprocessing.

Cycle slips cannot be identified and corrected from a single series of phase measurements alone. Basically, the phase observations are compared and combined with other quantities, and the behavior of the differences is analyzed. The differences can be regarded as "test quantities" that must show a smooth behavior. Any discontinuity in the time series of the test quantities indicates a cycle slip. Different approaches are in use to identify the discontinuities:

 – a low degree *polynomial* is fitted to the time series; this method is widely used,
 – a dynamic model is set up to predict subsequent observations by *Kalman filtering* (in new approaches also by *wavelets*); a comparison between predicted and observed data indicates possible cycle slips, and
 – a scheme of first, second, third and fourth differences can be set up; discontinuities show rather strong signals in the higher order differences.

The methods applied have not changed much since the 1980s, hence it is worth studying the cited publications. In the following some of the principal procedures are described:

(a) analysis of double differences and computed ranges,
(b) analysis of ionospheric residuals,
(c) analysis of code/carrier combinations, and
(d) methods usable in kinematic applications.

(a) *Analysis of double differences and computed ranges*
This method requires knowledge of approximate satellite and station coordinates. Slant ranges R_{ij}^{pq} computed from the coordinates are compared with the observed double differences. The test quantity is the residual, $r_{\nabla\Delta}(t)$:

$$r_{\nabla\Delta}(t) = \nabla\Delta\Phi_{ij}^{pq}(t) - \frac{1}{\lambda}\left[R_j^q(t) - R_j^p(t) - R_i^q(t) + R_i^p(t)\right] + \nabla\Delta A. \quad (7.95)$$

The time dependent behavior of the test quantity, $r_{\nabla\Delta}(t)$, is analyzed by fitting a curve to the series, e.g. a low degree polynomial. A discontinuity in the curve fit identifies the cycle slip.

The term $\nabla\Delta A$ contains the residual atmospheric effect, in particular due to the ionosphere. The ionosphere either has to be modeled, or neglected for short interstation distances. The method hence suffers from ionospheric variations. Other disadvantages are that two stations are involved, and that approximate coordinates are required for stations and satellites. The method is furthermore rather sensitive to observation noise. The main advantage is that the technique works with single frequency receivers. For more details see e.g. Bock et al. (1986); Beutler et al. (1988); Lichtenegger, Hofmann-Wellenhof (1990).

(b) *Analysis of ionospheric residuals*
If dual frequency receivers are available, the differences between the L1 and the L2 signal can be analyzed. Based on a proposal of Goad (1985), see also Cross, Ahmad (1988); Landau (1990), the following test quantity may be used:

$$r_\Phi(t) = \Phi_{L1}(t) - \frac{f_{L1}}{f_{L2}}\Phi_{L2}(t). \quad (7.96)$$

The main advantage of this test variable is that only frequency dependent parameters are present in equation (7.96). The variable, $r_\Phi(t)$, only depends on the ionosphere (and multipath), and varies very slowly with time. No a priori knowledge of station or satellite coordinates is required. This "geometry-free" method works for a single receiver and can hence be applied in undifferenced phase processing [7.3.4].

If cycle slips ΔN_{L1}, ΔN_{L2} occur on both frequencies they cannot uniquely be identified because they are connected by the Diophantine equation,

$$\Delta N = \Delta N_{L1} - \frac{f_{L1}}{f_{L2}} \cdot \Delta N_{L2}. \quad (7.97)$$

ΔN is the total cycle slip effect resulting from an analysis of equation (7.96). Equation (7.97) can be solved if the cycle slips in the L1 and L2 signals are known from other methods with an accuracy of about 6 to 8 cycles (e.g. Bastos, Landau, 1988).

(c) *Analysis of code-carrier combinations*

For low-noise code receivers the unambiguous code result can be compared with the phase measurements. The difference between both data sets depends only on the number of integer ambiguities, N:

$$r_{CD,\Phi}(t) = \lambda\Phi + \lambda N - R_{CD}(t). \qquad (7.98)$$

The equation holds for both carriers. With a code noise level of the order of ± 0.6 m, the method allows the determination of cycle slips to about ± 3 cycles. This is sufficient for input values to method (b).

If the noise level of the code measurements is sufficiently low, the cycle slips can be determined and removed immediately. The same is true for methods that smooth the P-code measurements over a few epochs and allow quasi-real-time ambiguity resolution through code-carrier combination [7.3.2.3]. In such cases the cycle slip problem no longer exists, because a new ambiguity can be introduced for each cycle slip, and be immediately resolved. The main advantages of code-carrier combination are the simplicity of the underlying model and the possibility of using it in kinematic applications (e.g. Landau, 1988).

A new proposal uses the widelane carrier phase minus the narrowlane pseudo-range as test quantity. This observable is geometry-free as well as ionosphere-free. It is insensitive to receiver motion with good results for automatic cycle slip correction both in static and kinematic mode (Gao, Zuofa, 1999; Bisnath et al., 2001).

(d) *Methods usable in kinematic mode*

A particular situation evolves for kinematic surveying, i.e. if at least one of the participating receivers is moving [7.3.5.3], [7.3.5.4]. In early applications, the integer phase ambiguities were resolved before the start of the moving receiver. As long as the phase lock is maintained, kinematic positions of the moving antenna can be derived. If more than four satellites are tracked, and at least four satellites continue to stay on lock, the cycle slips can be corrected with respect to the remaining satellites, and the survey can continue.

Loss of carrier lock is detected by most GPS receivers within 20 milliseconds. Once a momentary loss of lock is detected a multichannel state-of-the-art GPS receiver should be able to determine the number of carrier cycles which were lost when track has again been established. The redundancy available in the additional satellites permits the reconstruction in the missing carrier cycles. In addition, if separate tracking of both carriers is performed within the receiver, this information can be used to determine the number of slipped cycles on one carrier, as long as both carriers do not lose lock over the same time period. If less than four satellites are left, the following procedures can be applied to recover the ambiguities.

1. Return to the last coordinated point and start with a new determination of the initial phase ambiguity. This early method is not practicable for marine and airborne applications, and it is today also not used for land applications because of the availability of efficient and rapid ambiguity fixing techniques.

2. With dual frequency receivers the ionospheric residuals (method (b)) can be used, together with a Kalman filter, to predict the dynamic behavior of the observation platform and to close the gaps in the observations. The technique is very sensitive to high dynamics and low signal/noise ratio.

3. A new ambiguity "on the fly" (OTF) is determined after loss of lock. With low-noise code receivers the extra-wide laning technique can be applied. If a sufficient number of satellites can be tracked after the data gap (6 or more) then advanced ambiguity search techniques [7.3.2.3] are a powerful tool. For each cycle slip, a new ambiguity parameter is simply introduced into the adjustment.

4. The integration of an additional sensor helps to bridge the gaps caused by cycle slips. An external atomic clock (rubidium oscillator) replaces one satellite. An inertial sensor package can be used to interpolate the GPS positions if signals to particular satellites are shaded by obstructions (e.g. Colombo et al., 1999; Böder, 2002).

7.3.3.2   The Receiver Independent Data Format RINEX

Each receiver type has its own binary data format, and the observables are defined following the manufacturers' individual concepts. Time tags may be defined in transmission time, or in receiver time; phase measurement may be expressed in whole cycles, or in fractional parts of cycles; code and phase may have different or identical time tags, and satellites may be observed simultaneously or at different epochs.

As a consequence, data of different receiver types cannot easily be processed simultaneously with one particular GPS data processing software package.

To solve this problem, either all manufacturers have to use the same data output format, or a common data format has to be defined that can be used as a data interface between all geodetic receiver types, and the different processing software systems. The first has not been realized to date. However, a successful solution has been found to define and accept a common data format for international data exchange.

Based on developments at the University of Berne, Switzerland, the *Receiver Independent Exchange Format* RINEX was proposed by Gurtner et al. (1989) at the Fifth International Geodetic Symposium on Satellite Positioning in Las Cruzes. The proposal was discussed and modified during a workshop at this symposium, and recommended for international use. More discussions followed in 1989 and 1990, and brought some modifications and extensions to the data format. A review of the historical development is given by Gurtner (1994).

RINEX has indeed been accepted by the international user community and by the community of receiver manufacturers. For most geodetic receivers *translator software* is provided by the manufacturers that converts the receiver dependent data into the RINEX format. In addition, all major data processing software requires RINEX data as an input. RINEX hence serves as a general interface between receivers and multipurpose data processing software.

With RINEX, one of the most serious obstacles to the routine mixing of data from different receiver types is removed. It is an important precondition for large

international cooperative projects like the IGS [7.8.1], and it found its first important application in the EUREF campaign in 1989 for the establishment of the European Reference Frame [7.6.2.1].

Since the first publication, in 1989, several revisions and modifications have been introduced. The current revision status is version 2.10. A detailed document is available, for instance via the IGS server (Gurtner, 2001). The following definitions are taken from this document.

RINEX defines three fundamental quantities in the GPS observables: *Time, Range*, and *Phase*. The *time* of measurement is the receiver time of the *received* signals. It is identical for the phase and range measurements and is identical for all satellites observed at that epoch. It is expressed in GPS time (not in UT).

The *pseudorange* is the distance from the receiver antenna to the satellite antenna, including receiver and satellite clock offsets and other biases:

$$
\begin{aligned}
\text{Pseudorange} = {} & \text{Geometrical distance} \\
& + c \cdot (\text{RCVR clock offset} - \text{Satellite clock offset} + \text{Biases}),
\end{aligned}
\tag{7.99}
$$

so that the pseudorange reflects the actual behavior of the receiver and satellite clocks. The pseudorange is written in units of meters and is unambiguous; i.e., C/A code ranges add the correct number of milliseconds to obtain the definition of pseudorange given above.

The *phase* is the carrier phase measured in whole cycles at both L1 and L2. The half-cycles measured by squaring-type receivers must be converted to whole cycles and this fact is noted by the wavelength factor in the header records. The phase changes in the same sense as the range (negative Doppler); i.e. range increases equal phase increases. The phase observations between epochs must be connected by including the integer number of cycles. The phase will not contain any systematic drifts from intentional offsets of the reference oscillator.

The observables are not corrected for external effects like atmospheric refraction, satellite clock offsets, etc. The sign of the *Doppler shift* as additional observable is defined as usual, namely positive for approaching satellites.

The basic RINEX format consists of three ASCII file types:

1. *Observation Data File*
2. *Meteorological Data File*
3. *Navigation Message File*.

Each file type consists of a header section and a data section. The observation file usually contains the data collected by one receiver at one station during one session. Since RINEX version 2 it is also possible to include observation data collected in sequence by a roving receiver during rapid static or kinematic surveys. From the long list of revision details only some major items are indicated:

 − inclusion of GLONASS data (since 1997),
 − continuous numbering of the GPS week; no rollover (1998),
 − inclusion of navigation data from GEO satellites (2000), and
 − inclusion of navigation data from LEO satellites (2001).

For detailed information see the cited documents, in particular Gurtner (2001).

RINEX is the international exchange format for the postprocessing of GPS data. For the transmission of data corrections, in real-time, in relative (Differential) GPS applications, a particular data format is available: the RTCM format. Details of this data format are given in the section on Differential GPS (DGPS) [7.5.1.2].

7.3.4 Adjustment Strategies and Software Concepts

All observations made simultaneously and continuously in the course of a particular GPS project are called a *session*. A session may be as short as a few minutes, if fast ambiguity resolution techniques are applicable in small networks, or it may last several hours, or even days, if the highest accuracy is wanted in larger networks. During the development phase, with a limited number of satellites available, a typical observation session lasted between one and three hours. Since continuous worldwide coverage was established in 1993/1994, sessions can last several days. For practical reasons and for analysis purposes it may be advisable to break down the complete data set into individual sessions of several hours, for instance one session per day or three sessions of eight hours each per day. The following observation and evaluation strategies are in use:

(a) *single-station* adjustment,
(b) processing of single *baselines* and subsequent combination of baselines into networks,
(c) processing of all simultaneously-observed data of a single session in a joined adjustment (*multi-station adjustment*), and
(d) combination of several session solutions into a rigorous overall network solution (*multi-session adjustment*).

The *single-station adjustment* (a) provides absolute station coordinates referred to WGS84. Because of the low accuracy [7.4.1], the results are of little interest to geodetic applications, but they often meet the requirements for some tasks in geophysical prospecting, GIS data acquisition, or in remote sensing. The typical application field is navigation (cf. [7.6.2.7]).

In a rigorous geodetic adjustment relative *and* absolute information [7.4.3] is required. This is why a single-station solution is incorporated into many software packages for multi-station post-processing. The single-station adjustment is also used for preprocessing and editing the data (e.g. because of cycle slips, Earth rotation, relativity, ionosphere, troposphere and formation of normal points), before they enter the level of multi-station adjustment. More accurate absolute positions, at the level of a few meters or better, can be achieved if data from several days of observation are used. Along with the accurate modeling of orbits and clocks [7.4.3] the concept of *Precise Point Positioning* (PPP) has been developed (see below).

The *single baseline concept* (b) was widely used in early software development for the processing of GPS data. The observations from two simultaneously-operating receivers are processed in a joint adjustment, mostly in double difference form [7.3.2.1].

Results are the components ΔX, ΔY, ΔZ of the baseline vector and the associated variance-covariance matrix.

The individual baselines can be used as input data for a network adjustment program and combined into larger networks. The procedure is rigorous, if only two GPS receivers observe simultaneously and if all the stochastic information of the complete variance-covariance matrix is exploited. However, if the station pairs are selected from a larger number of simultaneously operating receivers, the possible baseline combinations are not all independent of each other. Fig. 7.42 gives an example for the case of three receivers. If baselines a and b are considered as independent, baseline c is not. It is called a *trivial baseline* because it can already be derived from the results of baselines a and b. A general rule is given in terms of the number of simultaneously operating receivers, r:
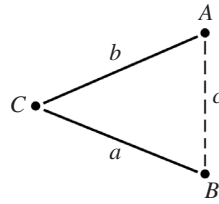
Figure 7.42. Independent and trivial baselines in the case of three simultaneously operating receivers

$$r(r-1)/2 \quad \text{number of possible baselines, and} \qquad (7.100)$$

$$(r-1) \quad \text{number of independent baselines} \qquad (7.101)$$

If only baseline processing software is available, the independent "non-trivial" baselines have to be identified using suitable selection criteria like the baseline length or number of observations. Nevertheless, the procedure is not rigorous for network solutions because the stochastic information between the simultaneously observed baselines is neglected. Careful weighting and decorrelation is necessary to improve the solution (e.g. Goad, Müller, 1988).

In the simplest case of the baseline concept only the length of the baseline vector is utilized, and GPS is used as a method of distance measurement. This technique was frequently applied in the early years of geodetic exploitation of GPS. The procedure is correct; however one part of the information contained in GPS, namely the spatial orientation of the baseline, is neglected.

Most manufacturers offer software along with the receiver equipment that utilizes the baseline concept. This software is suitable for small projects, for in-the-field data verification and for real-time kinematic (RTK) surveying [7.3.5.4], [7.5.2].

In a *multi-station adjustment* (c) all data that have been observed simultaneously with three or more participating receivers are processed jointly. No baselines are determined, but rather coordinates in a network with the associated complete variance-covariance matrix. Hence, it is a rigorous adjustment of the observations using all mutual stochastic relationships. For geodetic purposes, the multi-station adjustment has conceptual advantages over the baseline approach, because the accuracy potential of GPS is completely exploited. If the observations stem from one session, one speaks of a *session-solution*.

Several session-solutions can be combined into a *multi-session adjustment* (d) or, more precisely, into a *multistation-multisession solution*. This is the usual procedure, if larger networks have to be broken into parts because of a limited number of available GPS receivers. The basic condition is that each session is connected to at least one other session of the network through one or more identical stations where observations have been carried out in both sessions. An increasing number of identical stations increases the stability and the reliability of the total network (cf. [7.6.1.3]).

The multi-session solution is completely rigorous and equivalent to an all-in-one joint adjustment, if the variance-covariance matrices of the individual session solutions are properly used. The stepwise procedure, starting with session solutions, has the advantage of requiring less computer capacity. In addition, comparison of the individual session results provides an excellent insight into the network's accuracy if sufficient redundant observations at identical stations have been included. Software packages for GPS data processing of large networks are usually based on the multistation-multisession concept.

The development in the field of GPS software is fast, hence only a few basic considerations are made here.

A first classification is possible into *commercial software*, provided from receiver manufacturers, and multipurpose *scientific software*, that originates from developments at scientific institutes. Software of the first group is primarily designed for processing of data from a particular receiver type. Advanced packages, however, also accept data from other receivers via the RINEX interface. As a rule, only the executable object code is available to the user, and the basic mathematical models are mostly not documented in detail. Commercial software is adequate for everyday surveying work. It usually offers a large variety of possible applications and can be operated easily enough by personnel with an average background in engineering and GPS technology. In some cases the basic software includes baseline adjustment, and additional software is necessary for network computation. Usually this kind of software allows for static and kinematic [7.3.5] applications and includes extensive mission planning capabilities [7.6.1.1]. The "Real-Time Kinematic" (RTK) capability with OTF ambiguity techniques (see [7.3.5.4]) is today considered to be a standard option. Current examples of this first group are

    SKI-Pro   from Leica Geosystems
    TGO   (Trimble Geomatics Office) from Trimble Navigation, and
    Pinnacle   from Javad/Topcon Positioning Systems.

The development of a general-purpose GPS post-processing system (second group) is a major operation (Beutler et al., 1990). It requires several man-years of development and consists of a large number of individual programs, adding up to tens of thousands of lines of code. Usually, these software packages are not restricted to just one receiver type but accept data from a large variety of geodetic receivers. The packages serve in most cases for

    − professional standard use in smaller networks for rapid processing,
    − professional use in high accuracy surveys, also over large distances,
    − scientific use in research and education, and

   − data analysis and scientific investigations including geodynamic research and
      analysis of permanent arrays.

Besides the standard options for rapid processing, these kinds of software packages
offer many particular alternatives for scientific processing. Interactive operation is
essential. Some packages include options for orbit determination, or the estimation of
atmospheric models.

   Scientific processing requires a lot of experience and a deep understanding of
GPS signals and error behavior. Data processing is particularly difficult if the data
are contaminated by ionospheric disturbances (cf. [7.4.4.1]) and when the highest
accuracy over large distances is required from noisy data. The mathematical models,
and the rationale behind the scientific general-purpose software packages, are in most
cases well documented and discussed in published literature. In some cases, the user
has access to the source code and can include modifications or new parts. Current
examples of this second group of GPS software are

BERNESE   developed at the University of Berne, Switzerland
                      (e.g. Beutler et al., 1988; Hugentobler et al., 2001)

GEONAP   originally developed at the University of Hannover, Germany
                      (e.g. Wübbena, 1989, 1991), and

GIPSY-OASIS II   developed at the Jet Propulsion Laboratory, USA
                      (e.g. Blewitt et al., 1988; Webb, Zumberge, 1993).

A multi-purpose software package consists of several parts. Three main groups can be
identified:

   − the *pre-processor level* prepares the data for the main processing,
   − the *main-processor level* deals with the estimation of unknown parameters, and
   − the *post-processor level* summarizes various information in tables or in graphical
      form, and combines sessions to networks, if required.

Fig. 7.43 shows a simplified functional flow diagram of a generic software package
for static multistation - multisession GPS processing. The structure of the GEONAP
software is very similar to this diagram.

   Starting from the *raw data* of all receivers involved in a single session, these data
must be acquired, translated into a readable ASCII format, and tested for rough errors
(blunders). In most cases the RINEX format is used as a data interface between receiver
and software. RINEX requires additional information that is not always provided by
the receivers, e.g. antenna height, approximate station coordinates, meteorological
data etc. These data can be introduced into a *database*.

   The broadcast message can be separated from the observation data, checked and
organized in a session dependent *message file*. Smoothing algorithms [3.3.3.2] for
different portions of the message can be applied. At this level, external orbital infor-
mation, e.g. IGS orbits [7.4.3.2], [7.8.1], can be introduced if required.

   Single station solutions from code measurements or carrier-smoothed pseudo-
ranges are usually generated at the level of the *main program*. Necessary data correc-
tions can be applied at this stage; these are, for example, corrections for ionosphere,
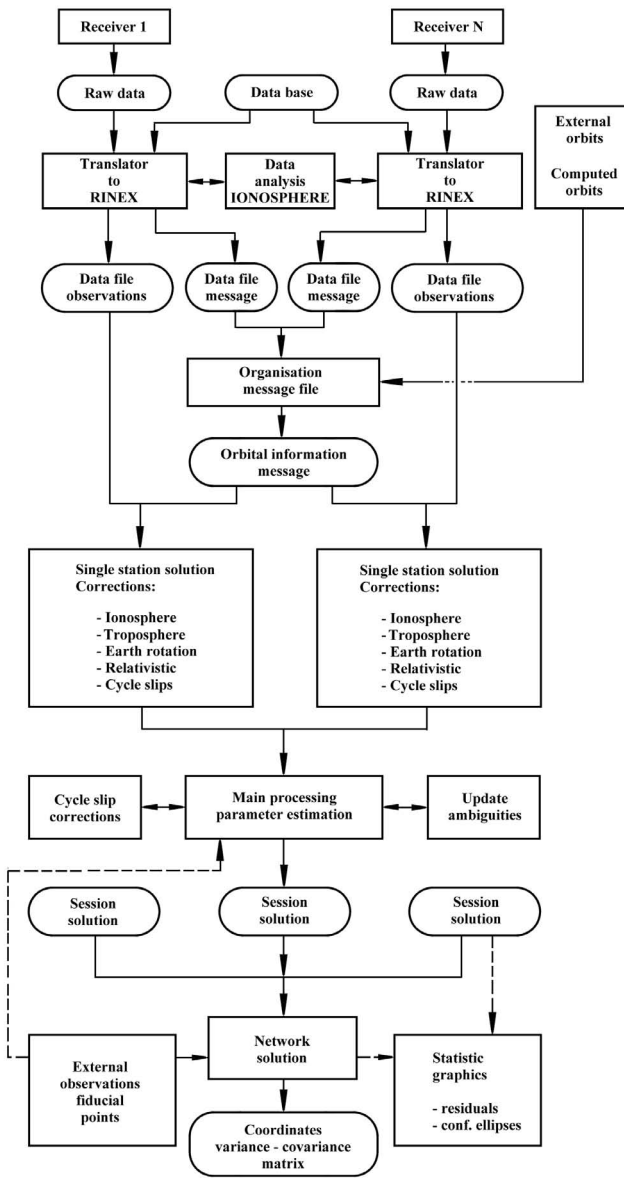
Figure 7.43. Simplified functional flow diagram of a generic GPS software package

troposphere [7.4.4], antenna phase center [7.4.5], Earth rotation and relativistic effects [7.4.1]. The data can be controlled for cycle slips, and be condensed to normal points [3.3.3.2].

Parameter estimation in the *main program* follows either the concept of parameter elimination, or parameter determination, or both [7.3.2]. Basic observables may be the undifferenced phase data (e.g. GEONAP, GIPSY-OASIS II), or double differences (e.g. BERNESE).

If double differences are used, sophisticated decorrelation and weighting techniques have to be applied (Goad, Müller, 1988; Beutler et al., 1990) to account for the mathematical correlations between the double-differences in a GPS network. It may also be advisable to generate an optimum set of independent double difference observables with respect to the shortest interstation distances and minimum influences of data gaps at individual receivers (Rothacher, Mervart, 1996).

Parameters to be estimated can be
- station coordinates,
- biases of satellite clock and receiver clock (second degree polynomial plus stochastic clock model [2.2.5]),
- hardware delays in the satellite and receiver electronics (second degree polynomial plus 1 stochastic parameter),
- orbit improvement with short arc model (e.g. up to 6 Keplerian parameters, solar pressure model),
- tropospheric scale parameter for each station (constant or stochastic process),
- local ionospheric corrections (e.g. improved parameters in the Klobuchar model, time varying parameters of a single layer ionospheric model), and
- parameters for each non-resolved ambiguity and each unrecovered cycle slip.

More parameters can be included. The Bernese GPS software version 4.2 allows, among others, to estimate precise orbits, Earth rotation parameters, precise ionospheric models and the precipitable water content of the atmosphere (Hugentobler et al., 2001). The adjustment process in the main program can be repeated in order to fix as many ambiguities as possible. The principal results of the session solution are the coordinates of all participating stations and the variance-covariance matrix.

At the *post-processing level*, all session solutions can be rigorously combined in a network adjustment program, if this is not already provided in the main program. The results of the network solution can be re-substituted into the session solution with the objective of fixing remaining ambiguities. External information, like coordinates of fiducial stations, can be introduced at this level. Also, improved orbits can be introduced again into the ephemeris file, and a second processing run can be initiated.

The final results from the network solution can be transformed into a local or global datum [2.1.5], [12.1], and compared or combined with existing terrestrial data sets [7.6.2.1]. A final evaluation of the results is supported by statistical tests and a series of graphical representations. These may include statistics on usable simultaneous observations, and the behavior of particular signals and linear combinations, such as the ionospheric signal (7.80).

A new development is *Precise Point Positioning* (PPP). This is a powerful strategy for estimating the coordinates of a single station using precise satellite orbits and satellite clocks. The necessary information can be taken either from the IGS [7.8.1] or from other sources like the JPL [7.4.3.2].

The idea behind PPP is as follows: Precise orbits and satellite clocks are estimated, based on observations from a high quality global fiducial network. This information is taken to solve for station parameters of any site in the world (position, clock and wet troposphere). Only one single station is processed at a time. A disadvantage is that the method is unable to take account for correlations between stations, and that orbits are assumed to be perfect, which is not true. The final formal errors hence have to be scaled to more realistic values.

If only code measurements are used, the observation noise level is well above the precision of orbits and clocks. The positioning accuracy hence depends mainly on the code observation and reaches several decimeters. With the use of phase measurements (e.g. with GIPSY-OASIS II software) the achievable accuracy is comparable to the accuracy in global networks, namely 1 cm or better (e.g. Völksen, 2000).

One particular advantage of the PPP strategy, when compared with network adjustment, is that the processing time increases only linearly with the number of stations. Note that original undifferenced observables are essential for this approach.

For international cooperation in the processing of global and regional networks it is necessary to exchange the results from processing centers, that use different software packages, with the objective to combine global and regional solutions. To this end a *Solution (Software/Technique) Independent Exchange Format*, SINEX, has been defined (IGS, 1996). The SINEX format contains coordinate estimates and the corresponding covariance information, as well as additional information like receiver types, antenna types, phase center values, eccentricities, and a priori weights. SINEX is mainly used by the IGS community [7.8.1].

Information about the latest developments in the software sector can be taken from the proceedings of the series of GPS symposia, such as ION GPS and IAG Symposia, or from related journals (e.g. *GPS World*, *GPS Solutions*, also *Journal of Geodesy*).

7.3.5   Concepts of Rapid Methods with GPS

7.3.5.1   Basic Considerations

Various techniques have been developed in recent years that exploit the capability of GPS to provide precise coordinates after a very short observation time, or even while the receiver (including the antenna) is moving along a trajectory. Sometimes misleading, the related rapid methods were named *kinematic GPS*. In addition, different terms describing particular types of rapid GPS surveying procedures have been created, such as *semi-kinematic, pseudo-kinematic, true kinematic, rapid kinematic, pseudo-static, stop-and-go kinematic* etc. In some cases, different terms were used to describe the same procedure or, more confusing, the same term was used to describe different procedures. The related literature must therefore be read with care. A clarifying overview is given by Kleusberg (1990).

Rapid methods require the resolution of ambiguities in order to exploit the high accuracy potential of GPS phase measurements. Otherwise the noise level of real-valued solutions for the short observation times would be too high. One prerequisite

for the rapid solution of ambiguities is that the distant-dependent errors (see [7.4.3], [7.4.4]) be small. Hence, the rapid methods only work well for short distances (up to several kilometers) between the participating stations. For longer ranges, it is necessary to model the distance dependent errors, e.g. in active reference networks (see [7.5.3]).

Different possibilities exist for subdividing the rapid methods of GPS. The scheme used here is into
  − rapid static methods,
  − semi kinematic (stop and go) methods, and
  − pure kinematic methods.
The rationale behind this subdivision is whether the receiver is taking measurements while it is in motion, and the coordinates of the trajectory can be determined (*kinematic mode*), or whether the receiver is switched off during transportation, and coordinates can only be determined when the antenna is stationary (*static mode*). A third mode is in between these possibilities, in that the receiver has to maintain lock during the times of transportation, but coordinates are not usually derived for the trajectory (*semi-kinematic mode*).

A further distinction between static and kinematic surveying can be seen with respect to accuracy issues (Kleusberg, 1990). In static GPS surveying, most random measurement errors are absorbed in the residuals after adjustment, while in kinematic surveying, most random measurement errors are absorbed in the coordinates. This is why the accuracy potential of static GPS cannot completely be reached with pure kinematic methods.

Only precise methods are considered here, i.e. with an accuracy level of a few centimeters, for kinematic surveys. This implies the use of carrier phase data as the basic observables. Less accurate methods for determining coordinates of a trajectory, i.e. when code measurements are used as primary observables, are discussed in [7.3.6] and [7.5.1]. The dividing line between kinematic and navigational use of GPS is debatable.

As in nearly all geodetic applications at least two receivers are needed to determine relative coordinates. In the concept of rapid methods, one receiver usually remains fixed during the operation, while a second receiver, the *roving* receiver, moves between stations or along a trajectory.

The first two methods were frequently applied after 1990, when GPS had developed into a powerful technique for detailed surveying (cadaster, GIS). With improving satellite coverage after 1994, and the availability of rapid OTF algorithms, currently mainly the third (pure kinematic) method is applied for local surveys. The stop-and-go technique has nearly disappeared from use. It is, however, explained in this book because of its conceptual importance and because it is still offered by a number of manufacturers.

7.3.5.2   Rapid Static Methods

Two different modes can be distinguished (Fig. 7.44):
  (a) rapid static mode with single station occupation

(b) rapid static mode with station *re-occupation* after about one hour.

In the first mode (a) fast ambiguity resolution techniques are required (cf. [7.3.2.3]). These can be for example
— code/carrier combination with dual frequency, low code-noise receivers, and
— ambiguity search methods with 6 and more satellites.

Figure 7.44. Modes of rapid static surveying

Basically, the same techniques are used as for classical static positioning. Depending on the receiver type, satellite coverage, and interstation distance, observation times of several minutes up to 15 minutes are sufficient. The method is particularly powerful over short distances, with dual-band low noise receivers, a high number of visible satellites, and fast ambiguity resolution algorithms. The key factor is the necessary time to fix ambiguities (TTFA) and the *ambiguity success rate*. The procedure is very flexible and effective, and is widely used in surveying applications [7.6.2.4], mostly together with near real-time data processing.

Rapid static applications are of particular interest with respect to reference services like SAPOS [7.5.1.3]. In order to augment the possible distance to the nearest reference station, concepts like *active networks* or *virtual reference stations* play an increasing role.

The demand for surveying equipment with rapid static capability will grow further. Equipment will be assessed mainly on the basis of its capability to resolve ambiguities and to provide precise position results after as short a time as possible.

In the second mode (b) each station has to be re-occupied after an interval of about 50 to 120 minutes. The observation time required at each station is relatively short, about four to eight minutes. Tracking during the transitions is not necessary; the receiver might be turned off while traveling. The rationale behind this procedure is that data from a different geometric configuration are required to resolve the ambiguities (geometrical method, [7.3.2.3]), but not because there is a need for extra observations. Ten minutes of data are completely sufficient to absorb most random measurement errors in the adjustment residuals. Both data sets are considered as one set with one cycle slip in between. The same processing software is used as in static GPS surveying; the "cycle slip" can be fixed with triple and double difference techniques.

Cycle slip fixing over more than 30 minutes, however, only works properly if the data quality is high (low noise, low multipath, low ionospheric effects) and if the repeated station occupations are identical (forced centering). A further requirement is that the same satellites have to be observed for both station occupations.

The re-occupation method was widely used in the early years of rapid GPS methods. Today it has nearly disappeared from the surveying market because of its rather complicated procedure and the high efficiency of method (a). Note that it is a static
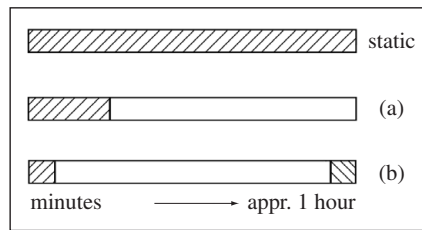
method, although the procedure is frequently called *pseudo-kinematic*. Other names are *broken-static*, *intermittent-static*, or *snapshot-static*.

7.3.5.3   Semi Kinematic Methods

This procedure can be traced back to the pioneering work of Remondi (1984, 1986), and it is often also referred to as *stop and go surveying* or simply *kinematic surveying*. The fundamental idea is that there is basically no difference between static and kinematic methods once the ambiguities are resolved and maintained. Kinematic surveying can hence be understood as the "transfer of ambiguities from one station to the other". In the concept of "semi-kinematic" or "stop and go" surveying, the antenna is mounted for a short time (seconds to minutes) on the survey marker that is to be coordinated. The required stop-time comes rather from the need to identify the station and to mount the antenna pod vertically than from the need to gather sufficient GPS data at the station.

The trajectory between stations is usually of no interest (other than in the concept of "pure kinematic surveying" [7.3.5.4]). However, it is essential that phase lock to at least four satellites with a geometrically strong configuration is maintained during antenna movement. The fundamental problem is to determine the initial phase ambiguity before the survey starts. This can be achieved with static initialization procedures. Three main techniques have been developed:
  (1) determination of a start baseline with a static survey prior to kinematic operations,
  (2) short observation on a known baseline, and
  (3) antenna swapping.
The *first method* is rather time consuming, but rapid static methods [7.3.5.2] can be applied as well. The *second method* requires a precisely known baseline. Note that the three dimensional Cartesian coordinate differences of the baseline must be predetermined with an accuracy level of a few cm. The method is very fast because only data from about one minute of observations are required. After initialization, one receiver remains at the station and the second (roving receiver) starts with the survey.

The mathematical background is simple. With the notation of (7.64) the double difference carrier phase data, collected for a short time, are given by

$$\nabla \Delta \Phi = \nabla \Delta R + \lambda \nabla \Delta N. \tag{7.102}$$

The range double difference, $\nabla \Delta R$, can be computed because the station coordinates and the satellite coordinates are known. The computed double difference range is subtracted from the observed carrier phase double difference and yields the ambiguity

$$\nabla \Delta N = (\nabla \Delta \Phi - \nabla \Delta R)/\lambda. \tag{7.103}$$

The *third method* (Fig. 7.45) has been widely used because it is fast, precise, reliable, and it does not require a priori knowledge of a baseline. The procedure was first described by Remondi (1985).

Two receivers are set up close-by at two stations. One station should have known coordinates because it serves as a reference station for the survey. The second station can be arbitrarily selected. About one minute of common data are observed. Then, both antennas are exchanged, maintaining phase lock on the satellites, and again about one minute of data is collected. One receiver remains at the reference station, and the second (roving) receiver is taken to the survey markers.
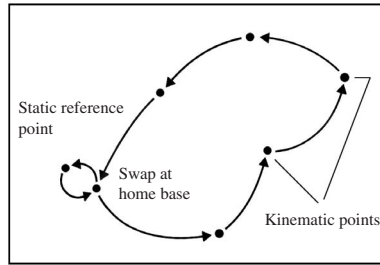
Figure 7.45. Antenna swapping technique

Again, the mathematical background is simple. The first antenna set-up yields the ambiguous double difference phase observation given by

$$\nabla\Delta\Phi(t_1) = \nabla\Delta R(t_1) + \lambda\nabla\Delta N. \tag{7.104}$$

After the antenna exchange the observations provide a second double difference equation:

$$\nabla\Delta\Phi(t_2) = -\nabla\Delta R(t_2) + \lambda\nabla\Delta N. \tag{7.105}$$

Subtraction of the observation equations eliminates the ambiguity term so that

$$\nabla\Delta\Phi(t_1) - \nabla\Delta\Phi(t_2) = \nabla\Delta R(t_1) + \nabla\Delta R(t_2). \tag{7.106}$$

From equation (7.106) the baseline can be determined and substituted into equation (7.103) (known baseline method), to derive the ambiguity term $\nabla\Delta N$.

In essence, equation (7.106) describes the triple difference. Without the antenna exchange, the right-hand side of (7.106) would nearly vanish, because the sum would change to a difference of nearly two identical quantities, and hence could not be solved. The geometrical message of equation (7.106) is that, due to the antenna swap, the geometry has changed enough to solve the triple difference equation after a very short time. A detailed derivation of the antenna exchange technique, with complete expressions, can be taken from Remondi (1985) or Hofmann-Wellenhof, Remondi (1988).

The semi-kinematic method is highly efficient in open areas where no loss of lock, due to signal obstructions, is to be expected. In cases where cycle slips occur, and cannot be recovered, the roving receiver has to go back to the last coordinated survey mark. The stop and go method is still meaningful for local surveys, when two GPS receivers are available but no data link for real-time application. Otherwise, today, RTK techniques [7.3.5.4], [7.5.2] or the use of active reference network services [7.5.3.2] is advisable.

7.3.5.4   Pure Kinematic Method

For many purposes, precise coordinates of the trajectory of a moving GPS receiver have to be determined, in particular in marine and airborne applications. In these cases, a loss of lock without the possibility of recovering cycle slips or ambiguities while the antenna platform is moving cannot be accepted. Hence, methods are required that are independent of static initialization techniques, and that include the capacity to recover cycle slips and/or to resolve ambiguities during motion. These techniques are referred to as ambiguity solution *on the way* (Seeber, Wübbena, 1989), or *on the fly* (e.g. Abidin, Wells, 1990; Cocard, 1995). Only with such methods at hand can kinematic surveying be *purely* or *truly kinematic*. Suitable methods for ambiguity resolution while the receiver is moving are [7.3.2.3]

- − code/carrier combination using the *extra wide laning technique*, and
- − *ambiguity search functions* for six or more satellites.

The efficiency of these techniques will be improved when low noise code receivers [7.2.5], or combined GPS/GLONASS receivers [7.7.1], are available, as well as new satellite signals [7.1.7].

For the ambiguity resolution on the fly a real-time data link with sufficient capacity is required [7.5.1.2]. Methods for cycle slip recovery in true kinematic mode are

- − use of redundant satellites (≥ 4 four satellites),
- − use of dual frequency data, and
- − use of code/carrier combination.

The inclusion of external sensors can support the recovery of cycle slips and the resolution of ambiguities as well, for example the use of (e.g. Lachapelle, 1990; Colombo et al., 1999; El-Sheimy, 2000; Böder, 2002)

- − high quality clocks (e.g. rubidium),
- − inertial navigation systems (INS), and
- − barometric altimeters.

The accuracy level of pure kinematic surveying is well below 10 cm and can reach a few centimeters under favorable conditions (satellite coverage, low noise receivers, no multipath, low platform dynamics).

The fields of application are broad and continuously broadening [7.6.2]. They include land, air, and ocean surveying, traffic and machine control, engineering-surveying and GIS. In many cases it is sufficient to process data afterwards. Most major software packages offer options for kinematic data. When the results are required in real-time, it is essential to establish a data link between reference and user station. Either the reference data are transmitted from an active network of control points like SAPOS [7.5.1.3], or a particular data link between a local reference station and the roving receiver is established. The latter solution is known as the RTK option (*Real Time Kinematic with OTF (On the Fly) ambiguity resolution*) and it is mainly used for limited ranges (several km). RTK systems are routinely applied for surveying tasks, and they form part of every modern GPS receiver system [7.2.4.2]. For more details about RTK technology see [7.5.2], [7.6.2.4]. Real-time kinematic methods over long distances are still under development.

When accuracy requirements are less demanding, it is not necessary to resolve the ambiguities, and the code measurements can be used as the primary observable. These techniques are discussed in the following section [7.3.6].

7.3.6   Navigation with GPS

GPS was primarily designed as a navigation system with a worldwide real-time capability. The following modes are in use (see also [7.5.1]):
   (1) absolute observations with code phases,
   (2) absolute observations with code and carrier phases,
   (3) relative observations with code phases,
   (4) relative observations with carrier-smoothed code phases, and
   (5) relative observations with code and carrier phases (the carrier phases are the primary observables).

Absolute observations with code phases (1) are important for general navigation purposes if the accuracy requirements are not too high (cf. [7.4.1]). The observation equations are given with (7.35). After deactivation of Selective Availability (SA) the accuracy level is about 10 m. This is sufficient for most purposes of general navigation, but not for particular tasks in land navigation, in marine geodesy and certainly not in hydrography (cf. [7.6.2.7]). It is therefore not very meaningful to use carrier-smoothed code observations (2) for a single receiver, because a series of observations (relative in time) is affected by time variable effects and also by change in the satellite constellation. Carrier phases can, however, be used for the determination of the instantaneous velocity. Using (6.4)

$$f_r(t) = f_s \left(1 - \frac{1}{c} \frac{dr}{dt} \right),$$

with f_s the frequency emitted from the satellite and $f_r(t)$ the Doppler-shifted frequency, measured in the receiver, it follows that

$$\frac{ds}{dt} = c \cdot \left(1 + \frac{f_r(t)}{f_s} \right). \tag{7.107}$$

In all navigational applications with accuracy requirements better than, say, 10 m, relative observations are essential, either in mode (3) or in mode (4). In both cases, simultaneous observations at a fixed reference station with known coordinates are required (cf. Fig. 7.60, p. 326), [12.3.2].

From the reference observations differences are computed, either between the actual position and the known position or between the actually observed pseudoranges and the ranges derived from the satellite coordinates and the known station coordinates. The differences are transmitted to the moving platform and are used as corrections to the navigation solution. The procedure is known as *Differential GPS* (DGPS) and is treated at length in [7.5.1]. Fig. 7.46 gives a schematic view of a generic navigational software package for differential navigation.
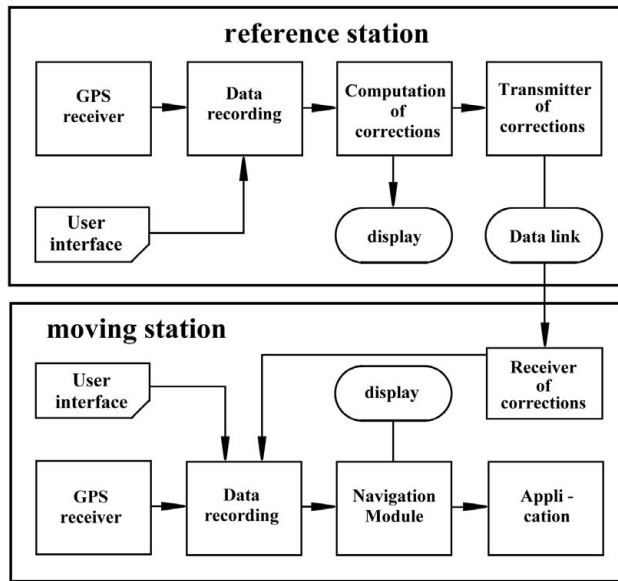
Figure 7.46. Schematic view of a software package for differential navigation

Modern high performance algorithms for shipborne positioning are nearly always based on a combination of code and carrier phase measurements (mode (4)). The carrier phase observations are considered as time-differenced pseudoranges with a much higher accuracy level than the pseudoranges from code measurements. A combination of both observables with proper weighting yields a smoothed series of pseudoranges (Fig. 7.47).

The recursive algorithm was proposed by *Hatch* as early as 1982 and is as follows (Lachapelle et al., 1986; Lachapelle, 1991), see also Westrop et al. (1990):
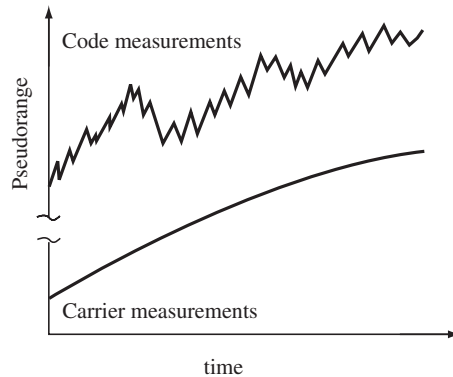
Figure 7.47. Carrier smoothed pseudoranges

$$PR_\Phi(t) = w_{PR}(t)PR_r(t) + w_\Phi(t)(PR_\Phi(t-1) + (\Phi(t) - \Phi(t-1))), \quad (7.108)$$

where

$PR_r(t)$ raw observed pseudorange at t,
$PR_\Phi(t)$ phase-smoothed pseudorange at t,
$w_{PR}(t)$ weight of raw pseudorange at t,
$w_\Phi(t)$ weight of $(PR_\Phi(t-1) + (\Phi(t) - \Phi(t-1)))$,

$$w_\Phi(t) = 1 - w_{PR}(t). \tag{7.109}$$

The use of this filter progressively increases the weight of the phase measurements, Φ. In practice, the ambiguity is determined after some time with a resolution of three to five cycles for C/A-code receivers, corresponding to an accuracy level of about 1 meter. Cycle slips may cause discontinuities if less than three satellites remain on lock. In such cases the algorithm has to be reset. Positions and velocities are usually estimated in a Kalman filter approach (e.g. Gelb (ed.), 1974; Schwarz, 1991).

In essence, the code pseudoranges are used for a rough estimation of the position, and the relative carrier phase measurements provide precise position change estimates. These position changes are used to refer all position estimates to one epoch for averaging. The technique works particularly well now after the deactivation of SA. A combination of precise IGS orbits and carrier smoothed pseudoranges provides high precision navigation results with a single receiver (Bisnath et al., 2002).

The mode (5) yields the highest accuracy and is identical to the pure kinematic method [7.3.5.4]. Navigation with GPS is treated in more detail in the sections [7.5.1], [7.6.2.7] and [12.3]. For an overview on the achievable accuracy see Fig. (7.61, p. 327).

7.4   Error Budget and Corrections

7.4.1   Basic Considerations

From a general point of view, errors are introduced into the process of parameter estimation if the modeling is too simple and does not conform to physical reality. The simple concept of pseudorange measurement and navigation in Fig. 7.2 and Fig. 7.29 fails to consider some physical circumstances, or does so insufficiently, namely:

- the Earth-fixed geocentric reference system (CTS) is not an inertial system,
- Newtonian mechanics are not strictly applicable, and
- the signals are not propagating in a vacuum.

It is therefore necessary to correct the satellite coordinates, the satellite clocks, and the observations for

- Earth rotation,
- relativistic effects, and
- ionospheric and tropospheric propagation effects.

In addition, corrections may be necessary for imperfect orbit modeling, signal propagation delays inside the satellite and the receiver hardware, and multipath effects. Finally, the error propagation is affected by the geometric distribution of the satellites. Since most influences that affect the accuracy of GPS measurements have already been treated in full earlier in this book, a summary presentation will be sufficient for this section.

Corrections to the observations are based on measurements or model assumptions. If corrections are applied, we are left with a certain residual error budget (cf. Table 7.12). The error budget can be reduced by refined modeling and by additional observations.

Usually, the contribution of a particular error source is analyzed in terms of its effect on the range determination. The combined effect of ephemeris uncertainties, propagation errors, clock and timing errors, and receiver noise, projected onto the line connecting observer and satellite, is called *User Equivalent Range Error* (UERE) or *User Range Error* (URE). Sometimes the total error is divided into *Signal-in-Space* (SIS) URE, also abbreviated as SISRE, and the *User Equipment Error* (UEE). The rationale behind this division is that the Operational Control Segment (OCS) is only responsible for the SIS performance whereas the UEE depends on the particular user's equipment and correction models.

The SIS URE includes satellite clock and ephemeris prediction errors, OCS state estimate process noise, and some minor residual noise. SIS does not include instantaneous single-frequency ionospheric model errors, tropospheric model errors, receiver noise, receiver antenna phase center variations, or multipath effects. These influences contribute to the UEE.

Official statements can be taken from the document "GPS SPS Performance Standard" (DOD, 2001). This document contains the specific capabilities provided by the *Standard Positioning Service* SPS to all users on a continuous, worldwide basis without any direct user charge. Following the "Federal Radio Navigation Plan" (DOD/DOT, 2001a), access to the *Precise Positioning Service*, PPS, is restricted to U.S. Armed Forces, U.S. Federal agencies and selected allied armed forces and governments (see also [7.1.6].

With disabled *Selective Availability* (SA) the SIS performance for PPS and SPS is nearly identical. According to the cited document the accuracy standard for the SPS Signal-in-Space URE is

$$\sigma_{\text{URE}} \approx 6 \text{ m}.$$

The related accuracy standards for position and height are for a global average positioning domain (95%, SIS only):

$$\leq \quad 13 \text{ m} \quad \text{horizontal, and}$$
$$\leq \quad 22 \text{ m} \quad \text{vertical.}$$

For a worst site positioning domain the numbers are (95%, SIS only):

$$\leq \quad 36 \text{ m} \quad \text{horizontal, and}$$
$$\leq \quad 77 \text{ m} \quad \text{vertical.}$$

Experiences show that in practice the achievable accuracy is much higher (cf. [7.1.6]). The particular error sources are assigned to three main groups, namely
   − satellite position and clock errors,
   − signal propagation errors, and
   − receiver errors.

Table 7.12 includes average numerical values of the individual error sources as they are generally accepted for operational GPS.

Table 7.12. Main GPS error contributions to the single range observation

Error Source	RMS Range Error
Satellite	
– orbit	1 – 2 m
– clock	1 – 2 m
Signal propagation	
– ionosphere (2 frequencies)	cm – dm
– ionosphere (model, best)	1 – 2 m
– ionosphere (model, average)	5 – 10 m
– ionosphere (model, worst)	10 – 50 m
– troposphere (model)	dm
– multipath	1 – 2 m
Receiver	
– observation noise	0.2 – 1 m
– hardware delays	dm – m
– antenna phase center	mm – cm

Another separation is into
  – distance dependent errors (orbit, ionosphere, troposphere), and
  – station dependent errors (antenna phase center variation, multipath).
This latter grouping is used together with the error modeling in multiple reference station networks [7.5.3].

The *Earth rotation correction* is necessary if satellite coordinates are computed in an Earth-fixed reference frame at the epoch of signal transmission. During signal propagation from the satellite antenna to the receiver antenna the CTS coordinate system rotates with respect to the satellite; consequently the position of the transmission antenna changes in the rotated CTS system. The original satellite coordinates must be rotated about the Z-axis by an angle, α, which is defined as the product of the propagation time, τ, and Earth's rotational velocity, ω_e (cf. [7.1.5.3]):

$$\alpha = \omega_e \tau. \tag{7.110}$$

Let X', Y', Z' be the original, and X, Y, Z the corrected satellite coordinates, then

$$X = X' \cos \alpha + Y' \sin \alpha, \quad Y = Y' \cos \alpha - X' \sin \alpha, \quad Z = Z'. \tag{7.111}$$

The rotation angle, α, is smaller than $1.''5$. Hence, the trigonometric functions in (7.111) can be replaced by the first elements of a series expansion.

A correction for *relativistic effects* is required because the satellite clock and the main clock by which GPS system time is defined operate at places with different gravitational potential and are moving with different velocities. The relativistic effect causes an apparent frequency shift in the satellite oscillator. The main part of this effect is compensated because the satellite oscillator (10.23 MHz) is operated at a slightly reduced nominal frequency (0.0045 Hz less), see Van Dierendonck et al. (1980). What remains is a small constant component, due to different orbital heights, and a periodic component. The constant effect is absorbed by the satellite clock's drift parameter, a_1, cf. (7.4). Due to the periodic effect the satellite clock reading must be corrected (cf. [7.1.5.3]):

$$\Delta t_r[s] = -4.443 \cdot 10^{-10} e\sqrt{A}[m]\sin E, \qquad (7.112)$$

and the drift of the satellite clock:

$$\dot{\Delta} t_r[s] = -4.443 \cdot 10^{-10}\sqrt{A}[m]\cos E \frac{dE}{dt}. \qquad (7.113)$$

The effect of the correction (7.112) on the satellite time, for one revolution, is demonstrated in Fig. 7.48. The maximum value can reach 70 nanoseconds in time, and 0.01 nanoseconds/sec for the clock drift. Remaining relativistic effects are compensated in relative observation. For a deeper treatment of the subject see e.g. Ashby, Spilker (1996).
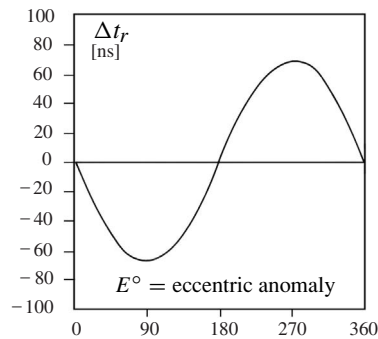
Figure 7.48.   Relativistic correction for one satellite revolution

7.4.2   Satellite Geometry and Accuracy Measures

The accuracy of GPS positioning depends on two factors:
- the accuracy of a single pseudorange measurement, expressed by the User Equivalent Range Error (UERE) [7.4.1] or by the associated standard deviation, σ_r, and
- the geometric configuration of the satellites used.

The relation between σ_r and the associated standard deviation of positioning, σ^*, is described by a scalar quantity which is frequently used in navigation and called DOP (*Dilution of Precision*):

$$\sigma^* = \text{DOP} \cdot \sigma_r. \qquad (7.114)$$

Different DOP designations are in use:

$$
\begin{aligned}
\sigma_H &= \text{HDOP} \cdot \sigma_r && \text{for horizontal positioning,} \\
\sigma_V &= \text{VDOP} \cdot \sigma_r && \text{for vertical positioning,} \\
\sigma_P &= \text{PDOP} \cdot \sigma_r && \text{for 3D positioning, and} \\
\sigma_T &= \text{TDOP} \cdot \sigma_r && \text{for time determination.}
\end{aligned}
\tag{7.115}
$$

The combined effect for position and time is called GDOP:

$$
\text{GDOP} = \sqrt{(\text{PDOP})^2 + (\text{TDOP})^2}.
\tag{7.116}
$$

PDOP can be interpreted as the reciprocal value of the volume, V, of a tetrahedron that is formed from the satellite and user positions (Milliken, Zoller, 1980):

$$
\text{PDOP} = \frac{1}{\text{V}}.
\tag{7.117}
$$

Fig. 7.49 gives a geometrical explanation. The best geometric situation exists when the volume is maximized, and hence PDOP in (7.117) is minimized.
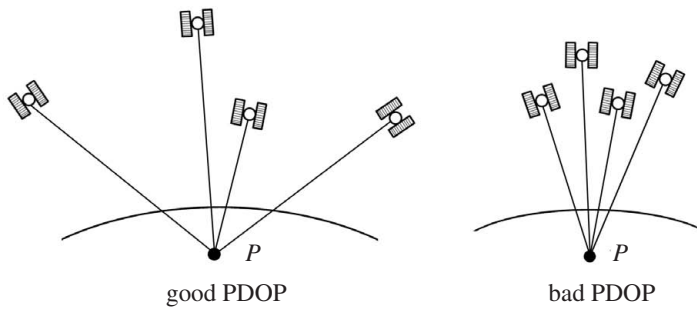
Figure 7.49. Satellite geometry and PDOP

The same result can be derived from the rules of adjustment and error propagation, e.g. Leick (1995); Misra, Enge (2001). Following the equations (7.86) to (7.91) we obtain the covariance matrix for GPS positioning:

$$
\boldsymbol{C}_{xx} = \sigma_r^2 (\boldsymbol{A}^\mathrm{T} \boldsymbol{A})^{-1},
\tag{7.118}
$$

with the following elements

$$
\boldsymbol{C}_{xx} = \sigma_r^2
\begin{pmatrix}
q_{xx} & q_{xy} & q_{xz} & q_{xt} \\
q_{yx} & q_{yy} & q_{yz} & q_{yt} \\
q_{zx} & q_{zy} & q_{zz} & q_{zt} \\
q_{tx} & q_{ty} & q_{tz} & q_{tt}
\end{pmatrix}.
\tag{7.119}
$$

From the covariance matrix, the variance of a position determination is found to be

$$\sigma_p^2 = \sigma_r^2 \, (q_{xx} + q_{yy} + q_{zz}), \qquad (7.120)$$

or

$$\sigma_P = \sigma_r \cdot \text{PDOP}. \qquad (7.121)$$

If we introduce a local Cartesian coordinate system (cf. Fig. 2.8) in which the x-axis points north, the y-axis east, and the z-axis to the zenith, it then follows for the horizontal position error, that

$$\sigma_H^2 = \sigma_r^2 \, (q_{xx} + q_{yy}), \qquad (7.122)$$

or

$$\sigma_H = \sigma_r \cdot \text{HDOP};$$

and for the vertical position error, that

$$\sigma_V^2 = \sigma_r^2 \, q_{zz}, \qquad (7.123)$$

or

$$\sigma_V = \sigma_r \cdot \text{VDOP}.$$

The fourth parameter t contains primarily the timing error. Hence

$$\sigma_t^2 = \sigma_r^2 \, q_{tt}, \qquad (7.124)$$

or

$$\sigma_t = \sigma_r \cdot \text{TDOP}.$$

In general

$$\text{GDOP} = \sqrt{\text{trace}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}}. \qquad (7.125)$$

Until the full satellite configuration was installed the DOP values for a given location changed considerably during the day, and the best "*observation window*" had to be selected. Since the operational 24-satellite configuration is complete, the variations of DOP values are less critical, and PDOP remains at most times below 3, or even 2.

Today, with the full constellation, the DOP numbers are in general of less interest to surveying applications, because all visible satellites can be tracked with modern receivers, and can be introduced into a rigorous adjustment process. Accuracy estimates result from the adjustment algorithm rather than from PDOP calculations. DOP indicators remain, however, an important tool for survey planning and control, in particular in the rapid static, kinematic or navigational mode [7.3.5], where short time signal blocking caused by obstructions may occur. For a detailed treatment of the DOP-issue see e.g. Spilker (1996c)

The accuracy measure usually applied in surveying and geodesy is the *standard deviation*, σ, mostly considered to be identical with the *root mean square error* RMS.

The probability of a location being within a certain region is described by a *confidence ellipse* (for two dimensions), or a *confidence ellipsoid* (for three dimensions) with the estimated position at its center. The axes of the confidence ellipse are a function of the standard deviation of the particular coordinates (e.g. σ_Φ for latitude and σ_Λ for longitude), and the level of probability. Usually applied levels of probability are 68.3 % (corresponding to 1 σ), 95.5 % (corresponding to 2 σ), or 99.7 % (corresponding to 3 σ). Accuracy numbers throughout this book are usually at the 1 σ level. For more information, see textbooks on adjustment and statistics e.g. Leick (1995); Strang, Borre (1997) or Niemeier (2002).

The accuracy measures usually applied in navigation are quite different. In most cases they are based on the accuracy of so-called *lines of position* LOP in a plane, or "surfaces of position" in three-dimensional navigation. Every navigation system defines its proper system of LOPs. The user position is located at the intersection of two or more such lines or surfaces of position. Some general information is given as follows, without going into details. For further reading see e.g. Forsell (1991) .

The relation between geodetic and navigational definitions is illustrated in Fig. 7.50. A common way to express two-dimensional accuracy is the *Distance Root Mean Square* (DRMS):

$$\text{DRMS} = \sqrt{\sigma_\varphi^2 + \sigma_\lambda^2}. \qquad (7.126)$$

The probability of being within a circle with radius DRMS varies between 63.2 % and 68.3 %. Alternative names for DRMS are *Circular Radial Error* or *Mean Squared Position Error* (MSPE). One pa-

Figure 7.50. Geodetic and navigational accuracy measures

rameter which is frequently used (for example in the *U.S. Federal Radio Navigation Plan* and related documents (DOD, 2001; DOD/DOT, 2001a,b)) is the 2 DRMS:

$$2\,\text{DRMS} = 2 \times \text{DRMS} = 2\sqrt{\sigma_\varphi^2 + \sigma_\lambda^2}. \qquad (7.127)$$

The probability level is between 95.4 % and 98.2 %. Note that 2 DRMS must not be confused with 2-D RMS, the *two-dimensional root mean squared error*, that is essentially identical with DRMS (7.126).

The *Circular Error Probable* (CEP) is also widely used for different levels of probability. The most used measure is:

$$\text{CEP} = 0.59\,(\sigma_\varphi + \sigma_\lambda), \qquad (7.128)$$

for 50 % probability. It defines the radius of a circle, centered at the true position, containing 50 % of the estimated positions. Other measures are:

$$\text{CEP}_{95} = \text{CEP} \cdot 2.08, \qquad (7.129)$$

for 95 % probability, and

$$\text{CEP}_{99} = \text{CEP} \cdot 2.58, \tag{7.130}$$

for 99 % probability.

Accuracy measures in use for three dimensions are the *Mean Radial Spherical Error* (MRSE):

$$\text{MRSE} = \sqrt{\sigma_\varphi^2 + \sigma_\lambda^2 + \sigma_h^2}, \tag{7.131}$$

with a probability of 61 %, and the *Spherical Error Probable* (SEP):

$$\text{SEP} = 0.51(\sigma_\varphi + \sigma_\lambda + \sigma_h), \tag{7.132}$$

with a probability of 50 %. The relationship between 2 DRMS, CEP, and SEP is

$$2\,\text{DRMS} = 2.4 \cdot \text{CEP} = 1.18 \cdot \text{SEP}. \tag{7.133}$$

It becomes evident that numbers, indicating accuracies, are only meaningful if the corresponding accuracy measure is identified. As an example, the achievable accuracies under SPS are

$$
\begin{array}{llll}
\approx 13\text{ m} & (\varphi \text{ and } \lambda) & 2\,\text{DRMS} & (95\,\%), \\
\approx 5\text{ m} & (\varphi \text{ and } \lambda) & \text{CEP} & (50\,\%), \text{ and} \\
\approx 11\text{ m} & (\varphi, \lambda \text{ and } h) & \text{SEP} & (50\,\%).
\end{array}
$$

Note that all three accuracy numbers describe an identical situation. A short overview of accuracy measures is given by Van Diggelen (1998).

7.4.3   Orbits and Clocks

7.4.3.1   Broadcast Ephemerides and Clocks

Discrepancies between the predicted ephemeris available to the user and the actual orbit propagate into the determined positions of the user antenna. It is evident (Fig. 7.51) that the radial component of the orbital error corrupts the range determination, and hence the location of the user position, to a much higher degree in single station positioning than in relative positioning. For nearby stations, most of the orbit errors are cancelled out in the differencing and in the determination of relative coordinates. As a *rule of thumb* we have for the effect, db, of the orbit error, dr, on the determination of the baseline, b:

$$\frac{db}{b} = \frac{dr}{\rho}. \tag{7.134}$$
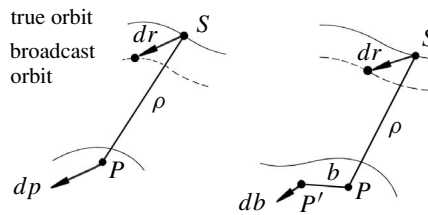
Figure 7.51.   Effect of an orbit error on the single-point position (left) and the baseline determination (right)

The baseline error, db, thus depends mainly on the ratio of the baseline length, b, to the satellite range, ρ.

The maximum range between a GPS satellite and the observer is about 25 000 km. If a baseline error of 1 cm is accepted, an admissible orbit error for a specified baseline length is summarized in Table 7.13. The table clearly shows that, for relative coordinate determination over short distances, the required orbit accuracy is not a critical factor. On the other hand, the requirement for 1 cm accuracy over very great distances, for example, in geodynamic applications over 1000 km and more, implies an orbit accuracy of better than 1 m which is not yet provided by the broadcast ephemerides.

Table 7.13. Relation between orbit errors and corresponding 1 cm baseline errors

Baseline length	Admissible orbit error
0.1 km	2500 m
1.0 km	250 m
10 km	25 m
100 km	2.5 m
1000 km	0.25 m

In many cases, the accuracy of the baseline determination is set as a ratio of the base length, estimated in parts per million (ppm). To give an example, an orbit error of 25 m propagates into a relative accuracy of 1 ppm, and an orbit error of 5 m corresponds to 0.2 ppm, i.e. 1 cm over 50 km. The last figure is the critical limit for reference station networks (see [7.5.3]).

The formula (7.134) implies a considerable approximation, and it is widely regarded as too pessimistic (e.g. Zielinski, 1989; Beutler et al., 1998, p. 104). Indeed, the equation is derived from geometrical considerations and only reflects an instantaneous situation. For a whole session, the changing geometric configuration for all satellites has to be included. The resulting error evolves as the sum of all individual satellite orbit errors, integrated over the whole observation period. Zielinski (1989) estimates the resulting baseline errors as too large by a factor of 4 to 10. The "rule of thumb" (7.134) should therefore only be used with care, and for a rough estimation. More detailed equations are given in the cited literature.

Because of the great altitude of the GPS satellites, their orbits are only slightly affected by surface forces and higher-order potential coefficients of Earth's gravity field [3.2.4]. For the computation of short orbit arcs a gravity field expansion up to degree and order (6,6) or (8,8) is sufficient. It is necessary to include the gravitation effects of the Sun and Moon, as well as the Sun's radiation pressure, in orbit computation models. In particular, the non-gravitational forces on GPS satellite orbits have to be modeled carefully (e.g. Fliegel, Gallini, 1989; Fliegel et al., 1992; Beutler et al., 1998; Ziebart et al., 2002), cf. [3.2.3.4]. The ROCK42 model is used for Block II/IIa satellites, and a particular new model for Block IIR satellites (Marquis, Krier, 2000). For precise

computations a model for thermal re-radiation and a particular y-bias may be included (e.g. Rothacher, Mervart, 1996).

Experiences with the accuracy of broadcast orbits (status 1998/2000) indicate a level of 5 to 10 meters (Jefferson, Bar-Sever, 2000). However, there are times where significantly larger orbit errors may occur. Broadcast orbits in 2002 are accurate to about 3 m, following the estimation of the IGS (cf. Tab. 7.31, [7.8.1], p. 402). Along with the Accuracy Improvement Initiative (AII) [7.1.7] an orbit accuracy of about 1 m can be expected.

According to Tab. 7.13, this accuracy is in general sufficient for work with point distances up to several tens of kilometers. When greater point distances are associated with high accuracy requirements, the accuracy of the operational broadcast ephemerides is usually not sufficient. Hence, orbit improvement techniques have to be applied along with the data adjustment in large scale networks [7.3.4], or a posteriori precise ephemerides, based on observations from globally distributed *tracking stations* have to be used [7.4.3.2].

GPS time is operationally defined as the time scale used by the GPS system (cf. [2.2.3]). Each GPS satellite carries clocks which act as the time and frequency base for the realization of the *GPS system time* in the particular satellite. Navigation signals and carrier signals are time-tagged to the particular *satellite time frame*. GPS time is monitored by the Main Control Station (MCS) [7.1.3]. Its relationship to the other atomic time scales is demonstrated in Fig. 2.14, cf. [2.2.3, p. 38].

The satellite clocks run fast by 38.5 μs/day; this correction absorbs more than 99.6% of the relativistic clock effect. If necessary, the MCS applies other corrections to synchronize the individual satellite clocks with the system time. A synchronization error of 1 μs in a satellite clock will produce an error of 300 m in the pseudorange. When meter-level position accuracy is required, the clock synchronization between the satellites must be controlled within a few nanoseconds. This is why rubidium and cesium oscillators are used in the GPS satellites [2.2.5]. These clocks have a short-term stability of 10^{-9} to 10^{-10}, and a long-term stability of 10^{-12} to 10^{-13}. Block II/IIa satellites carry two rubidium and two cesium atomic frequency standards, and each Block IIR carries three rubidium standards.

The performance of each clock is observed by the Control Segment [7.1.3], and one of the clocks is selected to generate the signals. The deviation of a particular clock from GPS system time is modeled as a quadratic function of time. The parameters of this model are estimated, uploaded to the satellite, and are broadcast within the navigation message [7.1.5.3]. The coefficients a_0, a_1 and a_2 in equation (7.4) are also called the bias, drift, and aging parameters of the clock. Typically, parameter a_0 is below 1 μs; $a_1 \approx 10^{-11}$ s/s, and $a_2 \approx 0$ s/s^2. With these correction terms of the broadcast message the satellite clocks can be kept synchronized within 5 to 10 ns (Misra, Enge, 2001, p. 92).

The actual behavior of each clock slightly differs from this model because of unpredictable, correlated frequency errors. For the highest accuracy requirements, the satellite clock parameters can be estimated in the adjustment process. A stochastic

correlation model (cf. [2.2.5], Wübbena (1989)) can be included for the growth of the random frequency error with time. Alternatively, a posteriori clock models based on observations can be used [7.4.3.2].

The requirements of the *receiver clock* are not very high. The user clock in the receiver need only be stable enough to do the pseudorange measurements with code-phases. A quartz oscillator of medium quality usually suffices. In most geodetic adjustment models, receiver clock errors are eliminated by means of double differences of the carrier phase observations [7.3.2.1].

However, the use of a more precise external clock (e.g. a rubidium oscillator) is of importance in cases where only few satellites are available. The clock substitutes one satellite.

7.4.3.2 Precise Ephemerides and Clocks, IGS

A posteriori precise ephemerides (PE) and clock parameters are based on observations from globally distributed *tracking stations*. At such stations, dual-frequency receivers are installed that can measure both the code phases and carrier phases of all satellites in view. Orbit errors can be separated from the station clocks' time errors through the use of high-precision oscillators (rubidium, cesium atomic standard). The tropospheric propagation delay can be determined with water vapor radiometers.

The data files usually conform to the *SP3 data format* (Standard Product 3), finalized by the U.S. National Geodetic Survey (NGS), (Remondi, 1991; Hilla, 2002). This format is precise to 1 mm and 1 picosecond. Several agencies provide precise ephemerides and adjusted clock parameters, among them the NIMA, JPL, and IGS.

The *National Imagery and Mapping Agency* NIMA (formerly *Defense Mapping Agency* (DMA)) generates precise ephemeris (PE) data files and improved clock parameters based on observations from 20 monitor stations. These are twelve NIMA and five Air Force stations (see Fig. 7.8, p. 217), and three IGS stations (Maspalomas, Kerguelen, Yakutsk).

The ephemerides files give position and velocity vectors for each satellite every 15 minutes. Two PE data types are available, one referred to the satellite's center of mass, and the other with respect to the satellite's antenna phase center. A comparison in 2001 between the IGS final orbits (see below) and the NIMA precise ephemerides showed differences of less than 20 cm.

Precise orbits and adjusted clock parameters can be used either for the post-processing of data in multi-station GPS networks; or they can be used for the processing of single receiver data in the *Precise Point Positioning* (PPP) mode [7.3.4]. The NIMA precise ephemerides are freely available via Internet (anonymous ftp).

Another resource for precise orbits are the NASA JPL (Jet Propulsion Laboratory) precise ephemerides and adjusted clock parameters. Positions and velocities are given for every 15 minutes, and clock parameters for every 5 minutes. Orbits and clocks result from the same estimation process and are hence completely consistent with each other. Final orbits are available after about 2 weeks. Rapid orbits are given within 20 hours; they agree with the final orbits at the level of about 20 cm. The JPL orbits

are also given as *non-fiducial* (NF) orbits, i.e the orbits are estimated in a free datum independent from the ITRF coordinates of the tracking stations. The datum instead is derived from the orbits and clocks. The adjusted network can then be transformed to any other datum without problems. Precise JPL orbits and clocks are primarily required for the processing of single receiver GPS data in the *Precise Point Positioning* (PPP) mode with the software package GIPSY OASIS II [7.3.4]. The necessary accuracy of the adjusted clocks is in the order of 100 picoseconds.

The most important source for precise ephemerides and other GPS products today is the IGS. The IGS, a service established by the IAG, officially started its activities on January 1, 1994, after a successful pilot phase of more than one year. In 1999 the name was changed from *International GPS Service for Geodynamics* to *International GPS Service*. Following the *Terms of Reference*, the primary objective of the IGS is to "provide a service to support through GPS data and GPS products geodetical and geophysical research activities". For more details of the structure, organization, and the various and growing services of the IGS see [7.8.1]. In the following, only the main information on data, orbits and clocks is given.

IGS collects, archives and distributes GPS observation data sets from more than 300 globally distributed stations. The stations have to meet certain quality criteria. About 120 stations are classified as *Global Stations* because they are regularly analyzed by at least three Analysis Centers. The IGS core products consist of weekly final products, namely

- GPS ephemeris and clock values, tabulated at 15-minutes intervals for each day (in SP3 format),
- Earth Orientation Parameters (EOP), and
- Geocentric station coordinates and velocities.

With respect to orbits and clocks, three different products are available (see Table 7.14, status August 2002). The *ultra-rapid orbits (predicted orbits)* are updated twice daily (at 03.00 and 15.00 UT) and are valid for a period of 48 hours. The first 27 hours are based on actual observations and the second 21 hours are a predicted orbit.

Table 7.14. Precise IGS GPS orbits and clocks

Orbits	Accuracy	Latency	Updates	Sample Interval
Broadcast	$\sim$ 260 cm/ $\sim$ 7 ns	real time	–	daily
Ultra-Rapid	$\sim$ 25 cm/$\sim$5 ns	real time	twice daily	15 min/15 min
Rapid	5 cm/0.2 ns	17 hours	daily	15 min/5 min
Final	< 5 cm/0.1 ns	$\sim$ 13 days	weekly	15 min/5 min

With the IGS products at hand, all requirements for precise orbits are completely fulfilled. Together with the precise station coordinates and the original observation data from IGS stations, it is possible to connect every new station worldwide directly to the geocentric reference frame (see [7.6.2.1]). Note that also the individual analysis

centers of the IGS provide precise orbits, for example the NGS in the U.S. and CODE in Europe (see [7.8.1]).

7.4.4   Signal Propagation

The GPS signals, when propagating from the satellite antenna to the user antenna are subject to the following propagation effects:
- − propagation delay in the ionosphere,
- − propagation delay in the troposphere, and
- − multipath propagation at the satellite and in the vicinity of the receiver antenna.

The atmospheric propagation delays are basically treated in section [2.3]. In this chapter, some of the more important properties and relationships with respect to GPS are pointed out.

7.4.4.1   Ionospheric Effects on GPS Signals

The propagation delay in the ionosphere (between about 50 km and 1000 km above the Earth's surface) depends on the electron content along the signal path and on the frequency used. The influencing parameters are mainly solar activity and the geomagnetic field. Hence, ionospheric refraction varies with frequency, geographic location, and time. The resulting range error, for GPS frequencies, can vary from less than 1 m to more than 100 m (Wells (ed.), 1986; Klobuchar, 1991, 1996). Dual frequency receivers make use of the fact that the L1 and L2 signals experience different propagation delays in the ionosphere. In addition, we have to note that the ionosphere is a dispersive medium [2.3.1.2], and that therefore the phase velocity (propagation of the carrier) is not the same as the group velocity (propagation of the codes).

To be exact, we observe the combined effect from the ionosphere and the *plasmasphere*, because the GPS orbits are located far above the ionospheric layers. The electron content below about 2000 km is also called the *Faraday content*. For a detailed study of the time-variable ionospheric behavior, e.g. in atmospheric physics for *ionospheric tomography*, it is hence advisable to combine measurements from Low Earth Orbiters (LEO) and GPS satellites, or to install GPS receivers in satellites at low orbital height. An observer at the surface of the Earth, who uses GPS as a tool for positioning or navigation, has no need to separate the ionospheric and the plasmaspheric propagation delay. In this book, as in most literature, the term *ionospheric delay* is therefore understood as the combined effect. For more information on the physics of the ionosphere with particular reference to the propagation of radio waves see e.g. Davies (1990) or Klobuchar (1996).

For carrier phase measurements we have the refraction coefficient from equation (2.95):
$$n_p = 1 - 40.3 \, \frac{n_e}{f^2},$$
with
  n_e electron content along the signal propagation path, and

f carrier frequency.

The ionospheric effect on code propagation (group delay) is, at first order, of the same size as the carrier phase propagation but has the opposite sign (2.99):

$$n_g = 1 + 40.3 \frac{n_e}{f^2}.$$

Integration over the entire propagation path, s, then yields the total effect of ionospheric refraction on the pseudorange measurement, R, with code phases:

$$\delta R_{\text{ION}_g} = \int_s (n_g - 1) \, ds \tag{7.135}$$

$$\delta R_{\text{ION}_g} \approx \frac{40.3}{f^2} \int_s n_e \, ds. \tag{7.136}$$

The corresponding expression for carrier phase measurements is

$$\delta R_{\text{ION}_p} \approx -\frac{40.3}{f^2} \int_s n_e \, ds. \tag{7.137}$$

Hence, the range from a code phase observation is measured as too long, and a range from a carrier phase observation is measured as too short.

The unknown integral can be determined by measurements of the ranges $R_1 = R$ (L1) and $R_2 = R$ (L2) on both frequencies:

$$R = R_1 - \delta R_{1,\text{ION}}, \quad R = R_2 - \delta R_{2,\text{ION}}. \tag{7.138}$$

By substitution of (7.136) into (7.138), omitting subscript g or p for simplicity, it follows that the expression of range correction for code phase measurements on L1, derived from dual frequency observations is

$$\delta R_{1,\text{ION}} = \frac{R_1 - R_2}{1 - \left(\frac{f_1^2}{f_2^2}\right)} = \frac{f_2^2}{f_1^2 - f_2^2} (R_2 - R_1). \tag{7.139}$$

Because of the approximation in (2.99), this equation (7.139) is called the *first-order ionospheric refraction correction*. The remaining model errors reach only a few centimeters (see Table 7.15). Therefore the ionospheric effect in GPS can be very largely modeled by dual frequency observations. Table 7.15 shows the maximum range errors that can be expected for both GPS frequencies, and for the dual frequency corrected signal, both in vertical direction. For inclined directions the influence increases with the appropriate mapping function (cf. [2.3.2]). For more detailed information see e.g. Wanninger (1994); Langley (2000a); Misra, Enge (2001).

Table 7.15. Maximum vertical ionospheric range error [m] (Wübbena, 1991)

Frequency	1st order effect $(1/f^2)$	2nd order effect $(1/f^3)$	3rd order effect $(1/f^4)$
L1	32.5	0.036	0.002
L2	53.5	0.076	0.007
L1/L2	0.0	0.026	0.006

Corresponding equations can be derived for carrier phase observations (e.g. Wells (ed.), 1986; Misra, Enge, 2001):

$$\delta\Phi_{\text{ION}}(\text{L1}) = \frac{f_2^2}{f_2^2 - f_1^2}\left(\Phi(\text{L1}) - N(\text{L1}) - \frac{f_1}{f_2}(\Phi(\text{L2}) - N(\text{L2}))\right), \qquad (7.140)$$

where $N(\text{L1})$ and $N(\text{L2})$ are the respective ambiguity terms. Equation (7.140) describes the *ionospheric phase advance* for L1 observations. Combining observations on L1 and L2 yields the *ionospheric free linear combination* (e.g. Leick, 1995; Hofmann-Wellenhof et al., 2001):

$$\Phi(\text{L}_0) = \frac{f_1^2}{f_1^2 - f_2^2}\Phi(\text{L1}) - \frac{f_1 f_2}{f_1^2 - f_2^2}\Phi(\text{L2}). \qquad (7.141)$$

Some authors call the ionospheric free combination the L_3 observable. The corresponding equation for ionospheric-free code range observations can be reached from (7.139) as (e.g. Misra, Enge, 2001):

$$R_0 = \frac{f_1^2}{(f_1^2 - f_2^2)}R_1 - \frac{f_2^2}{(f_1^2 - f_2^2)}R_2 = 2.546\,R_1 - 1.546\,R_2. \qquad (7.142)$$

If only single-frequency receivers are available, a correction according to equations (7.139) or (7.140) is impossible. An attempt can then be made to use an ionospheric correction model, the coefficients of which are transmitted as part of the GPS satellite message (Klobuchar, 1987, 1996), [7.1.5.4]. This *Klobuchar model* is described by eight coefficients α_n, β_n, and removes about 50% of the ionospheric delay at mid-latitudes. The correction formula is

$$\begin{aligned}\Delta T_{\text{ION}} &= DC + A\cos(2\pi(t - \Phi)/P) \quad [\text{day}]\\\Delta T_{\text{ION}} &= DC \quad [\text{night}],\end{aligned} \qquad (7.143)$$

in which

ΔT_{ION} vertical delay (ns),
DC constant night-day offset (5 ns),
A amplitude of the cosine function for daytime values,
Φ constant phase offset corresponding to the peak of the cosine function, fixed at 54 000 s or 14 h local time,
t local time,
P period of the cosine function, and furthermore

$$A = \sum_{n=0}^{3} \alpha_n \Phi^n \,(\text{seconds}); \quad P = \sum_{n=0}^{3} \beta_n \Phi^n \,(\text{seconds}).$$

The vertical ionospheric delay has to be scaled into the slant delay with the slant factor, F:

$$F = 1.0 + 16.0 \times (0.53 - E)^3. \tag{7.144}$$

E is the satellite elevation angle. Details of the calculations are given in Klobuchar (1996); Misra, Enge (2001), and in the GPS Interface Control Document (ICD, 1993).

The Klobuchar model is based on empirical data, but has severe limitations because the number of parameters is restricted to eight, and because it can only be updated once daily. Due to the fast changing ionospheric environment the remaining error in zenith delay is estimated to be about 10 m during the day at mid-latitudes, and much worse when the solar activity is high (Misra, Enge, 2001).

Alternative approaches are to model the ionosphere with the help of LEO observations, as with TRANSIT [6.2] or with PRARE on ERS-2 [4.3.3.3] (Flechtner, 2000), or with reference observations from one or more dual-frequency GPS receivers located in the working area. Both procedures were initiated early on, e.g. Lohmar (1985) and Georgiadou, Kleusberg (1988), but have severe limitations. With LEO satellites, only the Faraday content below 2000 km can be determined but not the electron content of the plasmasphere above 2000 km altitude (Davies, 1990). The contribution of the plasmasphere reaches, however, 10% to 50% of the total electron content. TEC measurements with GPS, hence, cannot directly be compared with TEC results from other radio systems. On the other hand, GPS occultation measurements with a GPS receiver on a satellite in low orbit (LEO) help to map plasma irregularities in the lower ionosphere (Hocke, Tsuda, 2001), see [7.6.2.9].

In the second technique, the vertical propagation delay is observed with dual-frequency equipment, and is introduced into a polynomial model describing the local ionosphere for the correction of single frequency observations. With one reference receiver located at the center of the working area, the method works well under homogeneous and moderate ionospheric conditions. However, it is not very effective in regions and times with strong *ionospheric disturbances* and/or *very high electron content*. The technique has been further developed using three and more dual frequency receivers (e.g. Webster, Kleusberg, 1992), and it is now well established in active multiple-reference-station networks. The actual behavior of the ionosphere can be measured and modeled in real-time for the complete working area. This concept

of "wide area augmentation" or "area correction parameters" or "virtual reference stations" is treated at length in [7.5.3].

At a global scale, ionospheric TEC models are derived from data of the International GPS Service, IGS (see [7.8.1]). Five so-called "Ionospheric Analysis Centers" (IAACs) deliver every 24 hours an "Ionospheric Map Exchange" (IONEX) file (Schaer et al., 1998) with 12 maps containing global TEC information with 2-hour time resolution. For the northern hemisphere, under normal conditions, the different TEC maps agree with the IGS mean by about five TEC units or less. At the equator and for southern latitudes, the situation is still more problematic because of poor station coverage. However, the use of regional networks for monitoring regional TEC behavior is being investigated, e.g. for South America (Fedrizzi et al., 2001).

The IGS is preparing for the establishment of an independent IGS ionospheric model and a near-real-time service (IGS, 2002a).

Residual errors in the ionospheric modeling are cancelled out, for the most part, through relative observations at two stations over short distances, since the satellites are observed through nearly the same atmosphere. The remaining error for single frequency equipment is estimated to be 1 to 2 ppm of the interstation distance, corresponding to 1 to 2 cm over 10 km (Campbell et al., 1984). These numbers are valid for a quiet ionosphere, and for observations in mid-latitudes only. The last periods of high solar activity have demonstrated that the residual error can be significantly larger. It is hence advisable to use only dual frequency equipment for high precision application.

Irregularities in Earth's ionosphere can produce short-term signal variations in amplitude and phase (e.g. Wanninger, 1992, 1994; Langley, 2000a). These *scintillation effects* mainly occur in a belt of ± 30 degrees either side of Earth's geomagnetic equator, and in the polar auroral zones (see Fig. 7.52). A very high electron content only occurs in equatorial regions.
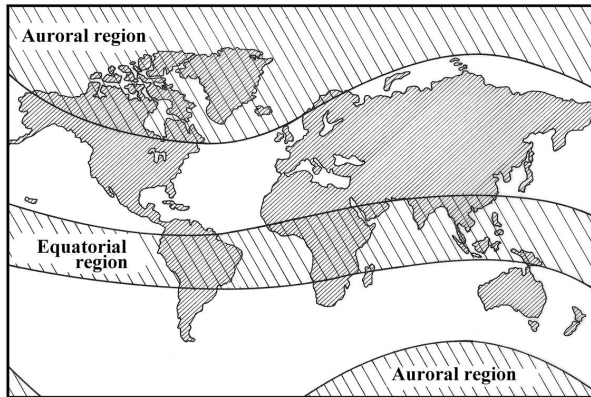
Figure 7.52. Regions of the world with high ionospheric activity

Equatorial scintillation effects have their maximum typically from approximately one hour after local sunset until approximately midnight (Klobuchar, 1991). Scintillation effects are less significant from April through August in the American, African, and Indian longitude regions, but maximize in the Pacific region. From September through March the situation is reversed.

Scintillation effects may cause a large number of cycle slips because the receiver cannot follow the short-term signal variations and fading periods. A very high electron content produces strong horizontal gradients and corrupts the ambiguity solution with the geometrical method, even over short distances, because the *ionospheric signal* (cf. [7.3.2.1]) overlaps even the wide lane wavelength within a few minutes of observation time (Wanninger, Jahn, 1991; Wanninger, 1994). In such situations, the only reliable possibility of ambiguity resolution so far found is the code-carrier combination [7.3.2.3] using data from low-code-noise receivers. Ionospheric effects are visible in double difference phase observations even over short distances. Relative errors up to 30 ppm have been observed in single frequency baseline determination over 10 km in Brazil (Campos et al., 1989).

At times, ionospheric perturbations also occur in mid-latitude regions (Wanninger, 1992, 1994). In particular, so-called *Medium Scale Travelling Ionospheric Disturbances* (MSTIDs) may generate serious problems for precise relative positioning in mean latitudes when the observation time is short (< 20 minutes).

In the short term, the situation will improve because the current sunspot cycle is now in the declining phase, with a minimum expected in 2006. As additional GPS frequencies become available, as part of the "GPS Modernization Program", multi-frequency receivers will enter the market, so that the local ionospheric delay can be directly measured and eliminated.

7.4.4.2   Tropospheric Propagation Effects

The tropospheric propagation delay is critical for precise position and baseline determination, in particular in the height component, because the tropospheric parameters are only poorly correlated over larger distances. Furthermore, it is difficult to separate error components stemming from the radial orbital errors, signal propagation errors, clock errors, antenna phase center variation, and errors in the station height. This is one of the reasons why the height component is much worse than the horizontal components in precise GPS positioning.

For frequencies in the radio spectrum [2.3] the tropospheric delay is independent of the frequency; hence it cannot be determined from dual-band measurements. The near-surface atmospheric structure has to be adequately modeled. Either mean atmospheric parameters, or measured data on temperature, atmospheric pressure, and water vapor content along the signal propagation path must be included in the model. Some of the currently accepted models are dealt with at length in [2.3.3.2]. For further information see also (Mendes, Langley, 1994, 1999; Spilker, 1996d; Langley, 1998b).

Usually, the influence of the neutral atmosphere on range measurements to satellites in the radio frequency domain is expressed by two integral terms: the dry component

and the wet component. The wet component depends on the distribution of water vapor in the atmosphere and is therefore harder to model. The wet portion, however, comprises only 10% of total tropospheric refraction. The total delay in the zenith direction comes to about 2.3 m, and increases near the horizon (10° elevation) to about 20 m. The dry component is precisely described (with an accuracy of ±1%) by the available models. The wet delay can be modeled, depending on the atmospheric conditions, with an accuracy no better than 1 to 2 cm (Langley, 1998b).

Most studies conclude that none of the available models has a clear priority over the others. For low elevation angles, the *Niell model* (Niell, 1996) is usually preferred (e.g. Hay, 2000). This is of particular relevance, because the observation of low satellites (down to 5° elevation) is essential in precise GPS height determinations (Dach, 1999), [7.6.2.3]. The Niell mapping function is of the Marini type (2.116) and uses coefficients depending on latitude and season. For details see also Schüler (2001, p. 157ff).

If the stations are close together, the tropospheric residual error almost completely disappears by differencing in the relative observation mode. It is hence not advisable to introduce the observed meteorological data separately for each station into the adjustment of a small network in non-mountainous regions. The local measurements usually do not represent the regional atmospheric situation with sufficient rigor, and hence introduce biases into the solution. Instead, appropriate identical standard atmospheric parameters should be used for all stations. In this respect, it is of interest that the Niell dry and wet functions are completely independent of surface meteorological measurements.

When station distances are greater (say > 50 km), or when the height differences are larger (in mountainous regions), atmospheric conditions are no longer sufficiently correlated with one another. Adequate modeling, hence, is of growing importance, in particular for precise DGPS or WADGPS applications (e.g. Collins, Langley, 1999) [7.5].

One way of determining the water vapor content of the atmosphere along the propagation path is direct measurement with *water vapor radiometers* (e.g. Nothnagel, 2000). The instruments are, however, very elaborate and expensive and can only be used for major tasks (cf. [2.3.3.2]).

Another approach is to introduce a station dependent *zenith scale factor* for each satellite pass. This parameter can only be estimated reliably after an observation time of 1.5 to 2 hours. To allow for the time variable behavior of the tropospheric zenith delay, stochastic modeling has been successfully applied (e.g. Völksen, 2000). Another option is elevation-dependent weighting (Rothacher et al., 1998). In global networks the introduction of a scale factor is self-calibrating because mis-modeling in the atmospheric delay would produce a scale-factor and hence result in mis-modeling of the orbits and in a violation of orbital mechanics (Nothnagel, 2000).

A very successful approach is real-time monitoring of the tropospheric effects in active multiple reference station networks and the immediate correction of user positions [7.5.3]. Along with the Accuracy Improvement Initiative [7.1.7] enhanced orbital data can be expected when new tropospheric mapping functions will be applied

in the Master Control Station (Hay, Wong, 2000).

Tropospheric modeling remains one of the most demanding tasks in the precise use of GPS. A wealth of information already exists, but research in this respect certainly will continue. The availability of near real-time global and regional high-resolution tropospheric models coming from ground-based (e.g. Schüler, 2001) or space-based (e.g. Reigber et al., 2002) GPS observations considerably contributes to improved data correction. The IGS provides a tropospheric product in the form of combined zenith path delay estimates for more than 210 sites at the level of 3 to 5 millimeters, which corresponds to ∼1 mm in water vapor (Gendt, 2000), see also [7.6.2.9], [7.8.1].

7.4.4.3 Multipath

Multipath propagation means that one or more reflected signals reach the antenna, in addition to the direct signal. Under particular circumstances only the reflected signal may reach the antenna.

There can be reflections off horizontal, vertical, and inclined surfaces (Fig. 7.53), possible examples being streets, buildings, waterways, and vehicles. This should be considered when selecting observation sites, in particular for permanent reference stations.

Multipath propagation affects both code and carrier measurements. The effect on *P-code observations* is two orders of magnitude larger than on carrier phase observations, and can reach decimeters to meters. The effect on *C/A-code observations* is at the order of sev-

Figure 7.53. Multipath propagation

eral meters, and can even reach, in extreme situations, 100 m or more (Braasch, 1996). Under the worst conditions the code signal multipath may cause the receiver to lose phase lock. Many cycle slips are produced by multipath effects.

The code signal multipath becomes particularly critical when the code/carrier combination technique is used for ambiguity resolution, e.g. the extra wide laning technique for "ambiguity solution on the fly" (cf. [7.3.2.3]) in a surveying airplane. Reduction of code multipath is essential for the precise code-only Differential GPS [7.5.1], for example when sub-meter accuracy is required in GIS applications.

Multipath influences on *carrier phase observations* produce a phase shift that introduces a significant periodic bias of several centimeters into the range observation. The direct and the reflected signals are, in a simplified presentation:

$$A_D = A \cos \Phi_D, \quad A_R = \alpha A \cos(\Phi_D + \Phi), \tag{7.145}$$

where
  A_D amplitude of the direct signal,
  A_R amplitude of the reflected signal,

α damping factor ($0 \leq \alpha \leq 1$),
  0: no reflection,
  1: reflected signal as strong as direct signal,
Φ_D phase position of the direct signal, and
Φ phase shift of the reflected signal with respect to the phase of the direct signal.

The superposition of both signals gives the expression

$$A_\Sigma = A_D + A_R = A \cos \Phi_D + \alpha A \cos(\Phi_D + \Phi) = \beta A \cos(\Phi_D + \Theta). \quad (7.146)$$

With $A_{D,\max} = A$, and $A_{R,\max} = \alpha A$, it follows that the equation for the resultant multipath error, Θ, in the observed carrier phase is

$$\Theta = \arctan\left(\frac{\sin \Phi}{\alpha^{-1} + \cos \Phi}\right). \quad (7.147)$$

The signal amplitude is expressed as

$$B = \beta A = A \sqrt{1 + \alpha^2 + 2\alpha \cos \Phi}. \quad (7.148)$$

Inspection of the above equations demonstrates that, for $\alpha = 1$, the maximum value of Θ is

$$\Theta_{\max} = 90°. \quad (7.149)$$

Hence, the maximum error in the L1 signal ($\lambda = 19.05$ cm) is about 5 cm. For linear combinations of L1 and L2 the values can be correspondingly larger or smaller. Their propagation into height errors may reach ± 15 cm (Georgiadou, Kleusberg, 1990). Due to the changing satellite geometry, the multipath effect in carrier phases shows a cyclic behavior. Typical periods are between 15 and 30 minutes, depending also on local reflectors. Fig. 7.54 gives an example for the double difference phase observable.

Figure 7.54. Observed and calculated (solid line) multipath effect

The multipath effect on position results can be minimized with observations over a larger time period, at least over one of the effective cycles. This is not possible in kinematic or rapid static surveying. It is hence important to avoid, or at least to mitigate, multipath propagation, in particular for reference stations, because multipath biases propagate into all rover positions. Possible measures to minimize the effects are

(a) observation design,
(b) receiver and software design, and
(c) station calibration.

(a) *Observation design*
Several measures and actions are possible:
  - − select sites carefully, avoid nearby reflectors,
  - − use antenna ground plane to avoid reflections from the ground,
  - − deploy absorbing material on the ground,
  - − select carefully designed antennas, e.g. choke ring antennas, and
  - − use multiple antenna arrays or controlled antenna motion to average out the multipath variation near the antenna.

One particular procedure is to observe sidereal differences. Since the satellite geometry repeats after 24 hours of sidereal time (23^h 56^{min} UT), the multipath effect for a given site also repeats. By forming *sidereal differences* between the observables, for example double differences DD, at two consecutive days it is hence possible to generate *multipath free observables* DD_{Sid}. These can, for example, be used for the absolute calibration of antennas [7.4.5.1] and also for highly precise control measurements (Seeber et al., 1997a).

(b) *Receiver and software design*
A number of methods for reducing multipath effects use real-time signal processing in the receiver. For an overview see e.g. Weill (1997). The basic idea is to use particular signal properties for improving the correlation process. GPS signals are left-hand polarized. Reflected signals change their polarization, and hence can be detected in the receiver. Reflected signals also arrive later at the antenna and hence can be discriminated. The various techniques have been given names like "narrow-correlator", "strobe correlator", "correlation function", "Everest technology", and so on. Usually, the details are not revealed by the manufacturers; some basic concepts, however, are published, mostly in the ION-GPS Proceedings.

(c) *Station calibration*
A rather new idea is to calibrate stations, in particular reference stations, for multipath effects. A first step is the detection of multipath. Several techniques are possible. In *double differences* over short baselines, most errors are eliminated. The remaining residuals mainly contain the multipath differences. However, it is not possible to separate between the participating stations. Multipath effects are also visible in the signal-to-noise ratio (SNR). The signal strength varies in a sinusoidal form depending on the multipath. Finally, the inspection of sidereal differences helps to analyze the variation of multipath effects.

A method for the absolute field calibration of multipath has recently been
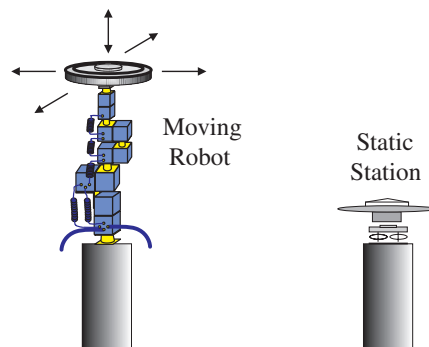
Figure 7.55. Calibration of a reference station for multipath

reported (Böder et al., 2001). One prerequisite is the availability of absolute calibrated antennas, because otherwise the multipath cannot be separated from the antenna phase center variations [7.4.5.1]. The basic idea of this method is to decorrelate the multipath through controlled motion of a robot (Fig. 7.55). The robot operates near the station to be calibrated. The fixed station senses the complete multipath. The moving station eliminates the multipath through the controlled motion. In the double differences only the multipath effects for the fixed station are present and can be described in a functional model, e.g. with spherical harmonics.

Multipath effects at satellites have been reported but seem to be less critical (Young et al., 1985).

7.4.4.4 Further Propagation Effects, Diffraction and Signal Interference

Two main influences have to be considered, signal diffraction and signal interference. *GPS signal diffraction* comes about when the direct GPS signal is obstructed but a diffracted signal is received. Fig. 7.56 explains the geometric situation. Following Walker, Kubik (1996) we distinguish the regions A, B, C, and D. In A, B, C we have direct reception of the GPS signals. In addition we may expect in

A: reflected signals from the ground in front of the obstacle and from the obstacle,

B: only little or no reflection, and

C: reflection from the ground behind the obstacle.

In region D, from the laws of geometrical optics, we have no signal reception except for signals diffracted at the obstacle. The increased signal path of the diffracted signal may produce a phase error of up to several centimeters or even

Figure 7.56. Signal propagation near an obstacle; after Walker, Kubik (1996)

decimeters. The effect can hence be considered as one of the dominant error sources in rapid static or kinematic GPS positioning (Wanninger, 2000).

A powerful means for the detection of diffracted signals is inspection of the signal-to-noise ratio (SNR). A proper weighting of the undifferenced phase observables, based on the SNR values, can be used for minimizing the diffraction effect on coordinate estimates.

Interferences with artificial signals from HF-transmitters occur for frequencies in or near the bandwidths of the GPS signals. The reason is that GPS signals are not transmitted at a discrete frequency but, due to the code modulation, they are spread over a certain bandwidth, namely 2.046 MHz for the C/A-code and 20.46 MHz for the P-code (spread spectrum technique, see [7.1.4]). The effect of disturbances from signals at nearby frequencies can be minimized by adequate filter technology, however

it cannot completely be avoided. The effects on GPS signals, for strong disturbances, are

- decreased SNR level,
- more difficult or impossible acquisition of the GPS signal, and
- loss of signal in phase tracking loop.

Interferences mainly occur for L2. L1-only receivers are usually not affected. Modern receivers, with enhanced filter technology, are better protected than older receiver types. Possible sources for signal interferences are

- VHF, UHF, TV transmitters with strong radiation power within a distance of 100 to 500 m,
- digipeater directional transmission of amateur radio (in Germany just 3 MHz off L2), and
- radar installations of aviation control services, up to 20 km distance.

The influence of high voltage power cables is small. GPS receiver antennas should stay distant by about 10 m, as from all other transmitters, to avoid direct disturbances. Cellular phones seem to have little influence. Possible interference with the forthcoming *ultra-wide-band* technology is under discussion (Akos et al., 2001). For more details on signal interference with GPS see Johannessen (1997), an excellent review in German is given by Kolb (1999).

A particular effect is *foliage attenuation* for stationary and mobile users. Depending on the type of trees and the length of foliage penetration, the attenuation can vary significantly. A detailed treatment of the subject is given by Spilker (1996a).

7.4.5   Receiving System

The main error sources in the receiving system are

- antenna phase center variations,
- receiver noise,
- interchannel bias, and
- oscillator instability.

7.4.5.1   Antenna Phase Center Variation

Positioning in navigation and geodesy refers to the *electrical phase center* of the antenna, that varies with the intensity and direction of the incident signals. For precise applications, the phase center positions of all antennas involved in a project have to be known exactly. This is of particular importance for determination of the height component because in the GPS adjustment the elevation-dependent effects are highly correlated with the height and the tropospheric scale parameter.

The mechanical center of an antenna is usually defined to submillimeter precision. It often coincides with the intersection of the vertical mechanical axis of symmetry and the ground plane (Fig. 7.57). The mean electrical phase centers for the L1 and L2 signals may be a few mm off from the mechanical center. The *antenna reference point* (ARP) is also defined mechanically, usually as the intersection of the vertical

mechanical axis with the lowest part of the antenna housing. For most antenna types, the 3D-coordinates of the offsets of the L1 and L2 mean electrical phase centers with respect to the ARP are given by the manufacturers. The actual electrical phase center depends on the azimuth and elevation of the observed satellites. The deviations of the actual phase centers from the mean electrical phase center are the *phase center variations* (PCV). They can reach millimeters to a few centimeters.

If antennas of the same type are used within one observation session over short baselines, the remaining phase center offsets and variations are eliminated in the differencing process. In cases where the phase center variation is azimuth dependent, all antennas have to be orientated prior to the survey. For this reason, some antenna types have an orientation mark directed to magnetic north.

Figure 7.57. Antenna phase center variation and reference points

If different antenna types are involved within the same project, as is often the case for precise DGPS with reference stations [7.5.1], the observations have to be corrected for the PCV. The same is true when identical antennas are used with very large baselines, because the satellite signals are observed under different elevation angles due to Earth's curvature. Note that different antennas of the same type may also show differences in PCV. For highest accuracy requirements only calibrated antennas should be used. Three major GPS antenna calibration methods are presently available (Rothacher, 2000a):

- anechoic chamber calibrations,
- relative field calibrations, and
- absolute field calibrations.

The *anechoic chamber calibration* (Schupler, 1994; Schupler, Clark, 2001) is a laboratory method, and it is rather seldom applied because not many anechoic chambers exist. A GPS antenna is tilted and dislocated with respect to an artificial GPS signal, generated in the chamber. Absolute PCV are determined under the assumption that they are also valid for observations in the field.

In *relative field calibration* the PCV and the mean offset of a specimen antenna is determined with respect to another antenna, the reference antenna. The PCV of the reference antenna (often the *Dorne Margolin T* choke ring antenna) are set to zero or taken as known. Both antennas are mounted close together on pillars with very precisely known coordinates. The calibration is based on single or double difference residuals. The elevation- and sometimes azimuth-dependent PCV model uses polynomials or spherical harmonics. The method has been widely used to calibrate all major GPS antenna types (e.g. Rothacher et al., 1995; Mader, 1999).

Methods of *absolute field calibration* have only been developed recently (Wübbena et al., 1997; Menge, Seeber, 2000; Wübbena et al., 2000). The basic idea is to elim-

inate multipath effects by either using sidereal differences between observations on two consecutive days, or to use a high-precision robot (Fig. 7.58) that rotates and tilts, at rather high speed, the antenna to be calibrated. Observations from a nearby stationary reference antenna are required to eliminate distance dependent errors; the results are, however, absolute PCVs independent on the type of the reference antenna. Fig. 7.59 shows one example. For details of the method see the cited literature.
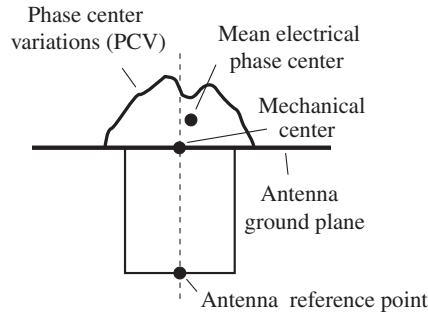
Figure 7.58. Robot for the absolute field calibration of GPS antennas

The advantages of absolute PCV values, from field calibrations, when compared with the traditional techniques are, among others

- they are available in real-time (robot technique) for L1, L2, GPS, GLONASS, future GNSS,
- they are independent of a reference antenna and reference coordinates,
- they are free of multipath,
- they cover the whole hemisphere and are independent of the "northern hole" [7.6.1.1],
- they support the absolute calibration of GPS reference stations, and
- they facilitate the separation from other error sources like troposphere, and satellite antenna phase center offset.

Absolute PCV of modern antennas are mostly below 10 mm, but they can also reach much higher values. The influence on height determination can be several centimeters. It is hence advisable to only use absolutely-calibrated antennas for active reference stations, and for all tasks where high accuracy is required.

In order to facilitate the use of reference data, a "zero-antenna" can be introduced, i.e. an antenna where all observations are corrected for the PCV. The rover then only has to apply its own PCV corrections. A particular RINEX format, *ANTEX*, for the distribution of antenna PCV information has been developed (IGS, 2002b).

Figure 7.59. Absolute PCV of a GPS antenna

At some sites, it is advisable to protect the antenna set-up with a radome against hostile environmental influences. Such radomes may change the antenna PCV characteristics; hence the antennas have to be calibrated together with the radome (Kaniuth, Stuber, 1999; Schupler, Clark, 2001).

Note that the azimuth dependent variation of PCV can also be used vice-versa to determine the orientation of the antenna and of the related platform (Tetewsky, Mullen, 1997). The resolution, however, is only in the order of several degrees.

With an absolutely calibrated reference antenna at hand, it is also possible to calibrate satellite antennas. A first result has been reported for the Block IIa satellites (Mader, Czopek, 2002).

7.4.5.2   Other Error Sources Related to the Receiving System

The *receiver noise* results from the fact, that GPS phase and code observables cannot be measured perfectly but are subject to random influences. For example, the observations are affected by unwanted disturbances in the antenna, amplifiers, cables, and the receiver itself. For details see e.g. Langley (1997b).

As a rule of thumb the observation resolution for classical receivers is about 1% of the signal wavelength. For the GPS signals we obtain:

C/A-code $\lambda \approx 300$ m,   noise ≈ 3 m,
P-code $\lambda \approx 30$ m,   noise ≈ 30 cm, and
carrier $\lambda \approx 20$ cm,   noise ≈ 2 mm.

Modern receiver technology tends to bring the internal phase noise below 1 mm, and to reduce the code-resolution to the 10 centimeter level [7.2.4.2]. Low noise code measurements are important for real-time ambiguity resolution.

Multichannel receivers exhibit different *signal propagation delays* for each hardware channel, since each satellite signal travels along a different electronic path. The instrument makers try to calibrate and to compensate these *interchannel biases*. Multiplexing and software receivers [7.2.1], [7.2.5] are free of interchannel biases. It is recommended that parameters for satellite and receiver hardware delays are included in the parameter estimation models, in particular if the concept of original undifferenced phase data is used (cf. [7.3.2.2]).

Oscillator instabilities play only a minor role in carefully designed receivers because the timing signal is taken from the satellite clock. They can be modeled in the adjustment process. For highest accuracy requirements, and in precise navigation, the use of external precision oscillators (rubidium or cesium) is recommended.

Further error sources that can be counted to the receiving system, are the stability of the station ground and of the pillars, as well as the quality of the station mark. These items are of particular interest in geodynamic networks.

7.4.6   Further Influences, Summary, the Issue of Integrity

Several more aspects exist that influence the achievable accuracy. Among them are the process noise and the tidal upload. *Process noise* means that some liberty exists in

the data analysis approach. We can identify

- *software noise*, the agency is free in the selection of a particular software package,
- *operator noise*, the operator is free in the selection of particular options of the software package, and
- *reference frame realization noise*, there exist various possibilities to select a set of fiducial stations, e.g. from the IGS, to connect a project with a given reference frame.

As a consequence a given data set will lead to slightly different results, when different operators work with different software packages. This is also true for the same package used in different laboratories. An impressive example based on the analysis of a 40-days-long data set from about 50 stations, analyzed with 4 different software-packages at 7 laboratories is given by Dietrich et al. (2001). The mean differences between solutions are 1 cm in horizontal position and 2 cm in height.

Tidal upload means that GPS stations show a vertical displacement due to the crustal deformation caused by oceanic and solid Earth tides (Dach, 1999). The effect can reach several centimeters but it is the same over large areas and hence will be cancelled by relative GPS. Considering today's high accuracy potential of GPS observations, corrections for tidal upload should be applied whenever highest accuracy of the results is attempted. For details see e.g. Dach (1999); Zahran (2000).

In summary, the accuracy achievable with GPS for geodesy, surveying, and navigation, depends on various conditions, for example

- single or multi-receiver operation,
- single or dual-frequency data,
- L2 high quality access under AS available or not,
- receiver noise level,
- static or kinematic positioning,
- real-time or post-processing results,
- accuracy of orbits used, and
- extent of data modeling.

Because of the many options and influences, and the eminent progress in error modeling during the last years, it is not possible and not meaningful to describe the accuracy potential of GPS with a single distance dependent formula, as has frequently been done in the past (e.g. Lichten, 1990). The statement of today is that 1 cm accuracy, at a global scale over all distances, is achievable with appropriate instruments, observation design, and data analysis models. For selected examples, see the section on applications [7.6.2].

A major issue when using GPS in navigation is the *integrity* of the system. For navigational purposes integrity is defined as the "ability of a system to provide timely warnings to users when the system should not be used" (Brown, 1990). The timely warning is, in particular, required for the navigation of civil aircraft. The GPS control segment does not provide sufficient warning when a component of the system fails. Different solutions to the problem have been discussed. With *internal methods of integrity monitoring*, GPS integrity is achieved using information available inside the receiver, such as redundant measurements to additional satellites. This technique is

known as *receiver autonomous integrity monitoring* (RAIM). Using *external methods of integrity monitoring*, the GPS signals are controlled in real-time through a network of ground monitoring stations.

The information is broadcast to users through a *GPS integrity channel* (GIC) via geostationary satellites such as INMARSAT. A further approach to assuring the integrity of the GPS navigation solution is possible by integrating GPS data with data from other sensors. These can be, for example, inertial navigation systems, Loran-C receivers, GLONASS and future GALILEO receivers. For more information see the discussion in the navigation literature, e.g. the journal *Navigation*, and also [7.7.2]. A good introduction to the topic of integrity is Langley (1999b).

7.5   Differential GPS and Permanent Reference Networks

The absolute position determination with GPS is, in general, much less accurate than relative positioning between two stations. This is due to the fact that most of the acting errors (biases) are highly correlated. Error sources can be grouped into three categories [7.4]:

(1) errors decorrelated with distance,
(2) errors decorrelated with time, and
(3) uncorrelated errors.

Errors of type 1, mainly ephemeris and propagation errors, are nearly the same for neighboring stations, as long as they are sufficiently close, and hence disappear in the differences. Errors of type 2 are coped with by synchronized or nearly simultaneous observations. Errors of type 3 affect both participating stations and need a calibration.

To minimize the effect of errors of type 1, instead of absolute coordinates, coordinate differences are determined with respect to a known reference station. Several concepts are in use; the basic strategies are

(a) use of the data of one or more reference stations for post-processing,
(b) use of corrections in position or range from code observations at the reference station in real-time,
(c) use of code-range and carrier phase data from the reference station in real-time, and
(d) use of reference data from a network of reference stations in real-time.

The option (a) is often referred to as *relative GPS*, whereas options (b) through (d) are called *Differential GPS* (DGPS) with different attributes. Option (b) is *ordinary DGPS*, in its proper sense, whereas option (c) is called *precise DGPS* (PDGPS) or also *Real-Time Kinematic* (RTK) GPS. Option (d) is known as the concept of *Multiple Reference Stations*, *Networked Reference Stations*, or also *Network RTK*. The wording, in general, is not uniform.

In this book the term "relative GPS" is used in a general sense, including all concepts where data from more than one station are processed simultaneously, either in post-processing or in real-time. The term "Differential GPS" means that processing of data from more than one station is performed in real-time, or near real-time. Usually,

the original measurements and/or correction data are transmitted in real-time from one or more reference stations to one or more user stations, also called *rovers*. Differential GPS is mostly applied in navigation (see Fig. 7.60). In the following the various concepts are discussed in more detail.
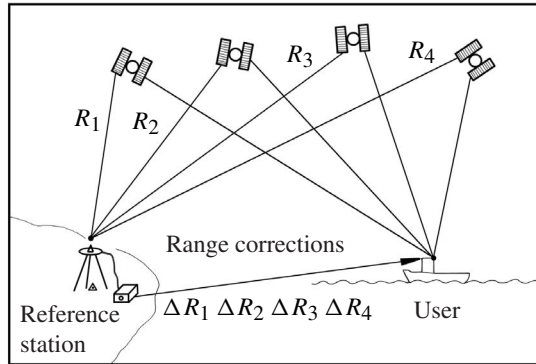
Figure 7.60. Differential GPS; range corrections are transmitted to the mobile user

7.5.1   Differential GPS (DGPS)

7.5.1.1   DGPS Concepts

Differential GPS (DGPS) is a technique that is used to improve the determined position of a roving station by applying corrections provided by a GPS monitoring station, also called *reference station*. Different procedures are in use for generating corrections:

(i) Corrections in the *position domain*

The GPS-derived position of the reference station is compared with its a priori known position. Position corrections Δx, Δy, Δz or $\Delta \varphi$, $\Delta \lambda$, Δh, are transmitted and used to correct the rover position.

(ii) Corrections in the *measurement domain*

The observed pseudoranges to all visible satellites are compared with ranges derived from known satellite and station positions. The differences are transmitted to the rover to correct its observed pseudoranges.

(iii) Corrections in the *state space domain*

Observations from several reference stations are used to estimate the state vector of the biases within the working area.

The first option (i) is rather simple, but also not very flexible. It only works if the same satellites are used at the reference and rover stations, and it is only efficient over short distances. It is therefore seldom applied.

Option (ii) is the ordinary DGPS procedure and is explained in more detail later on in this chapter. It is very flexible and works well within a radius of several hundred

kilometers about the reference station. Due to decorrelation of the biases with distance (orbit, ionospheric and tropospheric delay) the accuracy decreases roughly by about 1 m per 100 km.

Option (iii) is the most flexible procedure, and allows the use of DGPS over larger distances (Wide Area Differential GPS (WADGPS)) and for precise applications in surveying and geodesy (networked reference stations). It is explained in more detail in section [7.5.3]. Whereas option (ii) is based on observed pseudorange corrections (scalar corrections), option (iii) is based on correction vectors.

Several more classifications of DGPS are in use. Following the achievable accuracy we have:

Ordinary DGPS with code range corrections; accuracy 1 to 3 m, depending on the distance from the reference station.

Carrier smoothed DGPS; at the rover station, the carrier observations are used to smooth the coarse code observations with a suitable filter (cf. (7.108)) without solving for ambiguities. The achievable accuracy is < 0.5 m.

Precise DGPS (PDGPS); carrier phase observations, or carrier phase corrections, from the reference station are transmitted to the rover and are used to resolve ambiguities. This procedure is identical to the *Real Time Kinematic* (RTK), see [7.5.2]. Fig. 7.61 shows the accuracy potential of the different options.
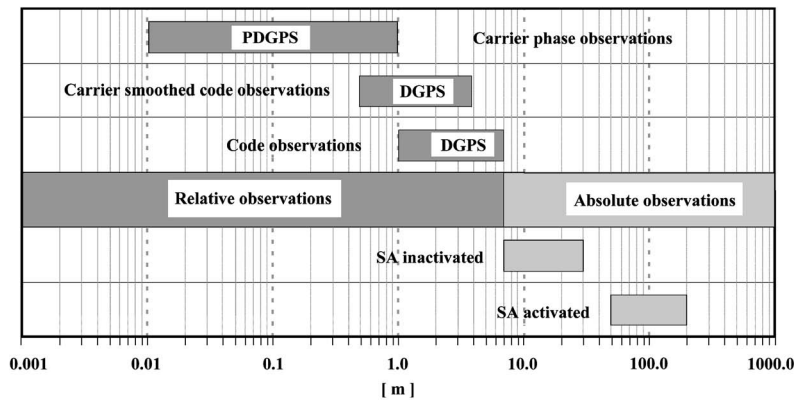
Figure 7.61. Accuracy potential with different modes of DGPS

Another classification is into
   - Local Area DGPS (LADGPS),
   - Wide Area DGPS (WADGPS), and
   - Carrier Phase DGPS (CDGPS), or precise DGPS (PDGPS).

Local Area DGPS (LADGPS) corresponds to the procedures (a) and (b), above. In option (b) a scalar correction to code-phase measurements is applied for each satellite. The corrections are used for areas up to 1000 km radius. However, with the de-activation of SA single receiver accuracy meets the same level of accuracy as DGPS

for distances larger than several hundreds of kilometers. In LADGPS, corrections of all the different error influences are put together in one value. If more than one reference station provides corrections, they can be weighted to form a mean value.

Wide Area DGPS (WADGPS) uses vector corrections for each satellite, derived from observations in a continental or global network of reference stations. The concept corresponds to option (c). The vector consists of individual corrections for the satellite clock, satellite position, and ionospheric delay model. Compared to a scalar correction, a vector correction is valid over much greater areas (Parkinson, Enge, 1996). The concept is discussed in more detail in [7.5.3.1].

Carrier Phase DGPS (CDGPS) is used for surveying applications (see RTK [7.5.2]) and also for attitude control of vehicles [7.6.2.7].

The basic model of the ordinary DGPS concept with code phases is as follows (cf. Misra, Enge, 2001). Starting from equation (7.44), we find expressions for the observed pseudoranges at the user station, PR_u, and at the reference station, PR_r:

$$PR_u = R_u + c\,(dt_u - dT) + d\,\text{ION}_u + d\,\text{Trop}_u + d\,\text{Eph} + \varepsilon_{PR_u}, \quad (7.150)$$

$$PR_r = R_r + c\,(dt_r - dT) + d\,\text{ION}_r + d\,\text{Trop}_r + d\,\text{Eph} + \varepsilon_{PR_r}, \quad (7.151)$$

with

dt_u, dt_r	the receiver clock errors,
dT	the satellite clock error with respect to GPS system time,
$d\,\text{ION}_u$, $d\,\text{ION}_r$	the ionospheric delays,
$d\,\text{Trop}_u$, $d\,\text{Trop}_r$	the tropospheric delays, and
$d\,\text{Eph}_u$, $d\,\text{Eph}_r$	the effect of the ephemeris error, at both stations.

The geometric range,

$$R_r = |\boldsymbol{x}_s - \boldsymbol{x}_r|, \quad (7.152)$$

is calculated from the known satellite position (from broadcast ephemerides) and the predetermined position of the reference station. Any variation with time in the above equations has been neglected for simplicity.

The error in the pseudorange observation at the reference station, the *Differential Correction DC*, is given by

$$DC = R_r - PR_r = -c\,(dt_r - dT) - d\,\text{ION}_r - d\,\text{Trop}_r - d\,\text{Eph}h_r - \varepsilon_{PR_r}. \quad (7.153)$$

In addition to DC, also the range rate of the correction, the *Differential Correction Rate, DCR*, is being determined and transmitted. DC and DCR refer to a reference epoch, t_k; they arrive at the user station with a certain delay or *latency*. Latency is hence defined as the elapsed time from the epoch of the measurement at the reference station until the use of the correction at the remote user site. At the user station, the corrections are predicted for the actual observation epoch, t, as

$$DC(t) = DC(t_k) + DCR \cdot (t - t_k). \quad (7.154)$$

The prediction of DC was critical with activated SA, because its changes with time were rather large. Today, a latency of up to 10 seconds would not be harmful. Typical

latencies are in the order of 3 to 4 seconds. In general it holds that DGPS corrections decorrelate with time if the application is delayed. With (7.150), (7.153) and (7.154) at hand, the corrected pseudoranges $\widehat{PR}_u$ for the user at the epoch of observation are, again omitting a time index, written as

$$
\begin{aligned}
\widehat{PR}_u &= PR_u + DC &(7.155)\\
&= R_u + c\,(dt_u - dt_r) + (d\,\mathrm{Ion}_u - d\,\mathrm{Ion}_r) + (d\,\mathrm{Trop}_u - d\,\mathrm{Trop}_r) +\\
&\quad\ (d\,\mathrm{Eph}_u - d\,\mathrm{Eph}_r) + (\varepsilon_{PRu} + \varepsilon_{PRr})\\
&= R_u + c\,(dt_u - dt_r) + \delta_{\mathrm{Ion}} + \delta_{\mathrm{Trop}} + \delta_{\mathrm{Eph}} + \varepsilon_{PRur}. &(7.156)
\end{aligned}
$$

For little or no latency, the satellite clock error, dT, is identical and hence disappears in (7.155) for no or only small latency. The biases, δ_{Ion}, δ_{Trop} and δ_{Eph}, are negligible for small station separations (a few kilometers) and grow (decorrelate) with increasing interstation distances.

For coordinate determination at the rover station the corrected pseudoranges from equation (7.155) are applied and lead to an improved positioning result. For nearby stations, the only remaining bias term in equation (7.156) is the combined user clock error, $c\,(dt_u - dt_r)$. With a latency of zero the equation is identical to the single difference equation in relative positioning (7.57). Note, that the terms DC and DCR are frequently named PRC (pseudorange correction) and RRC (range rate correction) in literature.

7.5.1.2   Data Formats and Data Transmission

Relative positioning requires the availability, at the rover station, of data from the reference stations. Several data formats have been developed for the transmission of such data. The two most important formats are RINEX, for post-processing or near-online purposes, and RTCM for real-time applications.

The *Receiver Independent Exchange Format* RINEX is described in [7.3.3.2]. It provides the complete data set and is accepted by most software packages. RINEX data can be obtained from the reference station provider via Internet and ftp, via mobile and fixed phone, CD ROM, or other data storage devices, depending on the service. Data from the IGS stations, for example, are available in the RINEX format [7.8.1].

For applications in real-time, in general, the transmission of the complete raw data from a reference station to the rover is not possible, because of the limited capacity of available transmission channels. Instead, the data are pre-processed at the reference station, and only a set of corrections is transmitted. Depending on the accuracy level, different types of corrections are required. For ordinary DGPS, the transmission of code-corrections is sufficient, for PDGPS carrier phase data are also required.

The standards established by the *Special Committee 104 of the Radio Technical Commission for Marine Services* (*RTCM SC-104*, or in short, *RTCM*) are today internationally accepted, and supported by nearly all receiver types. The preliminary version RTCM 1.0, from 1985, was replaced in 1990 by RTCM 2.0. This format

provides pseudorange and range rate corrections and is sufficient for ordinary DGPS (b) with an accuracy level of few meters or better. All necessary information can be transmitted with a bandwidth of 1200 bps (bits per second) or less. Depending on data rate and number of satellites, the required bandwidth can be reduced to 100 bps.

Version RTCM 2.1, from January 1994, additionally includes carrier phase data and hence provides the possibility to resolve ambiguities at the rover station. This version is the required standard for PDGPS and RTK. The necessary data rate is at least 4800 bps. Version RTCM 2.2, from January 1998, includes still further information, in particular the option to transmit correction data from more GNSS systems, e.g. GLONASS. Version RTCM 2.3, from May 2001, is a further refinement, allowing, for example, antenna phase center variation (PCV) data to be included. A new version, RTCM 3.0, is under discussion and will include capabilities for network RTK.

The RTCM message format is very similar to the format of the GPS navigation message. The messages consist of 30-bit words. Each word consists of 24 databits and 6 parity bits. Each message starts with a two-word header containing information such as the reference station identification, the message type, and a reference time for the parameters. In total, 64 message types are reserved, the majority not yet defined. Table 7.16 shows some of the message types valid in RTCM 2.3. For detailed information see e.g. Parkinson, Enge (1996), Kaplan (1996), Hofmann-Wellenhof et al. (2001) or the official document of the RTCM Special Commission 104 (RTCM, 2001).

Table 7.16. Selection of message types in the RTCM 2.3 format

Message type number	Current status	Title
1	Fixed	Differential GPS Correction
2	Fixed	Delta Differential Corrections
3	Fixed	Reference Station Parameters
18	Fixed	RTK Uncorrected Carrier Phases
19	Fixed	RTK Uncorrected Pseudoranges
20	Fixed	RTK Carrier Phase Corrections
21	Fixed	RTK High precision Pseudorange Corrections
31	Tentative	Differential GLONASS Corrections
32	Tentative	Differential GLONASS Reference Station Parameters
37	Tentative	GNSS System Time Offset
59	Fixed	User Defined

For precise DGPS, either message types 18/19 or 20/21 can be used. One advantage of types 20/21 is a higher compressibility when compared with the raw carrier data 18/19. Proprietary formats have been developed to transmit compressed – and possibly decoded – corrections together with other information in the user defined message, type

59. This compression allows the transfer of reference data for all satellites in view (up to 12 SVs) with a bandwidth of just 2400 bps (Wübbena et al., 1996).

The possible data links for the transmission of DGPS data may be categorized as follows:
- ground-based radio links,
- cellular phones,
- satellite communication, and
- internet.

For ground-based radio links there exists a set of general rules:
- the lower the frequency the larger the range,
- the higher the frequency the higher the bandwidth and data rate, and
- high frequency and short range installations are cheaper than low frequency and long range installations.

In the *Low Frequency* (LF) domain, < 300 KHz, data can be transmitted over several hundred kilometers because the waves follow Earth's curvature. One example in Germany, is the system ALF (Accurate Positioning by Low Frequency) where one transmitter near Frankfurt covers all of Germany and beyond (600 to 800 km) with DGPS data at a rate of 3 seconds.

Medium frequency (MF), 300 KHz–3 MHz, transmitters are cheaper than LF transmitters and also cover several hundred kilometers. They are often used as marine radio beacons in coastal areas. The data capacity is sufficient for ordinary DGPS services (several meters accuracy).

Very High Frequency (VHF), 30–300 MHz, and *Ultra High Frequency* (UHF), 300–3000 MHz, radios can communicate over short distances, limited by the line of sight. The data capacity is in the range of 2400 bps and hence suffices for PDGPS with a compressed data transmission format (see above). A further possibility is to use a frequency modulation subcarrier in the *radio data system* (RDS) from broadcasting services. The capacity is then in the order of 100 bps, which is sufficient for ordinary DGPS.

Mobile UHF radio systems, operating at 400 MHz and higher have a range of several kilometers, depending on their power, and can transmit 9600 bps, which is sufficient for RTK applications. A frequently-used wavelength is 70 cm (428 MHz) which, for example, at low power (0.25 W) can be operated without permission in Germany.

Cellular phone is spreading across densely populated areas to provide telephone and data services. With decreasing user fees cellular phones are an attractive alternative to radio frequencies. One disadvantage, however, is that with cellular phones the number of simultaneous users is limited by the number of modems at the reference station, whereas the broadcast system works for an unlimited number of users.

Globally-operating DGPS services use geostationary communication satellites like INMARSAT, or a network of Low Earth Orbiters like *Globalstar*, to transfer DGPS data. The DGPS corrections provide accuracies at the few-meters level. Some of the services use the WADGPS concept [7.5.3.1]. The capacity of INMARSAT also offers PDGPS applications.

A very powerful technique is the distribution of DGPS data via Internet. GPS data from global or regional permanent arrays like the IGS or EUREF networks are already available in the RINEX format on a routine basis (de Jong, 2001). Since quite recently data for real-time applications are also accessible via internet. The JPL is building up an "Internet-Based Global Differential GPS System" (Muellerschoen et al., 2000). Regional and local RTK services via Internet are under development (Weber, 2002). Table 7.17 gives an overview on some DGPS data channels.

Table 7.17. Frequently used data links for DGPS transmission

Name	User	Range	Capacity	Type
radio 2 m	unlimited	tens of km	2400 bps	P-DGPS
radio 70 cm	unlimited	few km	9600 bps	P-DGPS
radio LF	unlimited	hundreds of km	300 bps	DGPS
radio MF	unlimited	hundreds of km	100 bps	DGPS
radio UHF/RDS	unlimited	tens of km	100 bps	DGPS
satellite	unlimited	global	> 2400 bps	DGPS
mobile phone	1 per channel	variable	9600 bps	P-DGPS
internet	unlimited	global	> 9600 bps	P-DGPS

7.5.1.3 Examples of Services

During the last years a large number of GPS reference station services with different architecture and performance has appeared. We can distinguish between global, regional, national, and particular services, as well as between public and commercial, or post-processing and real-time services. In the following, some examples are given.

The most important global reference network is maintained by the *International GPS Service* (IGS) [7.4.3.2][7.8.1], a non-governmental scientific organization. More than 300 stations worldwide are continuously operating and provide, among other products, information on position and observation data, via regional and global data centers. The IGS network (Fig. 7.101, p. 399) is closely related to the ITRF reference frame [2.1.2.2], hence it is possible to connect new GPS observations everywhere in the world directly to the ITRF. The IGS stations generally do not transmit DGPS data in real-time. The IGS is a passive global reference network mainly used for post-processing purposes.

Another global service under development is the *Global Differential GPS System* (GDGPS) of the NASA JPL (Bertiger et al., 1999; Muellerschoen et al., 2001). Based on observations from about 60 stations in the NASA global network, state parameters are modeled and provided to users via Internet. The quasi real-time accuracy is estimated to be 10 cm for the horizontal position and 20 cm in height with a latency of about 1.5 to 3 seconds. The GDGPS approach belongs to option (iii) in [7.5.1.1], see also [7.5.3].

Examples of global commercial services are *Skyfix* and *Omnistar*. Skyfix maintains a network of about 80 reference stations within the reference frame ITRF92. Correc-

tion data, in the RTCM 2.0 format, are transmitted via INMARSAT communication satellites. The achievable accuracy is about 2 m. Omnistar runs about 70 reference stations, covering about 95% of the world. The correction data are distributed via 9 different geostationary satellites, also in the RTCM 2.0 format, and submeter accuracy is promised. Both services apply some state space modeling [7.5.3] in order to obtain the indicated accuracy over large distances.

At the continental level, the *EUREF Permanent Network* (Fig. 7.75, p. 358) can be considered to be a densification of the IGS in Europe. It consists of about 140 (status July 2002) permanent stations and has a similar structure to the IGS. The main purpose is maintenance and control of the European Reference Frame, ETRF 89 [2.1.2.2] (Ádám et al., 2000). Similar so-called "regional networks" are operated in other parts of the world. For example, in South America, in the SIRGAS project [7.6.2] a permanent network of about 30 stations, related to IGS, is continuously operated (Seemüller, Drewes, 2000).

National networks are being established worldwide. Existing maritime radiobeacons are used to broadcast DGPS data, in the standard RTCM format, to marine users in the LF- and MF-bands (285–325 kHz maritime radiobeacon band) with a data rate of 100 to 200 bps (Parkinson, Enge, 1996; Mangs et al., 2001). Beacon networks are mainly located along coastlines or large navigable inland waterways, but are in some regions also expanded inland. In the U.S. the beacon network is the responsibility of the Coastguard. There are plans to cover the whole U.S. with about 80 stations in the *Nationwide Differential Global Positioning System* NDGPS (DOD/DOT, 2001a). Complete coverage is expected for 2003. In Europe, the beacon network nearly covers the complete coastline of the European coastal states (see Fig. 7.62). Within the *EUROFIX* project, tests are underway to use existing Loran C stations for the transmission of DGPS data (Helwig et al., 1997).

The U.S. National Geodetic Survey (NGS) runs the CORS network. CORS stands for *Continuously Operating Reference Stations*. It consists of about 200 stations and is still growing. CORS will meet the post-processing requirements of positioning users by providing code phase and carrier phase observation data in the RINEX format. The data are freely accessible via Internet or anonymous ftp. Depending on the station, the data are recorded at 1, 5, 15, or 30 seconds intervals. For details see Snay (2000).

Canada is establishing the *Canadian Base Network* (CNB), with a spacing between 200 km and 1000 km, depending on the area. CNB consists only of pillar monuments, and is connected to the *Canadian Active Control System* (CACS). CACS basically consists of a number of active stations (14 in 2002), partly remotely controlled. The dual frequency pseudorange and carrier phase data are transmitted to a processing center in Ottawa and are being used as a backbone for estimating state vectors in wide area DGPS systems (e.g. the *Canada-wide Differential GPS* (CDGPS)), see [7.5.3].

In Japan, a nationwide GPS control network, also named *GEONET*, is being established under the responsibility of the Japanese Geographical Survey Institute (GSI). The network consist of about 900 sites, equipped with dual-frequency GPS receivers and additional sensors like tiltmeters and meteorological stations. The spatial density

Figure 7.62. DGPS Beacons in Europe

is very high; the mean distance between stations is about 25 to 30 km. The primary purpose of the network is the determination of crustal strain for earthquake monitoring and prediction [7.6.2.2]. The stations equally provide reference code and carrier data, for surveying purposes, both in real-time and for post-processing. In addition, the data can be used to map tropospheric zenith delay, and to contribute to weather forecasts [7.6.2.9].

In Brazil, a continuously growing network of active reference stations, the *Rede Brasileira de Monitoramento Continuo* (RBMC), coordinated with respect to SIRGAS, provides reference data for precise post processing (Fortes et al., 1997).

Relatively dense permanent networks are already running, or being established, in European countries. Examples are SWIPOS in Switzerland, SWEPOS in Sweden, SATREF in Norway, and SAPOS in Germany. In Great Britain the *Ordnance Survey National GPS Network*, consisting of 30 active stations, provides RINEX data via Internet for users who thereby can directly access the national coordinate system. Additionally, over 1000 passive GPS points are available to support surveying measurements (Cruddace, 2001).

The above examples demonstrate the many activities and different approaches for DGPS services all over the world. In the following, the SAPOS project in Germany

will be explained in more detail because it provides a large variety of services.

SAPOS stands for *Satellite Positioning Service*, and is organized by the German State Surveying Agencies (AdV). The final objective is to cover the complete area of Germany with a network consisting of about 250 permanent stations at a separation of about 40 to 70 km. The rationale behind SAPOS is to provide services to many users who have different requirements concerning accuracy of position results, required observation time, and coverage. SAPOS runs different services, providing different accuracy levels, namely (Hankemeier, 1996)

EPS	Real-Time Positioning Service,
HEPS	High Precision Real-Time Positioning Service,
GPPS	Geodetic Precise Positioning Service, and
GHPS	Geodetic High Precision Positioning Service.

EPS is similar to many commercial and national DGPS services, and provides an accuracy of 1–3 m, sufficient for a broad variety of navigational applications. The correction data are available free of charge via different communication channels.

HEPS is the principal precise real-time service, and it can be used for many applications in surveying and GIS, including cadaster. The position accuracy is between 1 cm and 5 cm, and depends on several influences, in particular on the behavior of distance dependent errors. In order to obtain 1 cm accuracy in real-time over distances larger than a few kilometers, it is necessary to model the error state in the working area (see [7.5.3]). For the application of HEPS a special decoder and payment of user fees are required.

GPPS provides 1 s data from the reference stations for a limited time (e.g. 10 days; thereafter the data are reduced to 15 seconds). Via mobile phone, the data can be transmitted directly to the user in the field for precise near real-time positioning. GHPS requires precise ephemerides, and is a post-processing service. An overview of the SAPOS services is given in Table 7.18.

Table 7.18. SAPOS products

DGPS Service	Positioning Accuracy	Positioning Mode	Data Format	Data Transmission	Data Rate
EPS	1 − 3 m	Real-time	RTCM 2.0	LF, UHF 2 m Band	3 − 5 s
HEPS	1 − 5 cm	Real-time	RTCM 2.1 Modified	2 m Band Cellular Phone	1 s
GPPS	1 cm	Quasi Real-time Post-processing	RINEX	Cellular phone Fixed Phone Data Network	1 s (15 s)
GHPS	< 1 cm	Post-processing	RINEX	Data Network Data Storage	1 s (15 s)

7.5.2   Real Time Kinematic GPS

Real Time Kinematic GPS (RTK) is another name for carrier-phase differential GPS (option (c) in [7.5]). Its eminent characteristic is that users can obtain centimeter-level positioning accuracy in real-time over short distances with an easy-to-handle and highly integrated instrumentation. It is the RTK technology that makes GPS a universal surveying tool, replacing traditional surveying techniques. RTK technology is based on the following features:

- transmission of pseudorange and carrier phase data from a reference station (base station) to the user station (rover) in real-time,
- resolution of ambiguities at the rover station "on the way" or "on the fly" (OTF), and
- reliable determination of the baseline vector in real-time or near real-time.

In a typical RTK configuration (see also Fig. 7.84, p. 369, in [7.6.2.4]), a local (usually stationary), GPS reference receiver transmits pseudorange and carrier phase data over a radio link to the roving station. The GPS receivers may be single- or dual-frequency receivers; dual band equipment facilitates ambiguity resolution (faster, more reliable). The data are transmitted via a data radio (modem). The software runs in the receiver or an external data processor. In general, for the sake of flexibility, identical equipment is used at both stations. Most manufacturers provide highly integrated systems for both the base and rover stations (see [7.2.4.2]).

For the transmission of RTK data, new message types were defined in the RTCM SC-104 version 2.1, in 1994. Message types 18 and 19 contain raw carrier phase and pseudorange information. Alternatively, message types 20 and 21 can be used, containing corrections to the measurements at the reference station.

The transmission of RTK data requires a much higher capacity than does broadcasting of pseudorange corrections. For a set of raw phase corrections from 12 satellites, in format RTCM 2.1, more than 4800 bits are necessary. For a data rate of 1 second the usually available transmission channels with 2400 bps are not sufficient, so that alternative channels with higher data capacity, or particular (mostly proprietary) data formats are required [7.5.1.2]. Depending on the legal situation in a given country, VHF and UHF channels for the required data capacity, but with rather low power, can be used, sometimes without permission. As a consequence, the range of commercial RTK systems is often reduced to a few kilometers.

Another reason for the short range of VHF and UHF transmission is its limitation to the line of sight. The theoretically achievable maximum distance, d, in kilometers between base and rover is (Langley, 1998c) calculated according to

$$d = 3.57 \sqrt{k} \, (\sqrt{h_t} + \sqrt{h_r}). \tag{7.157}$$

h_t and h_r are the heights in meters of the transmitting and receiving antennas above the average terrain level. k is a factor depending on Earth's curvature and the atmospheric refractivity. A mean value for moderate climates is 1.33. To give an example, for a transmitting antenna at 25 m, and a rover antenna at 2 m, above the terrain, the

theoretical maximum signal range is 26 km. In practice signals may be blocked by hills, trees or buildings (see [7.4.4.4]).

Data transmission for RTK applications via Internet is still in the experimental stage (Weber, 2002).

A key factor for RTK is the ability of the rover to resolve ambiguities while the receiver antenna is in motion. This feature is named ambiguity resolution "on the way" (OTW) or, more frequently, *on the fly* (OTF) (for details see [7.3.2.3]). As long as the ambiguities are estimated as real values, (the so-called *ambiguity float-solution*), the achievable accuracy ranges from the meter to the decimeter level, depending on the tracking time. With resolved (fixed) ambiguities, the accuracy numbers go down to the centimeter level. In order to prove the correctness of the ambiguity fixing, the algorithm is initialized at least twice.

Another important factor is the necessary *time to fix ambiguities* (TTFA). Many OTF algorithms use the wide-lane linear combination to accelerate their ambiguity search procedures. The search algorithms are more effective with a large number of satellites and a limited search space. The requirements for a suitable RTK receiver are hence:

   − dual frequency data for ambiguity resolution, also if the baseline is derived from single frequency data,

   − low-noise code pseudoranges to narrow down the ambiguity search space, and

   − all in view capability, to use as many satellites as possible for the search algorithms.

The necessary TTFA strongly depends on the behavior of the distance dependent errors [7.4.4]. For short distances (a few kilometers), and under favorable conditions, the TTFA can be as short as only one epoch.

For the estimation of the baseline vector between the base and the rover station, powerful real-time software is required. The data processing can either follow the concept of parameter elimination (single and double differences) or parameter estimation (undifferenced observables) [7.3.2.2]. Either raw data (RTCM message types 18, 19) or carrier phase corrections (RTCM message types 20, 21) can be used. The carrier phase correction, CPC_r, at the reference station, is given by

$$CPC_r = \Phi_r - \text{Frac}\left(\frac{R_r}{N_r \lambda}\right), \qquad (7.158)$$

with

Φ_r the raw phase at the base station,
N the resolved ambiguity at the base station,
R_r the geometric range at the base station, and
λ the wavelength.

Furthermore

$$N = \text{Int}\left(\frac{R_r}{\lambda}\right).$$

The corrected phase, $\hat{\Phi}_u$, at the rover station is given by

$$\hat{\Phi}_u = \Phi_u + CPC_r. \tag{7.159}$$

In analogy to (7.154), the phase correction rate PCR is also transmitted and applied. The prediction, from the reference epoch, t_k, to the observation epoch, t, is given by

$$CPC(t) = CPC(t_k) + PCR \cdot (t - t_k). \tag{7.160}$$

Carrier phase corrections have several advantages when compared with transmission of raw phase data (Wübbena et al., 1996, 2001b). Corrections are less receiver dependent and more flexible than raw data; this minimizes problems arising from the use of unequal receiver equipment. In addition, corrections for local errors like antenna phase center variations can be applied. Corrections can also be derived from several reference receivers, which makes network solutions possible (cf. [7.5.3]). Another important aspect is that phase corrections require much less bandwidth for transmission than the raw phase data. A capacity of 2400 bps is sufficient.

Most RTK algorithms are proprietary solutions and have a key impact on the performance of commercial RTK equipment. With modern RTK sets a wide variety of surveying tasks can be solved. A short list of possible applications is:
- GIS, cadaster,
- detailed surveying,
- staking out,
- machine control, and
- precision farming.

More applications are discussed in [7.6.2]. A significant limitation of RTK solutions is the fact that the errors and TTFA grow with increasing distance from the base station. A general rule of thumb for the achievable accuracy is

10 mm + 1 to 2 ppm   for horizontal coordinates, and
15–20 mm + 2 ppm   for the height component.

RTK applications are therefore limited to a range of a few kilometers (mostly below 2 km) with a TTFA of just a few seconds. Only in times of low ionospheric disturbances can the range be larger, up to 10 km; however the TTFA will significantly increase. For larger distances, the use of multiple reference stations solves the problem (see [7.5.3.2]).

For more details about RTK see e.g. Langley (1998c), Hofmann-Wellenhof et al. (2001) or the information brochures of receiver manufacturers.

7.5.3 Multiple Reference Stations

One of the serious drawbacks of DGPS is the fact that the influence of some error sources, such as orbit, ionosphere and troposphere, grows with increasing distance from the reference station. In other words, the *error correlation* decreases with station separation, and the *error decorrelation* grows. The effect is roughly 1 meter increase in the DGPS positioning error per 100 to 150 km, for single frequency code receivers.

It would require a very large number of individual DGPS installations to cover a single country or even a whole continent.

A solution to this problem is the idea to interconnect several reference stations, and to transmit their measurement data in real-time to a central processing station. All data are used in a common filter to estimate the error state for the whole area and to separate the error components. The state vector can then be applied to improve the corrections for the complete area, as a function of the geographic user location, resulting in much better accuracy and a much sparser density of reference stations. Alternatively, as a first step, simple interpolation algorithms can be used.

This concept was developed early on under the name *Wide Area Differential GPS* (WADGPS) for the use of code-range measurements in continental networks [7.5.3.1]. Only recently, a similar concept has been developed for high precision DGPS using carrier phase data in local, regional or national networks. The concept is known as *Networked Reference Stations*, *Virtual Reference Stations*, or *Area Correction Parameter Approach* [7.5.3.2]. Note that the terms are not uniformly used in the literature.

7.5.3.1   Wide Area Differential GPS

The term *Wide Area Differential GPS* (WADGPS) was coined by C. Kee and others, in 1991 (Kee et al., 1999). The basic idea was to establish a sparse network of reference stations, over an area as large as the continental U.S. (CONUS), to provide high quality DGPS corrections to navigation users at land, on sea, in the air, as well as in the near-space. A WADGPS system consists of (see Fig. 7.63)

— a sparse network of reference stations, equipped with dual frequency GPS receivers, high precision clocks and optional meteorological sensors,
— a master control station that receives all measurements and estimates the differential corrections,
— an upload station, to broadcast the corrections, possibly via GEO satellite, to the users, and
— monitor stations, to control the system.

The main objective is to overcome the decorrelation of the distance dependent errors by using suitable network algorithms. An excellent overview of the basic algorithms is given by Mueller (1994). A rough separation is into measurement and state-space domain algorithms.

Measurement domain algorithms do not estimate the individual error components, but form a weighted mean of all corrections from the participating reference stations. The weighting scheme may use

— a distance weighting (the nearest station gets the highest weight),
— an elevation angle weighting (higher elevation satellites get more weight),
— an age weighting (lower latency gets higher weight),

or other criteria.

State space domain algorithms try to identify the individual error sources and transmit the information to the user in a suitable form. The components are, in particular,
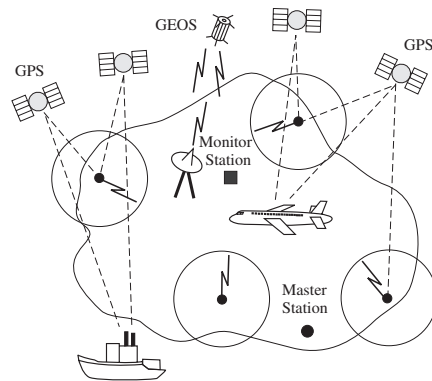
Figure 7.63. Architecture of an Wide Area DGPS installation

a 3-D ephemeris error, a satellite clock offset, ionospheric delay parameters, and tropospheric parameters. User apply this information as a function of their geographical location.

The measurement domain algorithms are the simplest, and hence cheapest, solutions. The corrections, however, are not independent of distance, but degrade with growing separation from the center of the network. The concept is hence seldom used. The state-space domain approach is baseline independent and provides the highest accuracy. The error components in the state vector also vary rather slowly, so that the data transmission rate can be low. For details about the algorithms see also Kee (1996); Kaplan (1996).

The advantages of using WADGPS when compared with a single DGPS reference station are obvious:

  − coverage can be extended over inaccessible regions like water areas,
  − the number of reference stations can be reduced, and
  − the biases are nearly or completely distance independent.

The WADGPS concept has been realized in various services. In the U.S., the Federal Aviation Administration (FAA) is establishing the *Wide Area Augmentation System* (WAAS) to meet safety-related integrity requirements and to support as well the en route as the precision approach phases of flight (FRNP (2001), see also [7.7.2]). In Europe, the *European GPS Navigation Overlay System EGNOS* has similar objectives [7.7.2]. The same is true for the Japanese service (MSAS). All three augmentation systems will be interoperable to provide seamless global coverage. In Canada the *Canada-Wide Differential GPS* (CDGPS) provides accurate differential corrections, via communication satellite, for the whole country to support positioning and navigation at the meter level. The service is, however, not intended for commercial aviation (no integrity channel).

The aforementioned commercial services, *Skyfix* and *Omnistar*, like other global or regional commercial services, also apply the WADGPS concept. The approach of

Omnistar is very similar to the idea of "virtual reference stations" VRS, see [7.5.3.2].

7.5.3.2   High Precision Networked Reference Stations

With the installation of reference stations providing carrier phase data for precise DGPS applications in real-time (e.g. SAPOS in Germany [7.5.1.3]) the problem of distance dependent errors became evident. When 1 cm accuracy is required, the number of reference stations with the necessary density would be unrealistically high, in particular during periods of strong ionospheric disturbances. A solution to the problem comes from interconnection of the reference stations, and the estimation of the error state in the working area in real-time (state-space domain approach, option (iii) in [7.5.1.1]). Fig. 7.64 demonstrates the problem and its solution. For "ordinary" baseline RTK (without network), the achievable accuracy decreases with increasing distance; at the same time the TTFA increases. In the networked solution the error state in the area is estimated and transmitted to the rover, where the measurements can be corrected accordingly. As a result, the accuracy and the TTFA remain at a constant level independent from the distance.
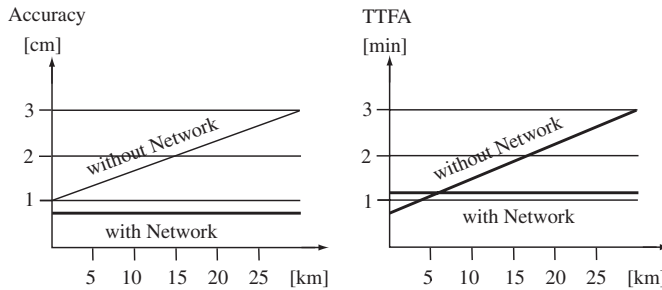
Figure 7.64. Modelling the distance dependent errors in an interconnected reference network

The current data formats for DGPS corrections do not allow to transmit the complete state vector to the roving station; hence the state vector needs to be represented by a simplified model. The basic idea of this approach is as follows (Wübbena, Willgalis, 2001). The observation equation for carrier phase observations between the antenna phase centers of satellite, i, and receiver, j, is

$$PR_j^i = |\boldsymbol{R}_j^i| + \delta B_j^i + \lambda N_j^i + \varepsilon_j^i. \tag{7.161}$$

The equation can be written for each particular signal, s; the index, s, is omitted for clarity. The bias term δB_j^i comprises the terms

δC_j^i for clock related errors,

δD_j^i for distant dependent errors, and

δS_j^i for station dependent errors,

hence

$$\delta B_j^i = \delta C_j^i + \delta D_j^i + \delta S_j^i. \tag{7.162}$$

The clock related errors δC_j^i contain components originating in the satellite and the receiver clock, and signal delays in the hardware of the satellite and the receiver. The distance dependent errors, δD_j^i, are composed of the ionospheric delay, δI_j^i, the tropospheric delay, δT_j^i, and the orbit error vector, $\delta \boldsymbol{o}^i$, hence

$$\delta D_j^i = -\delta I_j^i + \delta T_j^i + \frac{\boldsymbol{R}_j^i}{|\boldsymbol{R}_j^i|}\delta \boldsymbol{o}^i. \tag{7.163}$$

The station dependent errors, δS_j^i, finally, are composed of multipath and receiver antenna phase center variation (PCV). For completeness, multipath and PCV at the satellite antenna can be included.

Precise positioning with carrier phases requires the correct determination of the phase ambiguities, N. As has been outlined in [7.3.2], two approaches are possible:
 – parameter elimination, and
 – parameter estimation.

In the parameter estimation approach all biases have to be estimated together with the coordinates and ambiguity terms. The parameter estimation procedure with undifferenced observables has some advantages when compared with the parameter elimination process, for example (see also [7.3.2.2]):
 – biases can be constrained by specific models,
 – precise clock models can be used,
 – absolute information is maintained,
 – more flexibility with changes in network design, and
 – different receiver types and different GNSS signals can be adapted more easily.

It is obvious that the parameter estimation concept is particularly well suited to be used in the state space approach, because all biases can be separately modeled and the distance-dependent biases can be applied as corrections. Table 7.17 gives an overview of possible functional and stochastic models for the above mentioned error sources.

Once all state parameters are estimated with sufficient accuracy, they can be transmitted to the user, who can eliminate the corresponding error terms from the observation equation. As a result, precise absolute coordinates for the user antenna can be determined. This is basically the approach for precise point positioning (PPP) in global and regional networks (see e.g. [7.3.4] and Muellerschoen et al. (2001)). For operational multi-station networks, for the time being, a simplified *state representation* is used instead of the complete state vector.

In a first step a network solution with ambiguity fixing is established for the participating reference stations, and all states are properly estimated. The measured ranges at the reference stations are filtered using the state space model of the network. The differences between filtered and computed ranges, for all satellites, give residuals which

Table 7.19. Functional and stochastic description of GPS error sources, after Wübbena, Willgalis (2001)

Bias	Functional Model	Stochastic Model
Satellite clock	2nd order polynomial	white noise process
Signal delay (SV)	constant	white noise process
Satellite orbit	Cartesian elements	3D Gauss–Markov process
Ionospheric delay	single layer model	3D Gauss–Markov process
Tropospheric delay	modified Hopfield model	2 scaling parameter/station
Receiver clock offset	–	white noise process
Signal delay (rcv)	constant	white noise process
Satellite PCV	–	–
Receiver PCV	calibration	–
Multipath (rcv)	elevation dependent weighting	1st order Gauss–Markov process
Measurement noise	–	white noise process
Carrier phase ambiguity	constant after fixing	–

are separated into ionospheric, orbit and tropospheric residuals. This separation is possible because of the proper state estimation. In some cases, the orbit and tropospheric residuals are combined as *geometrical residuals*.

In a second step, the residuals are interpolated between the reference stations. Fig. 7.65 demonstrates that a rover experiences an error, δ_1, by using the range correction, ε_1, only, and an error, δ_2, by using the range correction, ε_2. With a linear interpolation between the reference stations, RS_1 and RS_2, the interpolation error is just $\delta\varepsilon$. Investigations show that up to distances of 100 km a linear representation is sufficient. Larger spacing between reference stations requires use of a polynomial of 2nd or higher order, depending on the spatial decorrelation characteristics of the particular error sources.

From the various proposals of how to correct the measurements at the rover station, two frequently applied procedures are the concepts of
- Area Correction Parameters (ACP), and
- Virtual Reference Stations (VRS).

The interpolation between two stations, as in Fig. 7.65, only models the errors along one baseline. For three reference stations, the state residuals can be represented by a plane (Fig. 7.66). For each epoch, the time variable parameters, $a_\varphi(t)$ and $a_\lambda(t)$, describing the inclination of the plane are determined. The parameters $a_\varphi(t)$ and $a_\lambda(t)$ are called *area correction parameters* (ACP). They are estimated at a rather slow rate, of about 10 seconds, separately for the two distance dependent error components, the *ionospheric component* and the *geometric component*, and transmitted in addition to the conventional PDGPS range corrections (e.g. RTCM 2.1) from the reference station
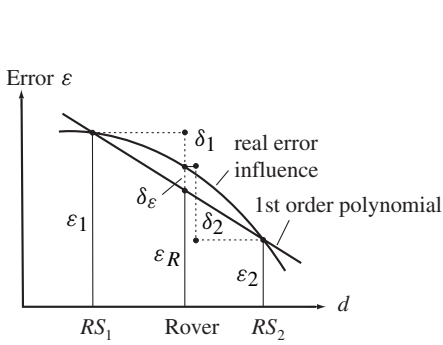
Figure 7.65. Interpolation of distance dependent errors (Wuebbena1998)
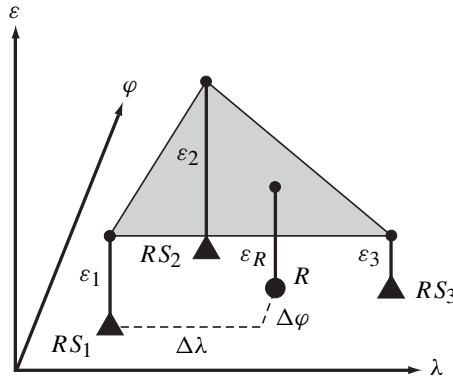
Figure 7.66. Linear modeling of area correction parameters (ACP) for three reference stations

located at φ_0, λ_0 to the rover. With

$$\varepsilon_R(t) = a_\varphi(t)(\varphi - \varphi_0) + a_\lambda(t)(\lambda - \lambda_0), \tag{7.164}$$

the user can then compute the corrections, $\varepsilon_R(t)$, valid for his approximate position, φ, λ.

Experiences show (Wübbena et al., 2001b; Willgalis et al., 2002) that the approach works well for distances up to about 50 km between reference stations. For larger station separation, more sophisticated representation techniques have to be developed.

An alternative approach is the use of so-called *virtual reference stations* VRS (Weber, Tiwari, 1995; Wanninger, 1998, 2000). Here, the user communicates his approximate position to the analysis center of the reference network. Based on a state estimation of the network, as discussed before, the analysis center computes a set of range corrections valid for the approximate rover positions and transmits these "virtual observations" to the rover. The rover accepts this data set as PDGPS corrections from a nearby reference station, and applies conventional RTK algorithms.

Both procedures have advantages and disadvantages. In the VRS concept only one data link is required between reference and rover station, and conventional RTK software can be applied. Disadvantages are that only a limited number of users can work at the same time, and that for moving rovers the virtual reference station is also moving. Most RTK software, however, only accepts fixed reference stations. The main advantages of the ACP concept are its unlimited number of users, higher flexibility and correctness in the error state modeling when undifferenced phase data are used, and the possibility to include moving reference stations.

As a result of either procedure, the rover is able to determine its position in real-time with an accuracy of about 1 cm, independent of the distance from the reference station, see Fig. 7.67. This situation helps to support many tasks in detailed surveying [7.6.2].
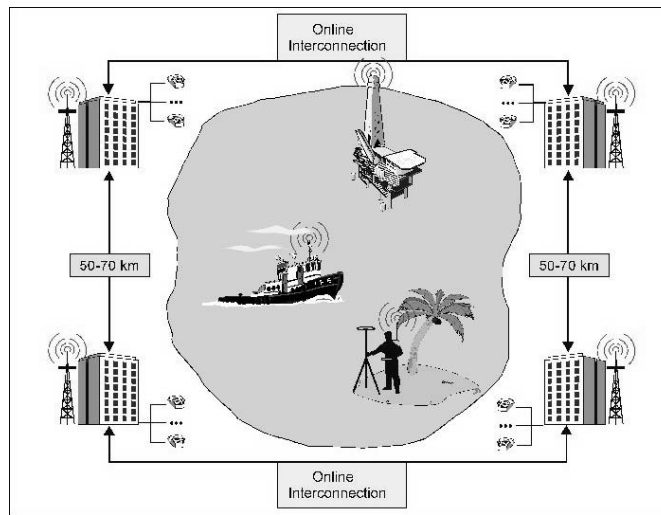
Figure 7.67. Concept of precise real-time positioning in an interconnected network

The concept of multiple reference stations for precise positioning in real-time is already applied in some areas, for example in Germany within the SAPOS service ([7.5.1.3], Jahn, Winter (2002)). Other developments are reported in Vollath et al. (2000) or Raquet, Lachapelle (2001). In total, however, development is still in its early stages. A next step will be the unification of existing networks, and an extension of services (Wübbena et al., 2001b; Wübbena, 2002). Global state parameters (orbit, global ionosphere, and troposphere) can be taken from global networks, like the IGS, and introduced into the state modeling of regional or local networks. The same is true for a combination of sparse national and dense regional networks. The state parameters of the national network, including solved ambiguities, are forwarded to regional or local networks that are only established in more densely populated areas. This concept, of an *adapted PDGPS network*, is in particular suitable for large countries where a uniform dense coverage is neither feasible nor required (Willgalis et al., 2002).

7.6 Applications

7.6.1 Planning and Realization of GPS Observation

In the early days of GPS, the planning and execution of field projects resembled, in many aspects, the execution of Doppler projects with TRANSIT [6.6]. Experience gained in the preparation, organization, and execution of Doppler projects could be transferred for the most part to GPS work, and has influenced GPS practice. It is hence of interest to study chapter [6] and to read some of the original TRANSIT publications. Due to the much broader field of applications, however, the issue of

GPS project planning and realization has its own significance, and is widely discussed in the literature, e.g. Jäger (1990); Santerre (1991), Hofmann-Wellenhof et al. (2001, chap. 7). The following gives the most important aspects.

7.6.1.1   Setting Up an Observation Plan

As long as the GPS system was not yet complete, a pre-computation of satellite coverage was an indispensable preparatory step in project planning. With the system completely deployed in 1995, sufficient satellites are visible above the horizon at any time; hence field campaigns can be planned independently of the constellation. For analysis purposes, and for kinematic observations, in particular in areas with obstructions, a pre-computation of the satellite constellation can still be of importance.

These so-called *ALERT-lists* can be computed with data from the satellite almanac. *Almanac data*, that is, low-accuracy orbit data for all available satellites, are transmitted in the fourth and fifth subframes of the navigation message [7.1.5.4]. These subframes have 25 "pages" each 30 seconds long, so that the complete almanac information can be read in 12.5 minutes.

With the aid of the almanac data, satellite positions can be precomputed over several months with sufficient accuracy for planning purposes. One must, however, occasionally expect larger orbit maneuvers, so that a regular check of the almanac data is recommended. With the almanac data, visibility diagrams (Fig. 7.68) and PDOP values [7.4.2] can be generated. Most manufacturers provide suitable software packages (*mission planning software*) on a PC basis. The almanac data are listed in Table 7.20. Almanac data are available from various internet sources, for example from the U.S. Coast Guard.

Table 7.20. Almanac data

PRN	Space vehicle identification number [–],
000	Health status (000 = healthy) [–],
e	Orbit eccentricity [–],
$\sqrt{A}$	Square root of the semi-major axis [$\sqrt{m}$],
Ω_0	Right ascension of the ascending node [degrees],
ω	Argument of perigee [degrees],
$\overline{M}_0$	Mean anomaly [degrees],
t_{0a}	Reference time for almanac data [s],
δ_i	Difference of orbit inclination from 54° [degrees],
$\dot{\Omega}$	Nodal rate [degrees/s$\times 10^{-3}$],
a_0	Clock correction [s$\times 10^{-9}$],
a_1	Drift of clock correction [s/s$\times 10^{-9}$], and
XXX	GPS week [–].

From the almanac data the current position vectors of the satellites can be calculated in the CTS coordinate system, using the formulas in [7.1.5.3]. With known approximate absolute coordinates (φ, λ, h) of the observation site, the satellite's azimuth and elevation can be found as a function of time with standard formulas, and used for the construction of visibility diagrams. A corresponding visibility diagram in stereographic projection (sky plot) is shown in Fig. 7.68. The related bar diagram is given in Fig. 7.15, p. 231.

The sky plot shows a certain lack of symmetry in the distribution of satellite tracks. This comes from the fact that the inclination of the GPS orbits is 55° and hence defines an area of the observer's sky (shadow area) where it will not be possible to make observations. The shadow area is a function of the observer's latitude and is equal for any longitude. Fig. 7.69 gives an example for equatorial, polar and mid-latitude observers (Santerre, 1991). For observers at northern mid-latitudes the shadow area is also called the "northern hole".
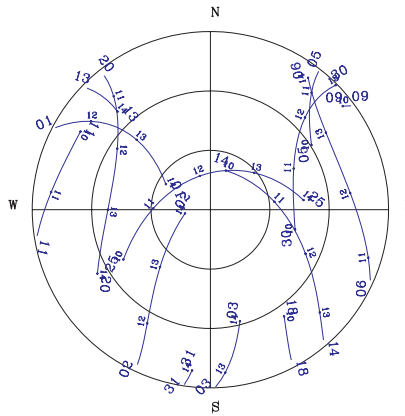
Figure 7.68. Visibility diagram (*Sky Plot*), 4 hour period, for Washington DC
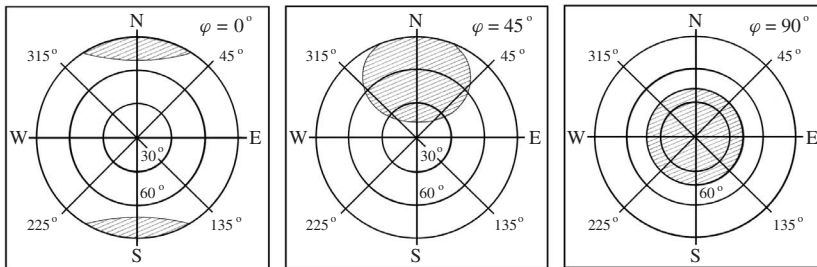
Figure 7.69. Shadow area as a function of the observers geographic location

Additional information for planning purposes is given by the computation of PDOP values that reflect the geometrical strength of the satellite configuration [7.4.2]. As long as the complete satellite coverage had not yet been installed the PDOP values indicated the best time periods for observations (*observation window*). With the current status of GPS, the importance of the PDOP criterion should not be over-emphasized; most

geodetic receivers can track all visible satellites and do not require any pre-selection, whilst for most geodetic applications the observation period is long enough to average out the influence of geometry. With the full GPS constellation the PDOP is sufficiently low for most of the day. Hence the PDOP criterion is of interest only for

- navigational purposes,
- kinematic surveying, and
- applications with satellites obscured by obstructions.

The necessary length of observation depends on the purpose of the survey, the instrument type, the desired accuracy, the software capacity, and logistic aspects. The basic requirement for precise surveys is the resolution of phase ambiguities. Once the ambiguities are resolved the observations can be finished. Over short distances (up to 10 km), with sufficient satellites (six or more), dual frequency receivers, and advanced software, this period can be as short as a few minutes or even less [7.3.2.3]. In kinematic surveying the few centimeter level can be achieved continuously (see [7.5.2]).

Over larger distances and under difficult environmental conditions (such as ionospheric disturbances, multipath), several hours of observations are required to obtain a precise ambiguity float solution [7.3.2.3]. For the establishment of national or continental fundamental networks, and for geodynamic purposes the observations can last 24 hours or even several days, to average out orbital, meteorological, multipath, and other time-variable effects.

Under difficult logistical conditions (e.g. in areas with difficult access) it is advisable to increase the usual observation time in order to avoid the need for re-occupation of sites in cases of poor data. Hence the following observation scenarios can be distinguished:

24 hours up to several days	fundamental networks, geodynamics,
several hours	highest accuracy over larger distances or under difficult conditions,
15 to 30 minutes	control surveys with short distances up to 10 km, and
continuous measurements	rapid methods and navigation.

With the installation of permanent networks and dense arrays of GPS receivers, continuous reference observations are available for many tasks. The number of specifically organized "GPS-projects" will decrease. Most observations in applied geodesy and surveying will be done with respect to existing reference stations or networks.

7.6.1.2   Practical Aspects in Field Observations

Advance local *reconnaissance* can be essential for successful observations. The observation sites should have unobstructed visibility and should be accessible to vehicles. As a general rule, a free line of sight down to the horizon is required in all directions.

In forested areas or near buildings, a satellite visibility diagram (sky plot, cf. Fig. 7.68) helps in the site selection. However, if the sites are to remain usable for later observations with other satellite constellations, it is recommended that the horizon be generally open, at least down to a 10° angle of elevation. Note that for precise

height determination observations down to 5° are advantageous [7.6.2.3]. Existing obstructions should be documented in the reconnaissance sheet in a *shadow diagram* (Fig. 7.70).

The GPS technique requires and permits selection criteria other than those of classical triangulation techniques. Control points no longer have to be installed on topographic elevations or towers with mutual station intervisibility, but rather where ever they are needed, on easily accessible sites with a minimally obstructed horizon. Also places near high buildings, towers, power lines, and transmitting antennas are not suitable. Nearby walls or other reflecting surfaces can cause multipath effects [7.4.4.3].

Figure 7.70. Shadow diagram, indicating obstructions with elevations $\geq 10°$

In wooded areas, the antennas can be mounted on light masts. In the case of non-centric observations, however, centering and plumbing must be done with the same accuracy with which the GPS measurements can be evaluated, that is centimeters to millimeters. Note that eccentricity calculations have to be done in the 3-D space (cf. [2.1.7]). In the interest of convenient future use for topographic and surveying purposes, points should be selected where centric observations can be made.

In many cases the follow-up surveying is done with conventional equipment, for example with electronic tacheometers. Points should be selected such that either a free sight is available to a nearby surveying mark, or an intervisible second GPS point has to be installed a few hundred meters away.

The *monumentation* of station marks usually follows the general rules of the responsible surveying and mapping authorities. Regarding the high accuracy potential of GPS the monuments should be established on stable ground, if possible on rock, or concrete blocks with sufficiently deep foundation. The station marker should be defined to at least 1 mm, for example with a fine grid mark on a corrosion-resistant metallic rivet. In such cases, the GPS stations can also be used for control purposes and engineering surveying. In addition, the station markers should be suitable as an exact vertical reference. The central survey marker is usually controlled by eccentric reference marks.

All essential information should be documented in a *reconnaissance sheet*. Possible elements are
  − station name and identification code,
  − description of site,
  − approximate coordinates and height,
  − accessibility (car, road conditions, walking distance),
  − necessary antenna height (tripod, mast),
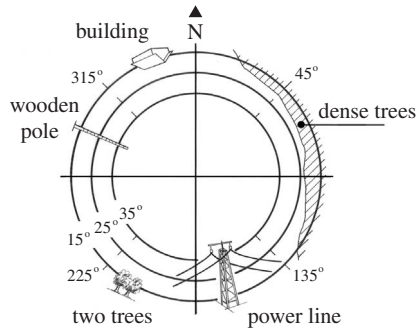
– orientation marks, and

– shadow diagram.

Power supply is no longer a major problem in practical field work as it was for older receiver types, such as the TI 4100 [7.2.4.1]. Modern instruments have a very low power consumption. Internal batteries usually last for a whole working day or even longer. For security reasons it is advisable to recharge batteries every other day.

Whereas older receivers could only be operated by skilled people, modern receivers work completely automatically. Dialogue with the receiver is possible, but not required in standard operation. Usually all visible satellites are tracked, and no pre-selection, or change of constellation, is necessary. The personnel should be able to carefully

– mount the tripod on the station mark,

– measure the antenna height,

– control the receiver operation,

– work according to a given time schedule,

– run the station control sheet (station log), and

– measure additional data if required (meteorological data, eccentric elements).

In most projects the measured GPS data have to be stored on suitable recording media for subsequent computation, e.g. for multistation adjustments. Modern receivers have built-in solid state memories, or plug-in memory cards. Depending on the memory capacity and the amount of data, the recorded measurements have to be downloaded to a computer once a day or at the end of a campaign; or the memory cards have to be exchanged. In larger field projects it is recommended that the data are transferred to suitable mass storage in the observation area, making a data check at the same time. Data can also be transferred via cellular phone and/or internet from the field to a central processing facility.

The amount of incoming data is enormous if the full data rate of modern receivers is exploited, i.e. once or twice per second. For most static applications a much lower data rate is completely sufficient, for example every 15 or 30 seconds. It is, however, essential that all receivers in one project sample at the same data rate. This condition may cause problems if different receiver types are used.

In some cases, the meteorological data are used in the subsequent multi-station evaluation. The data have to be recorded at adequate intervals, for example every 30 minutes: pressure (± 1 mm), temperature ($\pm 1°$ C), and relative humidity (± 1 %). However, note that meteorological station data can introduce biases into the multi-station adjustment because they are not representative for the working area.

It is advisable to keep a *station log*, entering not only the weather data but also the station identification code, the receiver and antenna identification numbers, the antenna position and height, the observation schedule, operation problems, and other significant information of relevance to future data processing.

7.6.1.3 Observation Strategies and Network Design

Three basic observation strategies can be distinguished (cf. [7.3.4]):

– point positioning concept (single receiver),

− baseline concept (relative observations at two stations), and

− multistation concept (three and more receivers operating simultaneously).

Various procedures can be selected within the last category. Of particular relevance are observations in connection with active multiple reference stations [7.5.3].

The choice of the observation concept depends on the objective of the survey, the required accuracy, the number and type of receivers available, and the logistic conditions. Hence a general classification is difficult and not appropriate. As regards the necessary accuracy, the following user classes may be defined, though the boundaries are debatable (Table 7.21).

Table 7.21. GPS user classes

	Category	Average required relative accuracy	Corresponding accuracy in [m], distance dependent
A:	Exploration geophysics Georeferencing low accuracy GIS	$1 \cdot 10^{-4}$	$1 \ldots 50$
B:	Topographic map surveys Small scale engineering Vehicle control systems	$1 \cdot 10^{-5}$	$0.2 \ldots 1$
C:	Cadastral surveys Engineering surveys of mean accuracy	$5 \ldots 1 \cdot 10^{-6}$	$0.01 \ldots 0.2$
D:	Geodesy, Control surveys High precision engineering surveys	$5 \cdot 10^{-7} \ldots 1 \cdot 10^{-6}$	$\leq 0.01 \ldots 0.05$
E:	Geodynamics Highest precision engineering surveys	$1 \cdot 10^{-7}$	$0.001 \ldots 0.02$

With a *single receiver*, an absolute position determination can be achieved continuously (navigation mode) with an accuracy of 5 to 15 m, without SA, under the Standard Positioning Service (SPS), cf. [7.1.6], [7.4.1]. Even after several hours of observation the achievable absolute accuracy is not better than several meters. Therefore, only group A activities can be undertaken with a single receiver. A new situation evolves with the use of precise ephemerides and clocks in the concept of *Precise Point Positioning* (PPP). In essence this is, however, an implicit form of differential GPS [7.3.4].

For all other user groups, only *relative observation techniques* with at least two simultaneously operating GPS receivers are worth considering. The terms *differential GPS* and *translocation observations* are also used equivalently.

The concept of relative observations is extensively discussed in chapter [7.5]. It applies for moving and static antennas. The essential strength of relative techniques lies in the fact that a part of the error influences at neighboring stations is strongly correlated and is therefore cancelled out when a difference is taken (cf. [6.5.3] for the TRANSIT system). This is especially true of orbit errors, errors of the satellite clock, and errors in the ionospheric modeling.

Comparison of Table 7.12 and Table 7.6 makes it clear that the systematic model errors and the observation noise of the code have more or less the same order of magnitude, namely 1 to 10 m. Hence, in the navigation mode, a relative navigational accuracy of ± 2 to 3 m is successfully achieved using code phase measurements and corrections from a reference station (*differential GPS* [7.5.1]). For a static receiver, the extremely low observation noise of the carrier phase measurement, which is three to four orders of magnitude less than the systematic error effects, can be used to advantage only if the systematic components are eliminated by relative measurements.

In this way, an accuracy increase by a factor 10^3 to 10^4 is brought about in the geodetic relative mode with at least two simultaneously operating receivers, as compared with the single receiver mode. Relative techniques are particularly effective when the station distance is small compared with the satellite range ($\sim 20\,000$ km). The amount of correlation decreases as the distance increases; however, the correlation is effective up to several thousand kilometers. The adjustment models for relative observations are discussed in [7.3.4] and [7.5].

If two receivers are available, a point field or network can be set up by the observation of *baselines*. One possibility is to operate one instrument at a central station, and occupy the adjacent points in a star-shaped pattern (Fig. 7.71). Adjacent central stations A, B, C,... are linked through baseline observations. The baselines between the non-simultaneously occupied stations can then be derived by computation. For control purposes, some of those "trivial" baselines (cf. [7.3.4], Fig. 7.42, p. 284) can be independently observed.

Another possibility is to occupy neighboring points and form triangles or quadrangles (Fig. 7.72). This method leads
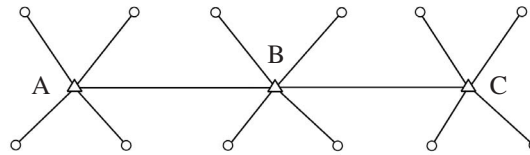
Figure 7.71. Baseline observations with two receivers
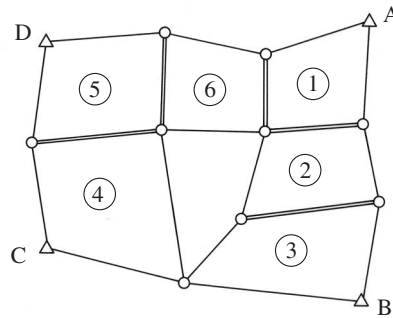
Figure 7.72. GPS network

to a high relative accuracy, in particular if the quadrangles in Fig. 7.72 are subdivided into triangles, but it is very laborious. It is clear that the use of more than two instruments is much more economical, even for small sized networks. The configuration in Fig. 7.71 can be worked with two instruments in 14 observation sessions, and with three instruments in seven sessions.

All observations made simultaneously during a given time period in the course of a GPS project are called a *session* (cf. [7.3.4]). Each session has to be connected to at least one other session of the network through one or more identical stations where observations have been carried out in both sessions. An increasing number of identical stations increases the stability, accuracy, and reliability of the total network.

When three or more receivers are used in a multi-session project, the design of an observation plan becomes an optimization problem between efficiency (economy), accuracy, and reliability. Some basic considerations are discussed here. We define

   r number of simultaneously operating receivers,
   n number of stations,
   m number of stations with more than one observation in two different sessions, and
   s number of sessions.

We know already from (7.100) that

   $r(r-1)/2$ number of possible baselines in one session, and
   $(r-1)$ number of independent baselines in one session.

The number of sessions required for a given network is:

$$s = \left[\frac{n-m}{r-m} \right], \tag{7.165}$$

with s being the next larger integer number. With two or more reoccupied stations in each session, some of the baselines are determined twice. In the total network we have

$$\begin{aligned} s(r-1) \quad &\text{number of independent baselines, and} \\ (s-1)(m-1) \quad &\text{number of double determined independent baselines.} \end{aligned} \tag{7.166}$$

Let the network example in Fig. 7.72 be observed by four receivers with two connecting points between consecutive sessions. From (7.165) and (7.166) we find:

   13 stations,
    6 sessions,
    5 double determined baselines, and
    9 repeatedly determined stations.

A detailed inspection of Fig. 7.72 shows that four stations are observed once and seven stations are observed three times. From the economic point of view, a homogeneous distribution of reoccupations would be favored, because it provides an equally distributed redundancy with the least number of sessions required.

Some software packages include the possibility of executing simulation calculations with a given station and receiver configuration. The importance of these features

must not be overemphasized because network configuration is only one aspect of a GPS mission.

As regards logistic and practical limitations, the choice of an observation strategy will often be guided by experience, with formal optimization criteria providing valuable aid. Since the accuracy of a local GPS network is only little dependent of the station distance, the design aspects are mainly governed by logistic, economic, and reliability factors. Some general rules from experience are that

- each station should be occupied at least twice, under different conditions, to identify blunders,
- neighboring stations should be occupied simultaneously because the ambiguity resolution works best over short distances,
- for medium-sized projects the use of 4 to 10 receivers is a good compromise with respect to logistics, production rate, and reliability, and
- a certain number of baselines should be observed twice for accuracy checks.

These rules are valid for independent projects. In active multiple reference station networks the situation is different, insofar as a new station is always determined by a single receiver with respect to the whole network [7.5.3].

Besides accuracy, the *reliability* of a GPS network is an important issue of network quality. Reliability means the ability of a network to self-check against blunders or systematic errors. Fig. 7.73 gives an example (Augath, 1988). Stations A and B are used
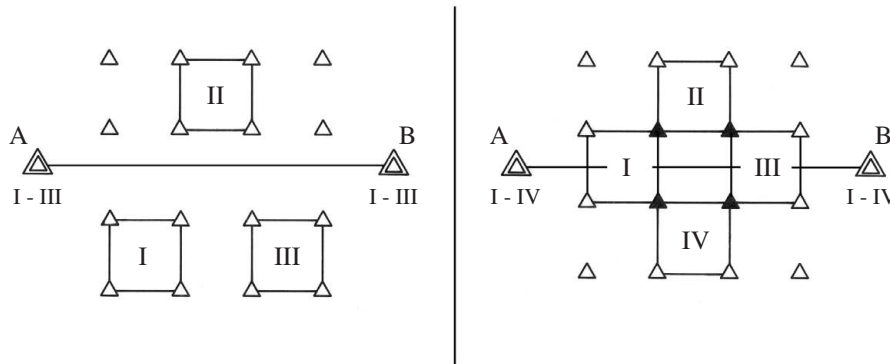
Figure 7.73. Network with accuracy criteria (left, three sessions) and reliability criteria (right, four sessions)

as reference points, and are occupied during all sessions. Four receivers are mobile. The left part of Fig. 7.73 demonstrates a network design with accuracy criteria only. The right part yields more or less the same accuracy, but in addition offers reliability, because each point (solid triangle) is used in two sessions. Additional constraints come through the permanent stations. An even more controlled network is shown with Fig. 7.74.

The data set which results from a GPS multi-station adjustment process has a high relative accuracy. The absolute coordinates, however, may have standard deviations of several meters because of the uncertainty in the realization of the satellite datum through observations (cf. [6.6.1] and [7.3.4]). As a rule, therefore, newly determined "GPS networks" must be tied to previously existing known points, either from the particular national control network, or from fundamental stations which are determined by precise global techniques such as VLBI, Laser, or GPS tracking networks. Examples of the latter group are

Figure 7.74. Generic network densification with GPS

- International Earth Rotation Service Terrestrial Reference Frame (ITRF) [2.1.2], [12.1.2],
- International Global Positioning System Service (IGS) [7.4.3], [7.8.1],
- National or regional GPS tracking networks like the Canadian ACS, the U.S. CORS, the Brazilian RBMC, or the German SAPOS [7.5.1], and
- Continental or national fundamental GPS networks like EUREF, SIRGAS, and DREF [7.6.2].

The tie can be made over one or several identical points, for example A, B, C, D in Fig. 7.72, or nearby permanent stations. For smaller working areas, a single connection point may be sufficient. The control points can be used as *fixed points* with minimum variances, or as *fiducial points* with a predefined, non-vanishing dispersion matrix. The network datum is derived from the pre-existing control points rather than from the GPS observations. In active reference networks the datum comes from the network datum. For detailed discussion see [12.1.1] and [7.6.2.1].

When planning GPS projects in remote areas, careful attention must be paid to connection to reference points with known precise geocentric coordinates. Otherwise, the errors in the absolute coordinates, inherent in the actual GPS observations, will propagate into the relative coordinates of the network solution [7.6.2]. A good solution to the problem is to connect new measurements with IGS stations.

With the evolution of sufficiently dense global, continental, and national fundamental networks, based on precise space techniques as well as on GPS, the reference point or fiducial point concept will be the technique usually applied when establishing GPS networks. In other words, GPS will be mainly used as an interpolation technique for network densification in the working area.

A final generic example (Fig. 7.74) highlights some of the essential items that have been discussed in this chapter. Stations A, B, C, D are points of the existing network,
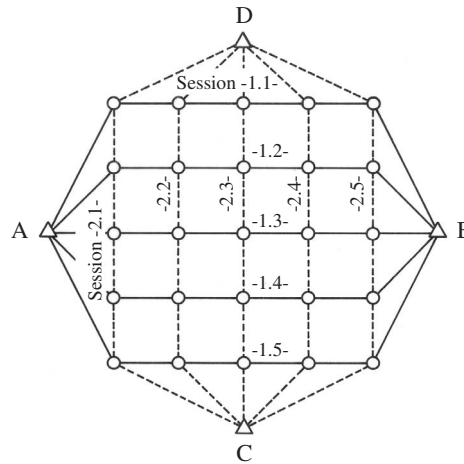
either from previous GPS campaigns, or from a precise classical terrestrial network. They are used as fixed or fiducial points. The datum is completely defined through these stations. Seven receivers are applied. Two of them are operated on the fixed stations, and five are moving. Five sessions are observed each project section. During the first section the fixed stations A and B are occupied, and the roving receivers move in five sessions according to the solid lines. During the second section stations C, D are occupied, and the receivers move according to the dashed lines. Depending on the interstation distances, the receiver types, the available software, and the project objective, the individual sessions can last minutes, hours, or even days. Note that the following principles are fulfilled for the newly determined points:

- *high accuracy*, caused by a sufficiently long observation period in each session (following the project goals),
- *highly economic*, because the session number for double occupancy, and the interstation travel times are minimized, and
- *high reliability*, because each new point is derived from two completely independent determinations (new antenna installation), tied to different control points, and (mostly) observed under a different satellite constellation.

For observation strategies using active reference networks see [7.5.3] and [7.6.2.1].

7.6.2 Possible Applications and Examples of GPS Observations

Since GPS is an all-weather, real-time, continuously available, economic, and very precise positioning technique, almost unlimited possibilities are opened up for its use in geodesy, surveying, navigation, and related fields, including

- control surveys,
- geodynamics,
- altitude determination,
- cadastral surveying and GIS,
- monitoring and engineering,
- precision navigation,
- photogrammetry and remote sensing, and
- marine and glacial geodesy.

Some typical fields and examples of GPS application will be discussed in the following. The use of satellite methods is further reviewed in [12].

It was recognized early on that GPS is a multipurpose system. One major advantage is its capability of forming a powerful building block in integrated systems. GPS together with a coordinate system and geographic information produces a map. GPS together with a map facilitates navigation. GPS together with a digital geometric data base, a geographic information system (GIS), and a communication link produces a command and control system (Gibbons, 1991).

With the establishment of continuously operating reference stations covering a whole country [7.5], the acceptance of GPS as a basic positioning tool will further grow. The availability of position information in real-time at any level of required

accuracy and at any place will be taken for granted as today is the availability of precise time or of communication links.

Because of the fast growing application market, only some basic concepts are described here. For more information on the current discussion see journals like *GPS World* or symposia proceedings like *ION GPS*. The statements in this chapter refer to NAVSTAR GPS. They are, however, also valid for other GNSS systems like GLONASS or the forthcoming European GALILEO [7.7].

7.6.2.1   Geodetic Control Surveys

The following objectives can be identified:
   (a) setting-up of a completely new field of control points,
   (b) densification or extension of existing networks,
   (c) inspection, analysis, and improvement of existing networks, and
   (d) establishment of a network of active reference stations.

The terms "network" and "control point field" are used as synonyms.

(a) New network
The installation of a completely *new network* can be performed in three steps. Since all densification work will be done with GPS techniques, it is advisable to select a global geocentric datum compatible with the World Geodetic System WGS 84 [2.1.6]. WGS 84 is now defined with an accuracy level of about ±1 cm (Merrigan et al., 2002) and corresponds at that level with the *International Terrestrial Reference Frame* ITRF [2.1.2.2], [12.4.2]. Areas with an insufficient coverage of ITRF sites, for example Africa, or some parts of Asia (see Fig. 2.4) are densified by stations of the IGS service [7.4.3.2], [7.8.1] with the same accuracy standard. For most practical purposes, the global network ITRF2000 and the IGS network can be considered as equivalent.

Starting from the global network, three basic levels of "Geodetic GPS Networks" may be distinguished, all with the same high accuracy standard, namely about 1 cm:
   − Level A: Continental (or Sub-Continental) Reference Frames,
   − Level B: National Fundamental Networks, and
   − Level C: All other GPS networks.

At *Level A*, a continental or sub-continental GPS network is installed, with the ITRF/IGS sites as fiducial points. The interstation distances are between 300 km and 500 km. The station coordinates have to be determined with the highest achievable accuracy, in general ±1 cm. This is possible with the fiducial point concept [7.4.3.2], about one week of observations, dual frequency receivers, precise orbits and advanced software.

As an example, see the EUREF (European Reference Frame) project. EUREF has been built up since 1989 by successive GPS campaigns. The existing ITRF stations in Europe (Laser and VLBI) were used as fiducial points. The first campaign was performed in May 1989 with about 60 dual-frequency receivers. In 1990, some 30 stations were added during the EUREF North campaign. After 1990, in several campaigns stations from Eastern Europe were included. The European Reference System

was defined as ETRS, in agreement with the ITRS, for the epoch 1989.0. Its realization is ETRF89 that coincides with the ITRF89 for stations in Europe. The basic idea is that ETRF89 rotates with the stable part of the European plate and hence can remain unchanged for a long time period. About 90 stations of the more than 200 EUREF sites form the permanent EUREF network [7.5.1.3] (Fig. 7.75), with the objective to maintain the ETRS and to densify the IGS network in Europe.
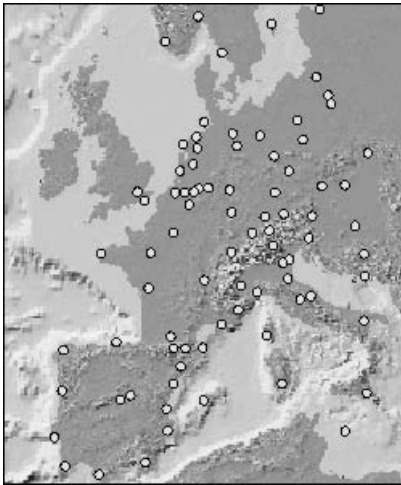
Figure 7.75. EUREF Permanent Network, source: BKG

Figure 7.76. SIRGAS 1995, source: DGFI

Similar basic reference frames have been or are being built up in other continents or subcontinents. In South America, the *Sistema de Referencia Geocentrico para America del Sur* (SIRGAS) was created in 1995 by 10 days of simultaneous GPS observations at nearly 60 stations (Fig. 7.76), Hoyer et al. (1998). The network was tied to ITRF94. Some of the stations continue as permanent stations, provide reference data, and maintain the frame. The data are processed in the IGS *Regional Network Associate Analysis Center* for SIRGAS (RNAAC SIR) [7.8.1] and contribute to a densification of the IGS global network. In 2000, the SIRGAS network was re-observed and enlarged including sites in Central and North America. In North America, the already mentioned U.S. CORS network [7.5.1] and the Canadian Active Control System (CACS) play a similar role. A continental network for Africa, AFREF, is under discussion.

At *level B*, nationwide, or statewide fundamental networks are installed with a spacing of 50 to 100 km, depending on the size of the country and the objectives. The stations of level A are kept fixed for use as fiducial points. The accuracy of the individual GPS station with respect to the neighboring stations is again ±1 cm, hence providing a homogeneous set of coordinates for the whole country.

One example is the DREF campaign (Fig. 7.77) in Germany. DREF was observed early in 1991 with 83 dual frequency receivers. The network contains 109 stations with a mean spacing of 70 to 100 km. Some 20 stations are EUREF sites from level A.

It is advisable to use as many receivers as possible to provide a homogeneous set of observations. In most cases it will not be possible to occupy all stations of a national network simultaneously. The total network has then to be broken down into sub-networks and sessions. The single sub-networks and sessions are interconnected via fiducial stations (from level A), and by selected "identical" points at the rim of the individual sub-networks. Most countries have established fundamental networks of this type or will do so within the near future.

Before using station coordinates from level A as a reference frame for densification at level B, the coordinates have to be corrected for crustal deformation, if applicable. Even small motions of say 2 cm/year will lead to a 10 cm deformation already after 5 years, which is not tolerable in precise geodetic networks. The procedure is as follows (Drewes, 1998).

Figure 7.77. The DREF (German Reference Frame) network

Step 1: Transformation of level A coordinates of the stations, S, used as connecting points (fiducials) from the epoch, t_0, of the reference frame (level A) to the epoch, t_i, of the new observations. Station velocities, v_S, derived either from repeated observations or from crustal deformation models, are applied according to

$$X_S(t_i) = X_S(t_0) + v_S(t_i - t_0). \tag{7.167}$$

Step 2: Network adjustment of the new stations, N, (level B) using the observations at epoch t_i together with the coordinates, $X_S(t_i)$, of the fiducial points from level A.

Step 3: Transformation of the new station coordinates, X_N, from observation epoch, t_i, back to the epoch, t_0, of the reference frame (level A), using

$$X_N(t_0) = X_N(t_i) - v_N(t_i - t_0). \tag{7.168}$$

This procedure ensures a homogeneous network of level B stations in the datum of level A. Since station velocities of the new stations are not always available, it is advisable to develop continuous deformation models for all continental plates [12.4.1].

At *level C*, all other control points have to be connected to stations of level B, again at the 1 cm accuracy level. One advantage, when compared with classical techniques, is that no systematic densification is necessary. Work can be done, following a priority schedule, where coordinates are required. The densification procedure can follow the scheme of Fig. 7.74. The classical division into geodetic networks of 1st to 4th order, within a country, will disappear, and be mostly replaced by two levels:

− the fundamental national reference frame (level B), and
− all other control points (level C).

(b) Densification of an existing network
This can be treated in different ways.

(1) A precise classical terrestrial network of 2nd or 3rd order exists. In this case GPS is used as a modern surveying tool for precise network densification. GPS is in nearly all cases much more economical than the classical methods. The procedure is as outlined in Fig. 7.74, see [7.6.1.3]. The existing control points are taken as fixed reference points. The existing national datum is maintained. This approach is used in many countries as a tool for rapidly providing precise geodetic control.

(2) A terrestrial network of medium or low accuracy exists; the old coordinates shall be maintained. In this case the distortion of the traditional network is introduced into the precise GPS results. GPS is only used as a method of cost-effective interpolation into the existing national framework. The GPS observations should be preserved for a rigorous adjustment once a fundamental GPS network has been established in the area at some later date. This procedure is acceptable as an intermediate solution, in particular in developing countries, until a completely new network and datum, based on satellite techniques, can be established.

(3) The existing terrestrial network is combined with new GPS observations. In this case the existing network datum is maintained; however the complete network is readjusted and strengthened with the inclusion of GPS measurements. New points are linked to the existing network in an optimal way. All network coordinates are slightly changed. The method only works if sufficient stochastic information on the existing network is available, see [12.1] and related literature, (Leick, 1995; Strang, Borre, 1997).

A particular problem arises when multiple reference stations [7.5.3.2] are established in an area with existing geodetic control. Even if the traditional network is of highest quality, discrepancies at the several centimeter level have to be expected when "distortion-free" new GPS points, derived from references stations at about 30 to 50 km distance, are established in the direct neighborhood of existing "distorted" surveying points. Such discrepancies are often not acceptable in cadaster or engineering projects. Two solutions are possible:

In a first step, for the whole area, local transformation parameters have to be derived from GPS observations at a sufficient number of existing control points. These parameters are either used to transform all existing surveying points into the distortion-free reference frame defined by the GPS reference stations, or they are used to transform the GPS determined coordinates of new object points into the existing distorted local frame, realized through the conventional surveying points. In the latter case the GPS results should be maintained in order to use them for a new coordination as soon as the former solution can be realized.

(c) Analysis of an existing network.
This procedure is of particular importance in countries where little information on the original observation and computation is available, for example in developing countries

(e.g. Campos et al., 1989). The analysis, however, also offers a very important insight into the present official networks of countries with an advanced cartographic tradition, such as Germany. For the analysis a certain number of existing stations is re-occupied with GPS. The residuals, after a seven-parameter Helmert transformation (2.46) are inspected.

Fig. 7.78 shows residual vectors between an early GPS campaign in Germany (DÖNAV) and the official terrestrial network DHDN (Seeber, et al., 1987). The residuals reach up to 1 m. A similar analysis is used to derive detailed expressions for a transformation formula between the datum of the GPS network and the existing local network.
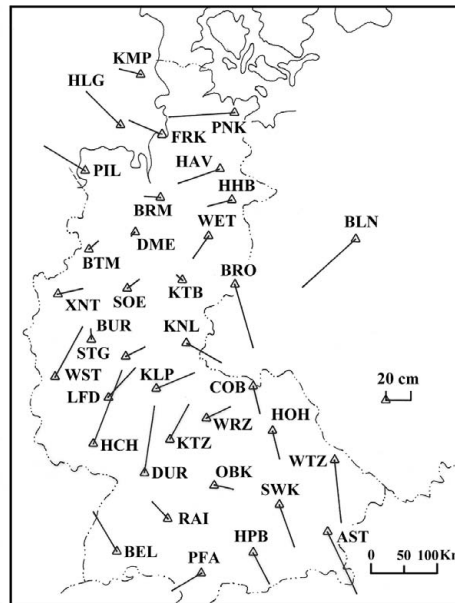
Figure 7.78. Residuals after a 7-parameter transformation between the DÖNAV network and the classical German network DHDN

(d) Active Reference Stations.

A modern tendency is to represent the fundamental reference frame in a country by a network of *active control points* that provide relative information, for any authorized user, on a routine basis (see [7.5.3.2]). This service can be operated under the responsibility of the national surveying authorities (for example SAPOS in Germany). The long-term rationale behind this concept is to substitute the reference frame exclusively through the active reference stations and to considerably decrease the number of monumented points.

Fig. 7.79 shows some possible concepts of active reference station networks at level C. In version (a) GPS data are collected at the reference stations and distributed to users via a control station. In option (b) range corrections in the RTCM 2.0 or RTCM 2.1 format are broadcasted to users from the nearest reference stations. In
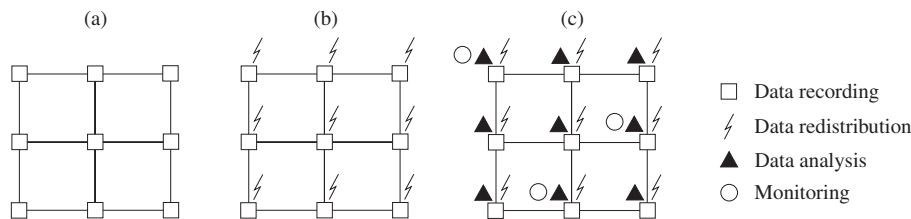
Figure 7.79. Different concepts of permanent reference stations

version (c) all stations are interconnected and work as monitor stations and analysis stations; they all transmit range corrections and ACP to the users [7.5.3.2]. The subject of network densification no longer arises.

7.6.2.2   Geodynamics

The very high accuracy potential associated with comparatively easily transportable equipment makes GPS a suitable technique for determining recent crustal movements [12.4.1]. Until about 1985, crustal movements were mainly analyzed with Very Long Baseline Interferometry (VLBI) [11.1] and satellite laser ranging (SLR) [8]. With VLBI, long-range baselines can be determined precisely; a few millimeters accuracy and precision over several thousand kilometers are achievable.

The main disadvantage of the VLBI method is the enormous technical expenditure and the limitation to a comparatively small number of fundamental stations; only very few transportable systems are available. With satellite laser instruments, very precise and reliable movement rates have been derived from many years' observation, for example, in the area of the San Andreas fault (Watkins et al., 1990), or along with the WEGENER/MEDLAS project in the Mediterranean region (Ambrosius et al., 1991). Transportable satellite laser ranging systems are also in use [8.3.3]; still, the use of SLR technology involves high costs and long mobilization times. For many areas of interest, in particular if a large number of points are to be determined for higher spatial resolution, GPS offers considerable advantages. This is why since about the late 1980s, besides VLBI and SLR, GPS is the technology of preference for the operational determination of crustal deformation and global plate motion.

In the early days of GPS one of the most important limiting factors in the error budget for precise baseline determination over large distances was orbit accuracy. Following the rule of thumb (7.134) an orbit error of about 2.5 m would propagate 1 cm error per 100 km into the baseline. In view of the known motion rates of a few cm/year or only mm/year, station spacing should then not be much greater than 100 km. With today's orbit accuracy of 5 cm or better for IGS products [7.4.3.2], the orbit is no longer a critical factor in crustal motion studies, even over large distances. Key factors of the error budget are rather [7.4.4][7.4.5]:
  − modeling of atmospheric propagation effects,
  − antenna phase center variations (PCV), and
  − multipath effects.
Much research has been invested in recent years into the modeling of tropospheric and ionospheric propagation effects [2.3.3]. The use of data from GPS observations in LEO (GPS-MET, see [7.6.2.9]) projects will further help to improve the models. In addition an attempt can be made to raise the accuracy level through the use of *water vapor radiometers* [2.3.3.2], [7.4.4.2].

Tectonically active areas near the geomagnetic equator or in high latitudes will experience large ionospheric disturbances [7.4.4.1], e.g. Wanninger, Jahn (1991), Völksen (2000). The use of dual-frequency receivers is hence essential. Long observation periods, over at least 24 hours, help to average out residual effects.

Site dependent effects can be minimized with absolutely calibrated antennas and multipath reducing observation techniques [7.4.5.1]. Even for identical antennas, the PCV variation will not be cancelled in relative observations over very long baselines, because of Earth's curvature (Menge, Seeber, 2000).

Two strategies are being used to determine station velocities:

 (i) repeated observations within dedicated campaigns, and
 (ii) continuous observations at permanent installations.

Strategy (i) was mainly used during the development phase of GPS, and it is still applied for smaller independent projects or in remote areas with difficult access. A first epoch measurement establishes a network of well demarcated stations, and repeated epoch observations are performed after one or several years. A typical example is given with the Iceland campaigns in Fig. 7.80.

With the availability of fully automatic, low power consumption GPS receivers and the possibility to transfer data over large distances, strategy (ii) is more and more applied. One main advantage, compared with option (i), is, that data are continuously available and sudden events, like displacements due to earthquakes, can be directly analyzed. Two eminent examples are the IGS network and the GEONET in Japan [7.5.1.3].

The following main fields of application for crustal motion monitoring can be identified:

 (a) global and continental plate motion and deformation analysis,
 (b) regional crustal motion analysis, and
 (c) local monitoring of deformation and subsidence.

Projects of *group (a)* show very impressive results after a couple of years of observations. Comparisons between GPS and other space techniques like VLBI and SLR demonstrate an agreement at the centimeter-level, and hence prove the capability of GPS for global geodynamics (Boucher et al., 1999).

A major break-through came with the establishment of the IGS [7.8.1]. More than 300 globally distributed stations deliver data on a permanent basis and as such provide a continuous monitor of deformation. The station velocities can be used to compute global stress maps and to determine a kinematic model of the individual plate rotation vectors (see [12.4.1], Fig. 12.13, p. 529, Tab. 12.3, p. 528).

Two examples are given for continental projects. The motion of the Antarctic plate was determined with two epoch measurements in 1995 and 1998 (Dietrich et al., 2001). Three weeks of observations, each time at about 45 stations on the Antarctic continent and the adjacent tectonic plates, were taken to establish a precise reference network linked to the ITRF96 reference frame, and to determine, besides of local deformations, the rotation of the Antarctic plate. Based on a data analysis with four different software packages at seven analysis centers, the combined solution yields an accuracy of 1 cm for the horizontal and 2 cm for the height components. For details see Dietrich (ed.) (2000). With horizontal velocities of about 2 to 3 cm per year an epoch difference of three years gives reliable results. For the detection of height changes the situation is more critical. A longer time span, and even more sophisticated modeling is required (see [7.6.2.3]).

The SIRGAS network in South America was observed in 1995 with 58 stations and again in 2000 with, in total, 184 stations. The results from the repeated stations are used to derive their velocity vectors. This information is also of high importance for follow-up geodetic work because SIRGAS 95 was adopted as a national datum by some of the participating countries. About 20 of the SIRGAS stations deliver data on a continuous basis. These data are included in the data set of the Regional IGS Network RNAAC SIR and continuously provide information on the motion of the South American Plate (DGFI, 2001).

An extremely challenging endeavour in this project group is the connection of continental control points with submarine control points near plate boundaries or sub-duction zones, because GPS measurements on floating platforms have to be integrated with underwater acoustic measurements [12.3.2], (Chadwell et al., 1998).

Projects of *group (b)* already show significant results. Investigations and epoch or continuous measurements have been started in nearly all tectonically active parts of the world. Well known examples are, among many others control networks in California, the CASA (Central and South America) and SAGA (South American Geodynamic Activities) GPS project, the GEODYSSEA (Geodynamics of South and South-East-Asia) project (Wilson et al., 1998), projects in the Mediterranean area (Kaniuth et al., 2001), and the neo-volcanic rifting zone in Iceland. Usually, displacement vectors are derived from the comparison of two or more epoch measurements if no continuous measurements are available. Fig. 7.80 shows the results derived from two early epoch

Figure 7.80. Displacement vectors from two consecutive epoch measurements in Iceland 1987–1990

measurements in 1987 and 1990 in the Northern Volcanic Zone of Iceland. About 50 stations were controlled with seven TI 4100 dual frequency P-code receivers. The epoch accuracy of adjacent stations is about 1 to 2 cm. The identified displacements in a post-rifting period reach about 3 to 5 cm/year. Subsequent epoch measurements in 1992, 1993, and 1995 provided a deeper insight into the mechanisms and enabled geophysical modeling and interpretation (Hofton, Foulger, 1996; Völksen, Seeber, 1998). Fig. 7.81 shows deformations after the $M = 8.1$ Antofagasta Earthquake on July 30, 1995 (Klotz et al., 1996).

One major difficulty in the analysis of a displacement field is the identification of "stable" reference points. Powerful methods have been developed in the field of network deformation analysis to address this problem (e.g. Mayer et al. (2000);

Niemeier et al. (2000)). One effective procedure is to relate all epoch measurements to ITRF. In order to demonstrate the local deformation behavior it can be helpful to select stations in the center of the deformation field, e.g. Fig. 7.80, Völksen (2000).

In areas of high risk (e.g. of earthquake, volcanic activities) like the San Andreas Fault in California, or in Japan, continuously monitoring GPS arrays have been installed (Bock et al., 1997). A fixed network of GPS receivers tracks all GPS satellites 24 hours a day. The data from all sites are transmitted via high-speed communication lines to the central facility, and are analyzed to obtain accurate "snapshots" of the relative positions of the network stations. Significant variations in these positions may indicate deformation caused by seismic

Figure 7.81. Deformation after the $M = 8.1$ Antofagasta Earthquake July 30, 1995, after Klotz et al. (1996)

or volcanic pre-event, co-event, or post-event activities.

Projects of *group (c)*, i.e. the monitoring of local deformation, belong in most cases to the field of deformation analysis in engineering surveying. Possible applications are the monitoring of

— land subsidence, e.g. in mining areas and oil fields,

— hang sliding, and

— local geotectonics.

In most cases the point distances are very small (about 1km), hence an accuracy of a few millimeters can be achieved, and very small deformations can be detected. Depending on the objectives of the control, and the expected rate of motion, the measurements have to be repeated after a given time period, for example days, weeks, or months. At least one stable reference station is required. In many cases, rapid methods can be applied [7.3.5]. In future, more and more continuously monitoring arrays will be built up. The data of the remote operating receivers have to be transmitted to the central station via cable, radio data link, or the internet.

A rather new and very promising field of GPS application in geodynamics is *Earth orientation monitoring*, in particular the variation of LOD and polar motion [2.1.2], [12.4.2]. Error analysis and comparison with other space techniques demonstrate the high potential of GPS to monitor daily and subdaily variations at the accuracy level of a few millimeters. Earth rotation monitoring, together with the delivery of precise orbits and station coordinates, is one of the major objectives of the *International Geodynamics GPS Service* (IGS) [7.4.3], [7.8.1].

7.6.2.3 Height Determination

GPS, like most other geodetic space techniques, yields three-dimensional coordinates X, Y, Z that can be transformed into ellipsoidal longitude, λ, latitude, φ, and height, h, see (2.37, p. 24). The vertical component is particularly sensitive to the geometrical configuration of the GPS satellites and to unmodeled errors in atmospheric refraction (cf. [7.4.4], e.g. Santerre (1991)). Simulation studies and experiences show that the error in the vertical component is about twice as high as in the horizontal components (Görres, 1996). Nevertheless, with improved receiver technology and careful data modeling, GPS is a powerful means for rapid and precise height determination. Considering the error budget, primarily height differences are of interest. For an elementary introduction into the GPS altimetry problem see also Schwarz, Sideris (1993).

It is obvious that the ellipsoidal height, h, coming from a GPS solution, is a purely geometric quantity. For most practical purposes, heights related to the gravity field rather than to the ellipsoid are required, namely *orthometric heights* or *normal heights*; for details on height systems see e.g. Torge (2001), also [2.1.5]. Note that orthometric heights are defined with respect to the geoid whereas normal heights refer to the quasigeoid. In the following no difference is made between orthometric and normal heights. The relation between

  h ellipsoidal height from GPS observations,
  H orthometric or normal height from spirit leveling, and
  N geoid height from a geoid computation,

can be written, according to Fig. 7.82, as

$$h_1 = N_1 + H_1; \quad h_2 = N_2 + H_2,$$

$$\Delta H = H_2 - H_1; \quad \Delta h = h_2 - h_1; \quad \Delta N = N_2 - N_1,$$

$$\Delta H = \Delta h - \Delta N, \quad \Delta N = \Delta h - \Delta H, \quad \Delta h = \Delta H - \Delta N. \tag{7.169}$$

If two types of information are known, the third one can be determined, namely

 − with precise geoidal heights the orthometric or normal heights can be derived from GPS, in order to control or to substitute spirit levelling, and
 − with precise levelling information and ellipsoidal heights from GPS, the geoid can be determined or controlled.
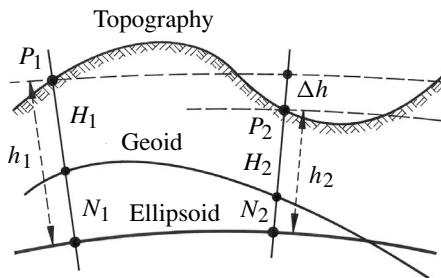
Figure 7.82. Relation between geoid height, N, orthometric height, H, and ellipsoidal height, h

If only height changes have to be analyzed the repeated determination of GPS heights without reference to the geoid is completely sufficient. Hence, three basic applications of GPS can be identified:

(a) height changes from repeated GPS control,

(b) transfer of orthometric or normal heights with known geoid, and

(c) determination of the geoid.

Height changes are mainly of interest for engineering purposes [7.6.2.6] or the detection of vertical crustal movements [7.6.2.2]. Particular engineering applications can be seen in the monitoring of subsidences in mining areas or for offshore drilling platforms. Vertical crustal movements are also of interest in connection with tide gauges to control sea level rise (Liebsch, 1996) [12.3.1].

The *determination of orthometric heights* (or normal heights) with GPS is a long-term goal in surveying and geodesy, in order to substitute time consuming and expensive spirit leveling. Fig. 7.82 and equation (7.169) demonstrate that we need two quantities for a height transfer from P_1 to P_2: the ellipsoidal height difference, Δh, determined with GPS, and the geoidal difference, ΔN, stemming from a geoid model.

In view of the high density of global and regional GPS reference networks [7.6.2.1] ellipsoidal height information is available at the centimeter-accuracy level over distances of several hundred kilometers or, in some areas, also several tens of kilometers. As a consequence, only ellipsoidal height differences over comparatively short distances have to be determined. This can be achieved with cm-accuracy by applying careful error modeling. For distances up to several tens of kilometers, even sub-centimeter accuracy can be obtained (Görres, Campbell, 1998). The main limitations for precise GPS altimetry obviously come from

  − modeling of the vertical error budget, and

  − the requirement for precise geoid information.

The *vertical error budget* is mainly governed by the tropospheric propagation delay and antenna phase center variations (PCV) [7.4]. PCV, delay and the height component are strongly correlated with each other. Absolute PCV can be determined through calibration. The tropospheric delay can be separated from the height component, when a tropospheric scale bias is estimated from several hours of observations, including measurements at low elevation angles (Rothacher et al., 1998). Nonetheless, the troposphere remains a critical factor when the highest accuracy in GPS altimetry is required (Kaniuth et al., 1998).

Local and regional geoid models reach the accuracy level of a few centimeters, as can be verified by comparison with GPS levelling. For Europe, the EGG97 currently gives the best solution (Denker, Torge, 1998). For the area of the United States the GEOID 96 (Smith, Milbert, 1999) gives a similar accuracy level. In general, however, for worldwide applications, the available geoid information is still not satisfying due to the lack of data. In order to exploit the potential of GPS for altimetry it is necessary to improve knowledge of the geoid. Here, the current and forthcoming gravity field missions CHAMP, GRACE, and GOCE [10] will contribute considerably, in particular for the geoid's long wavelength components, (see also [12.2]).

For the time being, local solutions and approximation techniques have to be applied, for example the use of mathematical interpolation algorithms between GPS stations with known leveled heights. In particular, for small areas with a good coverage of

control points, the method delivers satisfying results (Zhang, 2000). Very good results have also been obtained with the use of finite elements to represent a height reference surface (Jäger, Schneid, 2002).

Where heights in the gravity field are known from levelling lines it is possible to directly derive *geoid heights* from GPS results. This method can contribute considerably to the determination of a precise geoid. Other major problems to be solved with GPS altimetry are the connection of separated tide gauges, e.g. Kakkuri (1995), Liebsch (1996), and the establishment of a *global height datum*. This includes the determination of a precise marine geoid and of the *sea surface topography* [9.5.1]. In coastal areas, a precise geoid strongly supports the height determination for near-shore engineering and shore protection activities (Seeber et al., 1997b).

Very precise geoid profiles can be determined with a transportable digital zenith camera using the concept described in [5.2], in combination with GPS. The camera provides the direction of the plumbline in near real-time, and the GPS receiver generates geodetic coordinates as well as precise time. Using the technique of *astronomical levelling* (Torge, 2001), a high resolution geoid profile and orthometric heights are provided on-line (Hirt, 2001).

7.6.2.4   Cadastral Surveying, Geographic Information Systems

Because of the high accuracy in connection with short observation time, GPS can also be employed economically for *detailed surveying* in rural or urban environments. Main fields of applications are in connection with the installation or maintenance of multi-purpose cadaster or geographic information systems.

One major problem in detailed surveying is signal shadow caused by buildings, trees, towers, bridges etc. This is why the exclusive use of GPS in cadastral surveying will be restricted to open areas. With the presence of such obstructions, GPS will be mainly used to determine rapidly the standpoints for electronic tacheometers or other conventional surveying instruments. Fig. 7.83 illustrates the situation.

Figure 7.83. Combination of GPS with an electronic tacheometer

In areas of free sight, like most rural areas or urban areas with broad streets, low buildings, and low vegetation, rapid GPS methods can be used [7.3.5], in particular the RTK technique [7.3.5.4], [7.5.2]. Fig. 7.84 gives an artist's view of a detailed survey with GPS. The data can be stored in the moving receiver, or transmitted via a data link to the reference receiver, or vice versa.

With a continuously working data link, the setting-out of coordinates, or a re-identification of existing points or lost monuments will also be possible. The precise coordinates of the moving antenna are calculated in the field in real-time, and it is

indicated to the surveyor how far the antenna has to be moved to the final destination. Integrated systems of this type are available from most major GPS manufacturers.

The procedure depicted in Fig. 7.84 can be realized with a local, temporarily established reference station (case (a)), or with respect to a continuously operating reference station (case (b)).
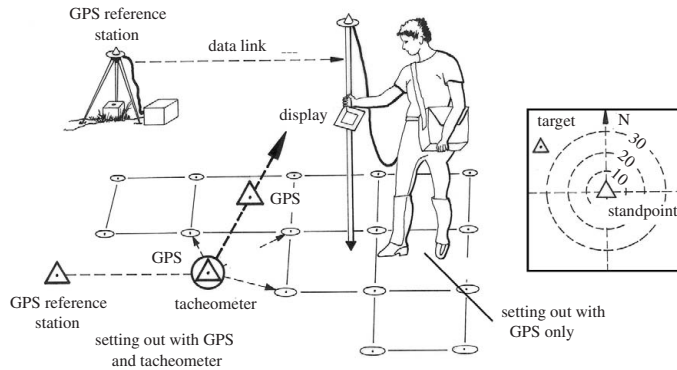
Figure 7.84. Use of GPS in real-time detailed surveying

Case (a) can be realized with conventional RTK equipment consisting of two GPS receivers and a radio. The reference receiver has to be installed on an existing demarcated surveying point, or the coordinates of the reference station have to be determined with respect to existing stations in the neighborhood. This can be realized when the roving receiver occupies two or three of such stations along with the survey. In modern surveying concepts it is no longer necessary to demarcate the temporary reference stations because the local field of surveying points is only represented by a strongly limited number of demarcated stations.

Case (b) has the advantage that only one GPS receiver is required in the field. Again, in most cases GPS will be used to establish standpoints for a tacheometer, whereas the object points (boundary marks or house-corners) are determined with conventional surveying tools. Another advantage is that all coordinates are immediately given with respect to the official reference frame and that no additional time is needed for the reconstruction of existing surveying marks. For high accuracy requirements it is necessary to work with networked reference stations [7.5.3.2]. For reliability purposes it is advised to occupy each object point twice.

GPS is a powerful means to support *Geographic Information Systems* (GIS). The role of GPS in this context is manyfold:
   − it contributes to a uniform basic geometric frame, for example a coordinate system, a digital map, or a digital terrain model,
   − it contributes to the geometric location of objects that enter the GIS, for example streets, buildings, power lines, proprietary boundaries,

  – it allows the GIS to be taken out into the field with GPS direct-entry, and

  – it forms an integrated building-block in a command and control system, for example for moving vehicles or machines that are navigating based on a digital terrain model.

In the following, only some examples are given. For all enterprises that provide services like energy, water supply or traffic information a geographic information system forms the basis of most decisions. As a first step, all spatially related data and object data have to be collected. Traditional maps are in many cases not sufficient. Here GPS provides an economic and efficient tool for an automatic data flow into the GIS. Vice versa, all objects that are selected in a GIS can be immediately identified in the field (e.g. Barrett, 1997). Integrated GIS - GPS concepts are offered by many manufacturers. The market is rapidly growing. Application examples are inventories for pipelines, power lines, fresh and waste water, streets, traffic signs, railway tracks, trees, contaminated locations, and so on.

Depending on accuracy requirements, GPS provides continuous position information at all scales of interest. In some cases, the accuracy of a single receiver (5 to 15 m) is sufficient. In most cases, ordinary DGPS will be applied (0.5 to 2 m). If highest accuracy is required (few centimeters), the services of multiple reference stations can be used [7.5.3], or even established for the purpose. Another advantage is that 3 D information is available. In connection with a digital geoid, gravity field related height information (e.g. orthometric heights) can be supplied for applications involving the direction of water flow.

Rapid digital data acquisition is possible with a *car driven survey system* for mobile mapping (Fig. 7.85). The positioning problem is solved by GPS in connection with an inertial sensor, or alternatively, wheel sensors, barometer and magnetic sensors. The data are acquired and analyzed automatically with several video cameras (Benning, Aussems, 1998; El-Sheimy, 2000).

Another fast growing field of application is *precision farming*. Based on a GIS including the topography, soil quality, and actual state data, all steps in

Figure 7.85. Car driven survey system

farming can be performed in an optimized way, like fine-tuned fertilization or spraying of infested areas. Computerized controllers and GPS-guided navigation form an optional part of farming equipment like sprayers or harvesters. Table 7.22 gives an overview of accuracy requirements (Demmel, 2000).

Many more examples could be given. The integration of GPS and GIS together with a communication link is increasing and widely discussed in the GPS literature (e.g. *GPS World*), as well as in the general surveying and GIS literature.

Table 7.22. Accuracy requirements for the use of GPS in precision farming

Task	Example	Required accuracy
Navigation	Search working area Search deposit place	± 10 m
Execution of work Information Documentation	Work in the field with – determination of returns – fertilization – plant protection – soil samples Automatic data recording	± 1 m
Vehicle guidance	Connected tracks Harvester-thresher	± 10 cm
Equipment guidance	Mechanical weed removal	± 1 cm

7.6.2.5   Fleet Management, Telematics, Location Based Services

These services present important new challenges, with a focus on real-time positioning, communication and information. They are mainly related to motorized vehicles like cars but also may concern pedestrians. The denomination is not yet clearly defined; all three terms are sometimes used for the same service. *Fleet management* means the control of a large quantity of vehicles like trucks, trains, police- and emergency cars, public buses and so on. *Telematics* is a new word composed from "telecommunication" and "informatics" and means the use of traffic-related information. *Location Based Services* (LBS) are mainly related to the use of cellular phones and mean the real-time availability of all kinds of position-related information to individual customers. The backbone of all services is composed of these elements:

 – knowledge of the position of the client,
 – knowledge of the position of other participants in the system, if required,
 – a geographical information system,
 – a personal digital assistant (PDA, palmtop) with the client, or a computer in the control center, and
 – a communication link.

The positions can either be provided by GPS (GNSS), or another positioning device like the cellular phone identification. The GIS is either available in the PDA of the client, or via cellular phone from a provider. Considering the rapid development of communication technology, and the high number of vehicles in industrial countries, the market promises to develop fast. Some examples follow.

Car navigation systems for individual users, based on a digital map and a location service, are already well established. The inclusion of information on congestion, snowfall, or roadwork, for instance, will improve the service. Additional features are automatic location transmission in case of emergency, or theft.

Fleet management is essential for shipping agencies, train and bus systems, police and emergency services, and fire brigades. In connection with a traffic management system, traffic light priority can be given to public transportation and emergency vehicles.

At large construction sites a logistic system can be installed to organize and guide the construction vehicle traffic. Each vehicle gets a certain time slot when entering the site, and a GPS based local navigation equipment is deployed in each car as long as it operates inside the construction area.

A particular application is the mobility of blind people. A precise DGPS system and a precise and detailed specific digital map, connected to a voice generator enables a user to navigate in an unknown environment aided perhaps only by a stick.

A large potential market is developing for location based services. Tourists can request information on nearby touristic highlights, restaurants and public transportation. Parents can supervise their children, and persons with a critical health status can be remotely monitored by a medical center.

A particular application will be the automatic location of a mobile phone in connection with the emergency calls E-911 in the U.S., or E-112 in Europe. A further step will be a combination of outdoor and indoor navigation within a single hybrid location device.

7.6.2.6   Engineering and Monitoring

Almost unlimited possible uses and applications may be conceived in this field. The corresponding observation and evaluation methods are as discussed in the previous sections. Since the distances are usually small it is possible to achieve mm accuracy with routine methods. Rapid methods [7.3.5], real-time solutions, and integration with electronic tacheometers may be required. Some fields of application are:

(1) *Determination of geodetic control points*
  − Geographic Information Systems (GIS),
  − cartography,
  − photogrammetry,
  − geophysical surveys,
  − inertial surveys,
  − antenna location in hydrographic surveying,
  − expeditions of all kinds, and
  − archaeological mapping,

(2) *Monitoring object movements by repeated or continuous measurements*
  − ground subsidence (mining, ground water withdrawal),
  − land slides,
  − construction of dams,
  − subsidence of offshore structures, and
  − settlement of buildings,

(3) *Setting out local networks for the control of engineering projects*
  − tunnel construction,

− particle accelerators,
− bridge construction,
− road construction,
− pipelines, and
− waterways,

(4) *Real-time guidance and control of vehicles*
− construction vehicles,
− large excavators in opencast mining, and
− forklifts in open storage areas (e.g. container yards).

If two antennas (and receivers units) are used, GPS can also be employed as a method of determining directions. Usually the direction is derived from the coordinates of the two antenna phase center positions, hence precise carrier phase resolution and carefully designed and calibrated antennas are required. Table 7.23 shows the relation between station spacing, azimuth accuracy and required GPS relative position accuracy. If 2 mm relative position accuracy is considered to be the accuracy limit, it is possible to determine a 1 arcsecond azimuth over 400 m distance. This may be of interest for setting out a tunnel axis.

Table 7.23. Azimuth reference control with GPS

station spacing (m)	azimuth accuracy in seconds of arc				
	1	2	4	6	10
	GPS relative position accuracy in mm				
100	–	1	2	3	5
200	1	2	4	5	10
300	2	3	6	9	14
400	2	4	8	12	19
500	3	5	10	14	24
600	3	6	12	18	29

For operational use, a much shorter baseline can be selected. With a 1 m antenna separation a directional accuracy of a few arcminutes can be achieved, even in kinematic mode. GPS can hence be used for compassing. With three antennas the attitude of a moving platform can be controlled.

From the above list of possible applications two examples are given, a control network for tunnel construction and a network for dam control. The advantage of GPS can, in particular, be demonstrated for the *tunnel network*. The main purpose of such a network is the setting-out of the bearing of the shaft center line at both entrances, P_W (portal west), and P_E (portal east), cf. Fig. 7.86. In classical engineering, both portals had to be connected via a precise network, covering the whole area. This could be an extremely difficult task in mountainous or heavily forested areas. With GPS it is sufficient to determine two control points each at both entrances for setting out the

bearing of the center line. For security reasons, it is advisable to establish a second target pillar at each portal for reference bearings. The distance should not be too large to enable sights under unfavorable atmospheric conditions.
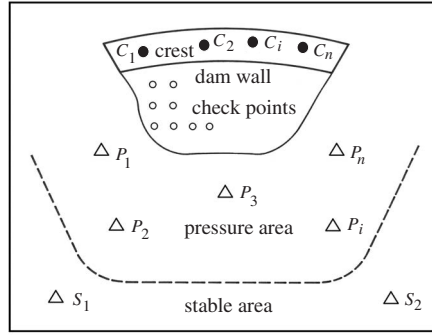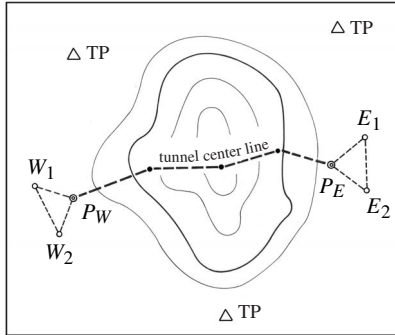
Figure 7.86. Generic tunnel network with GPS

Figure 7.87. Dam control with GPS

The relative location of the two *portal networks* can be determined with an accuracy level below 1 cm for distances up to 10 km. The results are given in three dimensions. In order to provide "levelled heights" via GPS, it is necessary to include a precise local geoid [7.6.2.3]. If required, the tunnel network can be easily connected with the nearest control points (TP) of the geodetic network via GPS techniques.

The second example refers to the permanent control of a dam, during construction and after completion (Fig. 7.87). A difficult task is the selection of stable control points, and the delimitation of the pressure zone from the stable area. Usually, the advice and support of experts is required. One advantage of GPS is that the stable control points S_i can be placed well away from the influence zone, and that no direct sight connection to the near-construction control points, P_i, is required. GPS can be used to

   − establish stable control points (S_i),
   − establish and monitor control points in the pressure zone (P_i), and
   − establish control points on the dam crest (C_i).

Checkpoints attached to the dam wall remain to be controlled by other techniques, either with electronic tacheometers or photogrammetry. GPS is suited to determining and monitoring the coordinates of the tacheometer or camera standpoints, P_i, in the pressure zone. GPS is also capable of identifying and analyzing point motion within the pressure zone. Regarding the high potential of GPS concerning accuracy and cost-effectiveness, it is possible to install a dense network of control points in the potential pressure zone.

Deformations can be derived from repeated observation, in intervals of days, weeks, or months, depending on the situation. In cases where there is suspicion of impending

structural distress, the establishment of a continuous monitoring array can be taken into consideration.

One critical factor is the limited visibility of satellites from stations near the dam wall. The situation will improve with the inclusion of other GNSS like GLONASS and GALILEO [7.7], but drawbacks result from unbalanced geometry and multipath effects. Instead, control points near the dam wall can be related to better placed control points by tacheometry.

GPS can also be used for deformation monitoring at the one millimeter or sub-millimeter level, when all acting error sources are eliminated or considerably reduced. The most critical part, multipath, can be eliminated by forming sidereal differences [7.4.4.3], because the satellite geometry repeats after 24 hours in sidereal time. The technique has been successfully applied for the monitoring of deformation during the filling process of locks (Seeber et al., 1997a; Wübbena et al., 2001a).

7.6.2.7   Precise Marine Navigation, Marine Geodesy, and Hydrography

Because of the real-time capability, continuous availability, and the high accuracy potential, this field of use is very broad, continuously growing, and is developing fast. In this chapter only a short overview is given. For more information see [12.3] [7.5.1] and the ample literature in symposia proceedings like *INSMAP 94, INSMAP 98* or journals like *Navigation, GPS World, Sea Technology*.

The possible applications, and the related accuracy requirements, can be divided into three user groups:

   (a) low accuracy requirements, about 10 to 100 m in position, and 1m/s in velocity,
   (b) medium accuracy requirements, about 1 to 10 m in position, and 0.1 m/s in velocity, and
   (c) high accuracy requirements, better than 0.1 m in position and height, and 0.01 m/s in velocity.

User inquiries indicate that highest interest is in the group (b), i.e. a position requirement of a few meters.

User group (a) can be fully satisfied with a single C/A-code navigation receiver aboard a ship. GPS will provide continuous two-dimensional position accuracy of about 10 to 30 m, or better, under the Standard Positioning Service [7.4.1]. Important areas of employment in user group (a) are, for example (cf. [12.3]):

   (1) general navigation tasks on the high seas,
   (2) research in oceanography,
   (3) ship's positioning in small scale bathymetry with swath systems, and
   (4) position and velocity in small scale gravimetric, magnetic, and seismic measurements.

For some applications of tasks (3) and (4), the accuracy of a single operating receiver is not sufficient. In these cases, and for the majority of applications *(user group (b))* in marine geodesy, hydrography, and precise navigation, GPS must be operated in the relative mode (*Differential GPS*, see [7.5.1]).

Typical fields of application in user group (b) are, for example (cf. [12.3]):

(1) precise navigation in coastal waters,
(2) harbor approach,
(3) sea floor mapping in the *Exclusive Economic Zone* (EEZ), for the delimitation of seaward boundaries and/or for scientific purposes (cf. Fig. 12.10, p. 524),
(4) hydrography,
(5) precise gravimetric and seismic surveys,
(7) positioning of underwater sensors and samplers in marine prospecting for mineral resources, and
(6) calibration of transponder arrays.

In cases where the data are not required in real-time, the final positions can be computed afterwards (post-mission) in a post-processing step. However, considering the huge amount of data it is advisable to determine the ship's position in real-time and not to store the original raw data.

A further option of the differential mode is to use the carrier phase data at the remote station to smooth the code phase observations [7.3.6] with an appropriate filter algorithm (7.108, p. 296). This method works on a routine basis if an appropriate receiver is used, and provides a continuous accuracy of 2–3 m for the moving antenna, or even better. The accuracy level satisfies most users of the above list, in particular in hydrography and precise surveying activities.

An increasing user market requires an accuracy level of better than 0.1 m, in particular in the height component (*user group (c)*). In this case, the carrier phase observable has to be used as the primary quantity, and the ambiguities have to be resolved. The *pure kinematic method* [7.3.5.4] with ambiguity resolution techniques "on the way" [7.3.2.3] has to be applied. The methods work well with postprocessing and also in real-time if a data link of sufficient capacity is available [7.5.1.2]. For larger areas, the concept of multiple reference stations can be applied to model the distance dependent errors [7.5.3].

Possible applications in user group (c) are:
(1) precise hydrographic surveying,
(2) monitoring silt accretion and erosion in rivers, lakes, estuaries, coastal waters, and harbor areas,
(3) real-time dredge guidance and control,
(4) support of coastal engineering,
(5) marine geodynamics.

Two further particular applications are:
(6) precise continuous height control, and
(7) attitude control of ships, buoys, floating platforms.

For precise echo-sounding and sea level monitoring a continuous height determination with an accuracy of a few centimeters is required and feasible (Goldan, 1996; Goffinet, 2000; Böder, 2002).

The actual sea level at the location of the surveying vessel must be referred to the height reference onshore (depth reduction). The conventional method is to estimate the *depth reduction*, *dh*, from tidal and hydrodynamic models with respect to tide gauges onshore. With GPS, the reduction can be determined directly (Fig. 7.88). The GPS
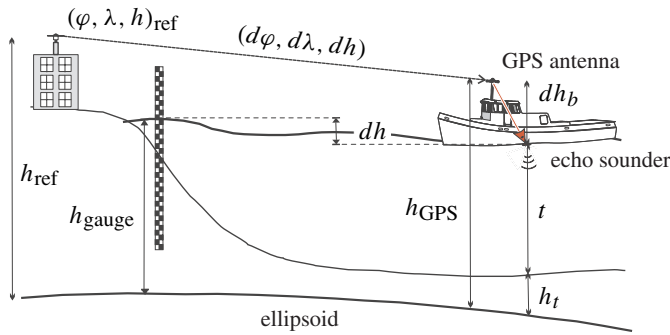
Figure 7.88. Depth reduction for echo-sounding; conventional and with GPS

antenna phase center does not coincide with the reference point of the echo-sounder (Fig. 7.89). The horizontal and vertical corrections are given by

$$dX = X - \sin(\beta + \gamma) \cdot S \tag{7.170}$$
$$dZ = Z - \cos(\beta + \gamma) \cdot S. \tag{7.171}$$

To minimize the effect of ship inclination on the depth correction, dZ, it is recommended to install the GPS antenna directly above the sounder ($\beta = 0$).

GPS onboard an anchored ship or a moored buoy can also be used to monitor tidal variation. The resolution is a few centimeters, depending on the size and behavior of the platform (a larger platform shows smaller noise), Goldan (1996). A challenging application is continuous height control in calibration areas for altimeter satellites [9.3.3], see Fig. 9.10

With three antennas/receivers on board a ship (Fig. 7.90) the time-dependent spatial behavior of the platform, its *attitude*, can be monitored in real-time (Seeber, Böder, 1998). The achievable accuracy depends on the baseline length between the antennas, and the noise in the GPS result (Table 7.24).
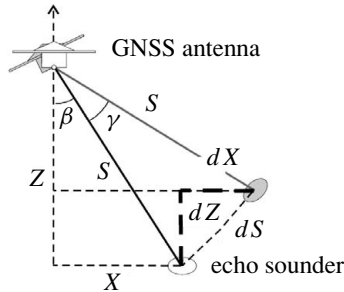
Figure 7.89. Inclination correction in echo-sounding
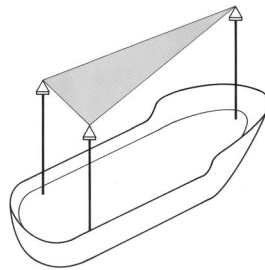
Figure 7.90. Attitude control with three GPS antennas

Table 7.24. Relationship between a height error, dx, baseline length, s, and GPS derived orientation accuracy

s \ dx	3 mm	1 cm	0.1 m
1 m	0.17°	0.57°	5.71°
5 m	0.03°	0.11°	1.15°
10 m	0.02°	0.06°	0.57°
30 m	0.006°	0.02°	0.19°

In general, a resolution of 0.1° is sufficient, for example for the correction in (7.171). Attitude control is of particular importance for the inclination correction of multibeam sonar systems in sea-bottom mapping ([12.3], Fig. 12.10, p. 524), and for the monitoring of floating GPS sensors at the sea surface in the precise location of submarine geodetic control markers (cf. [12.3.2], Fig 12.12, p. 526). For the mathematics of attitude determination see Kleusberg (1995); Cohen (1996).

Note that most developments in precise marine navigation with GPS can easily be applied in land navigation and remote vehicle control.

7.6.2.8 Photogrammetry, Remote Sensing, Airborne GPS

The use of GPS contributes in several different ways, for example:
- (a) determination of ground control points in photogrammetry,
- (b) navigation of sensor carrying airplanes, and
- (c) determination of sensor platform coordinates and orientation.

The determination of *ground control points* (group a) for photogrammetric map production corresponds completely to the procedures discussed in [7.6.2.1]. The technique and effort required depend on the desired map scale. For cadastral purposes, centimeter accuracy can be achieved with carrier phase adjustment. Usually the photogrammetric products have to be related to the official reference frame via at least one control point with known coordinates [12.1].

The accuracy requirements are much less for control points and ground truthing in *satellite images* (e.g. SPOT, LANDSAT). The level of 1 to 5 m can be achieved by differential techniques using code or carrier-smoothed code observations only, without resolving ambiguities [7.5.1]. The remote receiver can be operated over distances up to several hundred kilometers. It is sufficient to collect only a few minutes of data on the site.

For the *precise navigation* (group b) of a survey aircraft the differential mode and a real-time data link are required. Usually the transmission of range corrections [7.5.1.1] is sufficient, to assure an accuracy of several meters, as long as at least four satellites are visible. Conventional DGPS services are well suited to the task.

The most promising contribution of GPS to photogrammetry is the determination of the *sensor orientation*, in particular the *precise camera position* (group c) in order

to support *aerial triangulation* (Li, 1992; Lee, 1996; Schmitz, 1998), Fig. 7.91. GPS determined camera positions are introduced as precise observations into the combined block adjustment. As a consequence the required number of ground control points can be reduced to about 10 percent, or even less, of those required in conventional aerotriangulation (Jacobsen, 1997, 2000).

In order to achieve the required accuracy level of about ±5 cm it is necessary to

— operate in the differential mode,
— use code and carrier phase data, and
— resolve the phase ambiguities.

Because of the cycle slip problem, in particular in the turns between individual survey strips, it is necessary to use ambiguity resolution techniques "on the fly" [7.3.2.3]. Receivers that provide sufficient channels for all satellites, both frequencies, and low noise code observations are particularly suitable.

Figure 7.91. The use of GPS for camera positioning in aero-triangulation

The following problems or aspects have to be considered:
— simultaneity of receiver and camera operation,
— eccentricity between antenna phase center and camera projection center, and
— loss of satellite track or cycle slips in turns.

Modern GPS receivers and aerial cameras allow nearly *synchronous operation*. It is usually not possible for a GPS receiver to measure at arbitrary epochs, hence the camera shutter has to be triggered by an output signal from the receiver. For older aerial cameras it is advisable to operate the shutter manually, or by some external device, as near as possible to the GPS observation epoch, and to register the mid-open time of the shutter. Considering the average speed of a photogrammetric aircraft, asynchronous operation may introduce errors of up to several meters.

Another possibility is to interpolate the aircraft positions between the GPS positions with an *inertial platform* (INS). The integrated techniques of GPS and INS provide an accuracy of a few centimeters (Lee, 1996; Cramer, 2001).

The *3-D eccentricity* between the GPS antenna and the camera projection center (Fig. 7.92) includes the distance and the three orientation angles. The distance is invariable and has to be measured by conventional means. The orientation can be determined

— by a GPS based platform orientation unit with three GPS antennas (cf. [7.6.2.7]),
— as a by-product of an inertial package onboard, or
— with inclinometers.

If the camera is fixed to the aircraft body, the orientation angles serve at the same time for the eccentricity calculation and for the outer orientation of the photogrammetric camera.

Detailed investigations demonstrate that external information on the angles of orientation is of much less importance than the precise coordinates of the focal center. The orientation angles can be easily determined along with the bundle block adjustment (Jacobsen, 1992).

One particular problem is *signal loss* from some of the satellites, through the inclination of the aircraft whilst banking. If less than four satellites remain visible, a new cycle ambiguity has to be determined. Either powerful ambiguity resolution techniques *'on the fly'* have to be applied [7.3.2.3], or the data gap has to be bridged by inertial techniques. With the current constellation of more than 24 GPS satellites, signal loss is no longer a serious problem, because ambiguities can be easily fixed.

Because of the many restrictions in photogrammetric survey planning (weather conditions, vegetation period, etc.) the routine use of GPS requires the

Figure 7.92.   Eccentricity elements between camera, focal center, and GPS antenna (Li, 1992)

continuous availability of reference observations, for example from continuously operating reference stations [7.5.3]. Usually, post-processing techniques are applicable. Real-time results are also possible in areas with networked multiple reference stations like SAPOS in Germany [7.5.3.2].

The methods of precise platform positioning with GPS can also be used for related applications such as *laser bottom profiling* or *airborne gravimetry* (Schwarz et al., 1997).

7.6.2.9   Special Applications of GPS

As stated above, the possible applications of GPS in the field of engineering and geoscience are unlimited. Some further examples are
 − glacial geodesy,
 − time transfer,
 − GPS carrying satellites, and

– GPS meteorology.

In *glacial geodesy*, and *Antarctic research*, GPS can be employed successfully to determine the movement of glaciers or ice sheets (Hinze, 1990). If motion parameters (velocity and azimuth) are to be derived from repeated measurements over the years, a quasi-online reading suffices, while a route is traversed with snow mobiles, or a helicopter lands for a short time. In the relative mode, to a fixed station, sub-decimeter accuracy can be attained with short observation times, depending on the distance to the reference station [7.5.1], so that correct results can be expected from repeated measurements after about 1 month in the same season. Fig. 7.93 gives an example from the Ekström Ice Shelf near the German Antarctic station "Georg v. Neumeyer".
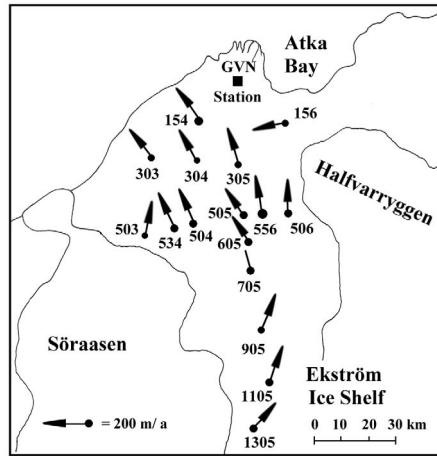
Figure 7.93. Ice flow from GPS observations

GPS is one of the most efficient means for operational *global clock synchronization*. Table 7.25 gives an overview of the achievable accuracy (Lombardi et al., 2001). In most cases, the so-called *common-view technique* is applied. The time of arrival of the same signal from one satellite is observed at two stations, and compared with the local reference clocks. Afterwards the data are exchanged. The signal travel time between the satellite and the station has to be calculated, based on precise coordinates for both stations and precise satellite orbits. *Single channel* technique means that the measurements follow a pre-determined tracking schedule. In the *multi-channel* common-view technique, data from all satellites in view are recorded without a schedule. The latter mode facilitates continuous comparison of standards with no gaps in the data.

Table 7.25. Accuracy of GPS time transfer

Technique	Timing Uncertainty 24 hours, 2 σ	Frequency Uncertainty 24 hours, 2 σ
One-Way	< 20 ns	< 2 × 10^{-13}
Single-Channel, Common-View	∼ 10 ns	∼ 1 × 10^{-13}
Multi-Channel, Common-View	< 5 ns	< 5 × 10^{-14}
Carrier-Phase, Common-View	< 500 ps	< 5 × 10^{-15}

A station position error of 3 cm enters 100 picoseconds into the error budget. The effect of orbital errors follows the rule of thumb (7.134), hence an orbit accuracy of

0.1 m is required for 100 picoseconds time transfer over 5000 km. The *International GPS Service* (IGS) [7.8.1] considerably supports operational high precision global time transfer through its station network and products. A number of the stations are connected with external oscillators like H-masers, cesium and rubidium clocks (IGS, 2000)

Several manufacturers offer dedicated GPS receivers for time-transfer. For the nanosecond accuracy level all error influences including hardware delays have to be carefully modeled. Real-time relative time transfer at the 100 picoseconds level has been demonstrated within the "Internet-Based Global Differential GPS" project of the NASA-JPL (Powers, et al., 2002), see also [7.5.1.2]. For a topical treatment of GPS time transfer see for example Schildknecht, Dudle (2000).

Very powerful GPS Earth Science applications result from the deployment of GPS sensors on *near Earth orbiting satellites*, so-called LEOs [3.4.2]. The GPS data received at the orbiting platform may serve for

 — precise orbit determination of remote sensing satellites, primarily altimeter satellites [9],

 — precise position and orbit determination of satellites probing Earth's gravity field [10],

 — attitude control of space vehicles, and

 — analysis of GPS signals after passing the atmosphere (GPS-MET).

One of the first demonstrations for precise orbit determination with GPS (see [3.3.2.3]) was with the TOPEX/POSEIDON mission (Melbourne et al., 1994a). Since then, GPS receivers have been included in a number of missions, in particular on LEO satellites like CHAMP, GRACE, JASON-1, and ICESAT. Precise orbit determination (POD) [3.3.2.3] is supported by the orbit and clock products of the IGS. The accuracy level is in the order of a few decimeters and may reach sub-decimeter after tailored gravity field improvement (Wickert et al., 2001). On the other hand, LEO data are of interest to IGS. The IGS has started a pilot project to study the inclusion of LEO data into the regular IGS products (IGS, 2000).

If the satellite carries three or more GPS antennas it is possible to determine its attitude. Since the baseline between the antennas is always very small, and only the carrier phase difference is required, single frequency C/A-code receivers can be used. For details of the technique, see for example (Purivigraipong, Unwin, 2001).

GPS contributes with two different techniques to the improvement of global weather data. The continuous observations at more than 200 IGS sites are used to model the total zenith delay at a level of 3 to 5 mm, that corresponds to better than 1 millimeter in water vapor (IGS, 2000). The data are available as a regular IGS product and can be used by meteorological institutions in numerical weather prediction models. For details of the subject see also Schüler (2001). For very dense networks of monitor stations, for example in Germany, with a spacing of about 50 km the accuracy of the integrated water vapor was found to be 1 to 2 mm, with a delay of only 40 minutes (Reigber et al., 2002).

The second technique uses the observations made between GPS satellites and LEOs equipped with GPS receivers. Fig. 7.94 demonstrates how GPS contributes to atmospheric research.
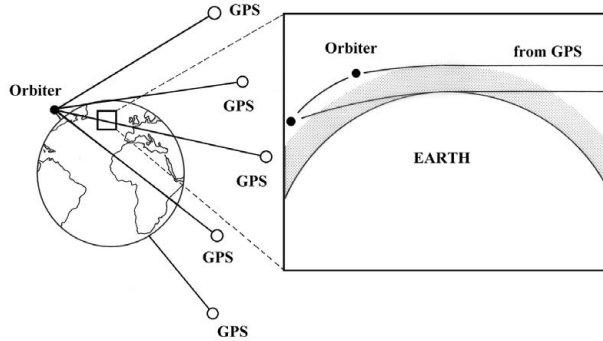
Figure 7.94. The use of GPS in atmospheric research (after Yunck, Melbourne (1990))

Due to the relative motion between the LEO satellite, and a GPS satellite setting behind Earth's disk, the tangential point of the radio link between the two space vehicles moves downward with a geocentric velocity of about 2.5 to 3 kilometers per second (Hocke, Tsuda, 2001) and scans the atmospheric layers from the high atmosphere down to Earth's surface. The signals are affected both by the ionosphere and the troposphere and can be used for ionospheric tomography as well as for mapping the integrated water vapor. The technique is known as *radio occultation* or *limb sounding*.

A first experiment (GPS/MET) was carried out with the launch of the MICROLAB satellite in 1995 (Hocke, Tsuda, 2001). Other suitable satellites for radio-occultations are ÖRSTED (launch January 1999), CHAMP and GRACE. For details on the technique see e.g. Kleusberg (1998). First results from CHAMP radio occultations are reported in Wickert et al. (2001). CHAMP measures at a rate of 50Hz and provides about 230 globally distributed vertical profiles of atmospheric parameters per day.

7.7   GNSS – Global Navigation Satellite System

GPS is not the only satellite-based navigation system. The Russian Federation is operating GLONASS, and the European Union together with the ESA is planning GALILEO. In addition, various augmentations to GPS are under preparation. The general name given to these systems is *Global Navigation Satellite System* (GNSS). Most of the material, outlined in chapter 7, is also valid for other GNSS systems. This is why, in many publications, instead of GPS the more general term GNSS is used.

The term GNSS was coined at the 10th Air Navigation Conference in 1991, when the *International Civil Aviation Organization* (ICAO) recognized that the "primary

stand-alone navigation system in the 21st century will be provided by the Global Navigation Satellite System (GNSS)" (Hein, 2000). As commonly understood GNSS includes more than just satellite-based positioning. Important features besides accuracy are *integrity*, *availability*, and *continuity of service*. GPS and GLONASS, being primarily military systems, do not guarantee these capabilities. On the way to establish GNSS, several steps have been defined.

GNSS-1 is based on GPS and/or GLONASS as backbone, and is augmented by additional civil components.

GNSS-2 is a second-generation satellite navigation system which fulfills the above requirements, such as GPS IIF or the European GALILEO.

In the following, some of the particular features of GNSS developments are explained.

7.7.1 GLONASS

The former Soviet Union (SU) has, since the 1970s, been developing a navigation system very similar in design to GPS under the name GLONASS (*GLO*bal *NA*vigation *S*atellite *S*ystem). The Russian denomination is *Global' naya Navigatsionnaya Sputnikowaya Sistema*. Today, GLONASS is continued by the Russian Federation. Like GPS, GLONASS is a military system, but it has been offered to civil use by several declarations of the Russian Federal Government (Slater et al., 1999). The system was officially put into operation on September 24, 1993 as a first-stage constellation of twelve satellites. By the end of 1995 the constellation was expanded to 24 satellites (standard constellation). Due to a lack of new launches, the constellation has since then decreased considerably. By the end of 2002 only 7 satellites were operational.

Similar to GPS with SPS and PPS [7.1.6], [7.4.1], GLONASS provides a *standard precision* (SP) and a *high precision* (HP) navigation signal. The SP signal is continuously available to all civil users world-wide. The specification for SP accuracy is 50 to 70 m horizontally and 70 m in height. Information for civil users is available via the *Coordinational Scientific Information Center* (CSIC) of the Russian Space Forces.

In this section, some basic information on GLONASS is given. For further reading with additional references see e.g. Kaplan (1996, chap. 10), Daly, Misra (1996), Habrich (2000), Roßbach (2001). A short introduction is given by Langley (1997a). Table 7.26 compares GLONASS with GPS and indicates similarities and differences. It is evident that the systems have strong similarities. The main characteristics and differences are as follows.

(a) *Satellite orbits*
In the baseline constellation, both systems consist of twenty-four satellites including three spares. Unlike GPS, the GLONASS satellites are arranged in 3 orbital planes 110° apart. Each orbital plane contains eight equally spaced satellites. Fig. 7.95 shows the complete configuration.

The ground tracks repeat every seventeen orbits, or eight sidereal days. The orbits are arranged in such a way that resonance phenomena do not occur and that in one

8-day period all satellite footprints pass through a given position. As a consequence, for an observer on Earth, a particular satellite will pass the same point in the sky after eight sidereal days, and one of the satellites in each orbital plane will appear at the same location in the sky at the same sidereal time each day (Forsell, 1991; Kaplan, 1996).

Both systems provide similar coverage for the 24 satellite constellation. Between 6 and 11 satellites are visible at any place on the Earth for either system, if fully deployed. Hence, both systems together can generate a coverage of about 12 to 16 satellites simultaneously for a given spot on Earth. Due to the differing inclination angle, the position of the "shadow area" (see [7.6.1.1],

Figure 7.95. GLONASS satellite configuration

Fig. 7.69, p. 347) is also slightly different and leads to a better overall coverage of the sky.

(b) *Satellite navigation signals*
GLONASS is, like GPS, a one-way ranging system. The radio-signal structure of both systems is very similar. Two carrier signals in the L-band are broadcast, and the signals are modulated by two binary codes and the message. In contrast to GPS, all GLONASS satellites transmit carrier signals at different frequencies. The L1 frequencies are

$$f_{L1} = f_0 + k \cdot \Delta f_{L1} \qquad k = 0, 1, 2, \ldots, 24, \qquad (7.172)$$

where $f_0 = 1.602$ MHz, and $\Delta f_{L1} = 0.5625$ MHz. L1 and L2 are related by

$$f_{L1}/f_{L2} = 9/7. \qquad (7.173)$$

Here, k means the frequency number of the satellite. Number 0 is designated as "technical frequency" and reserved for testing purposes. The procedure is called *frequency division multiple access* (FDMA). Satellites need not be distinguished by their unique satellite modulation, hence all GLONASS satellites use the same codes. The code frequencies are about half the corresponding GPS values, hence the range resolution may be slightly lower for GLONASS. The frequency ranges of both systems are close, thus permitting the use of a combined antenna and common input amplifiers in the user equipment, and allowing the development of combined receivers. The signal processing, however, is different (Roßbach, 2001).

Originally, each of the 24 GLONASS satellites should have transmitted on its own unique frequency. Since some of the GLONASS transmissions cause interference to radio astronomy in the frequency bands 1610.6–1613.8 MHz and 1660–1670 MHz, as well as to some communication satellites, it was decided to shift the GLONASS

Table 7.26. Comparison of GLONASS and GPS

Parameter	GLONASS	NAVSTAR GPS
Satellites		
Number of satellites in the baseline constellation	21 + 3 spares	21 + 3 spares
Number of orbital planes	3	6
Inclination	64.8°	55°
Orbital altitude	19 100 km	20 180 km
Orbital radius	25 510 km	26 560 km
Orbital period (sidereal time)	11 hours 15 min.	12 hours
Repeat ground tracks	every sidereal day	every 8 sidereal days
Navigation message		
Ephemeris representation	9 parameters (position, velocity, acceleration) in the ECEF Cartesian system	Keplerian elements and interpolation coefficients
Geodetic datum	PZ-90	WGS 84
Time base	GLONASS system time	GPS system time
Related system time	UTC$_{[SU]}$	UTC$_{[USNO]}$
Almanac transmission	2.5 minutes	12.5 minutes
Signals		
Satellite signal division	Frequency division	Code division
Frequency band L1	1.602–1.615 MHz	1.575 MHz
Frequency band L2	1.246–1.256 MHz	1.228 MHz
Codes	same for all satellites C/A-code on L1 P-code on L1, L2	different for all satellites C/A-code on L1 P-code on L1, L2
Code type	PRN sequence	Gold code
Code frequency C/A-code	0.511 MHz	1.023 MHz
Code frequency P-code	5.11 MHz	10.23 MHz
Clock data	clock offset frequency offset	clock offset frequency offset and rate

frequencies to a somewhat lower domain. In addition, the number of frequency channels is cut in half; so-called "antipodal satellites", i.e. satellites in the same orbital plane separated by 180 degrees in the argument of latitude, share the same channel. The re-organization of the frequency plan occurs in several steps. From 1998 to 2005 the frequency numbers $k = -7, \ldots, 12$ are applied, and after 2005 the numbers $k = -7, \ldots, 4$ (5, 6 for testing only). The bands will, hence, be finally shifted to 1598.0625–1604.25 MHz for L1 and 1242.9375–1247.75 MHz for L2.

(c) *Navigation message*

The navigation data are bi-phase modulated onto the carrier at 50 bits/s. The binary sequence has a total length of 2 seconds, called one line. The digital data structure is formed by navigation superframes of 2.5 minutes in length. A superframe consists of five frames of 30 seconds each, and each frame consists of fifteen lines (subframes).

As with GPS, the GLONASS message contains precise orbital information (ephemeris data) about the individual satellite's own position and status, and less precise "almanac" information about other satellite positions. Lines 1–4 of a frame contain the ephemeris data of the transmitting satellite, and line 5 general information concerning the entire system. Lines 6–15 contain the almanac data for five satellites. The almanac data of the complete system hence require one superframe, corresponding to 2.5 minutes.

Details of the data format can be found in the official *GLONASS Interface Control Document* (ICD-GLONASS) or in the literature cited above. The navigation message contains, for example

- coordinates for the i-th satellite in the ECEF reference frame for the reference time,
- speed vector components for the i-th satellite,
- acceleration vector components caused by Earth and Moon gravity,
- time scale correction to the GLONASS time scale for the i-th satellite, and
- satellite identification number, status information, reference time.

The GLONASS broadcast ephemerides are updated every 30 minutes and refer to the center of the 30 minutes time interval. For a measurement epoch in between these half-hour marks, the satellite position has to be interpolated using the position, velocity and acceleration data from the reference epochs before and after the observation epoch. These data are used as initial values for an integration of the equation of motion [3.3], (e.g. Roßbach, 2001).

(d) *Control Segment*

The ground-based control segment is responsible for (Kaplan, 1996)

- prediction of satellite ephemerides,
- uploading of ephemeris, clock correction and almanac data into each satellite,
- synchronization of the satellite clocks with GLONASS system time,
- estimation of the offset between GLONASS system time and UTC (SU), and
- spacecraft control.

The ground segment consists of

- the System Control Center,
- the Central Synchronizer,
- several Command and Tracking Stations, and
- Laser Tracking Stations.

The ground control center is in Moscow. The monitoring stations are uniformly distributed over the territory of the former Soviet Union, hence lacking global coverage. The navigation and control parameters are uploaded twice per day to each satellite. The central synchronizer forms the GLONASS system time and is related to the "phase

control system" (PCS), which monitors the satellite clock time and phase signals. Two laser tracking stations measure the distance and orientation of the GLONASS satellites with the objective to calibrate radio frequency tracking measurements. All GLONASS space vehicles are equipped with laser reflectors. The error specifications for GLONASS broadcast orbits are, following the Interface Control Document, 20 m (along track), 10 m (cross track), and 5 m (radial) for the position vector, and 0.05 cm/s (along track), 0.3 cm/s (cross track), and 0.3 cm/s (radial) for the velocity vector. The time scale synchronisation is within 20 ns.

(e) Geodetic Datum

The satellite coordinates are given in the PZ-90 (*Parametry Zemli 1990*) geodetic datum. Until 1993, the "Soviet Geodetic System 1985" (SGS 85) was in use. The main defining parameters of PZ-90 are given in Table 7.27. Note that slight differences exist compared to WGS 84 (see Table 2.1, p. 28).

Table 7.27. Main defining parameters of PZ-90

Parameter	Value
Semi-major axis	6 378 136 m
Flattening	1/298.257
Geocentric gravitational constant	$398\,600.44 \times 10^9$ m^{3}s^{-2}
Earth rotation rate	$7\,292\,115 \times 10^{-6}$ rad s^{-1}
2nd zonal harmonic	-1082.63×10^{-6}

For GLONASS only solutions, the user positions are determined in the PZ-90 frame. For combined GPS-GLONASS solutions a transformation between PZ-90 and WGS 84 (or ITRF) is required. Transformation parameters for a seven parameter transformation (2.46) [2.1.5] can be derived with collocated GLONASS/GPS observations at a number of stations and/or GLONASS observations at stations with known WGS 84 or ITRF coordinates (Table 7.28). An alternative approach is the computation of precise orbits with microwave and/or laser observation within the ITRF frame, and

Table 7.28. Transformation parameters from PZ-90 to WGS 84, 1: Misra et al. (1996), 2: Roßbach et al. (1996), 3: IGEX-98 (BKG), 4: IGEX-98 (ITRF-97)

Source	ΔX	ΔY	ΔZ	$\varepsilon_x['']$	$\varepsilon_y['']$	$\varepsilon_z['']$	m
1	0 m	2.5 m	0 m	0	0	−0.39	0
2	0 m	0 m	0 m	0	0	−0.33	0
3	0.06 m	0.07 m	−0.57 m	0.035	−0.021	−0.358	$-1 \cdot 10^{-8}$
4	0.3 m	0.0 m	−0.9 m	0	0.012	−0.354	0

comparison with GLONASS broadcast orbits. Several solutions have been published in recent years. For an overview see Roßbach (2001).

Of particular interest are the results of the *International GLONASS Pilot Experiment* IGEX-98 campaign (Slater et al., 1999). In general, the transformation parameters are smaller than the related standard deviation. Significant values have only been found for a rotation of about −0.35 arcseconds around the Z-axis. Note that the quality of the parameters depends on the realization of the PZ-90 frame through the GLONASS broadcast orbits. Analysis of the IGEX-98 data revealed an accuracy of about 5 m for the broadcast ephemerides.

(f) *System Time*

Navigation signals of GLONASS and GPS are tied to slightly different system times. The GPS system is related to the UTC standard maintained by the U.S. Naval Observatory ($UTC_{[USNO]}$), while GLONASS system time refers to the UTC standard in the former Soviet Union ($UTC_{[SU]}$). Unlike the GPS time scale, GLONASS system time considers leap seconds, and it has a constant offset of three hours (difference Moscow time to Greenwich time). GLONASS system time is generated and controlled by the *GLONASS Central Synchronizer*, based on a set of hydrogen masers. The relationship between UTC and GLONASS time is

$$t_{\text{UTC}} = t_{\text{GLONASS}} + \tau_c - 3^{\text{h}}. \tag{7.174}$$

The discrepancy, τ_c, comes from the different clock ensembles used and can reach several microseconds. τ_c is communicated to the GLONASS users in frame 5 of the GLONASS navigation message.

When using both navigation systems jointly, the difference in system time, $\Delta\tau$, depends on the clocks from both systems and has to be taken into account. The estimation of the clock offset, $\Delta\tau$, requires observations to one additional satellite. In other words, at least two GLONASS satellites must be added to the GPS configuration in order to contribute to the positioning solution.

(g) *System status*

The first satellite in the GLONASS System was launched on October 12, 1982. The satellites are carried into orbit three at a time by PROTON launchers from the Baijkonur Cosmodrome in Kazakhstan. The booster is first brought into a low altitude orbit, already with the final inclination of 64.8 degrees, and is then transported via an ascending ellipse (Hohmann transfer [3.4.3]) to the apogee height of 19 100 km. The orbital positioning of the three satellites is performed by their own thrusters. Launches 1 to 6 were pre-operational launches and also carried "dummy" satellites, without navigation payload. The operational deployment phase began with the seventh mission in 1985.

As of January 2001, with 30 launches, in total 74 GLONASS satellites were placed into orbit, along with two passive ETALON satellites [8.2]. The current operational satellites have a mass of 1400 kg, carry three cesium beam oscillators and have a design lifetime of three years. The physical configuration of the GLONASS spacecraft

is depicted in Fig. 7.96. The three-axis stabilized satellite is equipped with a propulsion system for station keeping and relocation, attitude control, and laser corner-cube reflectors. The antenna beamwidth of 35 to 40 degrees provides navigation signal reception up to 2000 km above Earth's surface.

The numbering scheme is manyfold. Besides the international satellite ID number, the GLONASS satellites are given numbers in the COSMOS series, a sequential GLONASS number, an orbital position number, and a channel number. The usual identification follows the channel number.
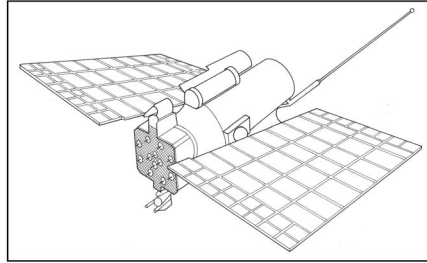
Figure 7.96. GLONASS spacecraft

A new generation of spacecraft, intended to replace older satellites, is under preparation and commonly referred to as GLONASS-M (M for modified). A first launch took place in December 2001, but as of December 31, 2002, no GLONASS-M spacecraft was operational. The main advantages of the GLONASS-M series are
    − longer lifetime of five to seven years,
    − enhanced clock stability,
    − intersatellite communication,
    − autonomous spacecraft operation, and
    − modified structure of the navigation format.
For details on GLONASS satellites see Johnson (1994); Kaplan (1996).

Due to the short design lifetime of the current spacecraft generation, frequent launches are required to maintain the constellation. During the first months of 1996 the constellation was fully deployed with 24 satellites. Thereafter, several spacecraft were withdrawn from service and not replaced. Between January 1997 and January 1999 12 to 16 satellites were always available. Since then, the number has continuously decreased. As of March, 2003, the following eleven satellites were in service:
    − Plane 1:   SV channel 2, 7, 8, 9, 12
    − Plane 3:   SV channel 3, 5, 5, 10, 10, 11.
Note the use of the same channel on pairs of antipodal satellites.

(h) *Use of GLONASS*
During the period of full deployment, GLONASS showed a similar performance to GPS. The advantages of GLONASS are that there has been no artificial signal degradation like SA, and that the P-code is fully available. The user range error (URE) shows a standard deviation of 8 to 10 m (Roßbach, 2001).

After 1992, several commercial receiver makes entered the market. Two groups of user equipment can be distinguished:
    − navigation receivers, L1 C/A code and L1 carrier phase, and
    − geodetic receivers, L1 C/A- and P-code, carrier phase, L2 P-code and carrier phase.

Several advanced receivers offer the possibility to observe both GPS and GLONASS satellites. Examples are the *Ashtech Z-18*, the series of *JPS* receivers, and the *Novatel Millennium* board. Fig. 7.97 shows the JPS "Legacy" dual frequency 40 channel receiver with the RegAnt antenna.

As with GPS, plans were developed in Russia to establish a differential GLONASS (DGLONASS) service. Because of the large size of the country the implementation went slowly (Ganin, 1995). Instead, developments started to combine GPS and GLONASS into a combined DGNSS service (Chistyakov et al., 1996). An important prerequisite was fulfilled with the inclusion of DGLONASS correction data into the format RTCM 2.2 from January 1998 [7.5.1.2].

Figure 7.97. Combined GPS-GLONASS receiver JPS Legacy with RegAnt antenna

GLONASS carrier phase data can be used, either alone or together with GPS data, for precise geodetic applications. For observation equations, modeling of observables, and ambiguity resolution see e.g. Habrich (2000) or Roßbach (2001).

In 1998/1999 a major effort was undertaken to exploit the potential of GLONASS for the geodetic community. Under the auspices of the IAG and the IGS and also ION and IERS, the *International GLONASS Experiment* (IGEX-98) was initiated and realized. The major objectives were to collect GLONASS data for several months from a worldwide network of tracking stations, compute precise orbits, evaluate receivers, and resolve geodetic reference frame and time system issues (Slater et al., 1999). The campaign lasted six months, from October 1998 until April 1999. Over 60 GLONASS tracking stations and 30 Satellite Laser Tracking (SLR) observatories in 25 countries participated. Precise orbits were computed by several analysis centers using the SLR and GLONASS receiver data, with accuracies of 20–50 cm. Datum transformation parameters relating PZ-90, WGS 84, and ITRF were analyzed. The most interesting results were discussed at a meeting in September 1999 and are published in a comprehensive report (Slater et al., 1999).

After the termination of IGEX-98, a number of stations (32 as of December 2000) continued dual frequency tracking within the *International GLONASS Service* (IGLOS) *Pilot Experiment* under the auspices of the IGS. The goals and objectives are (IGS, 2000)

- establish and maintain a global GLONASS tracking network,
- produce precise (10-centimeter level) orbits, satellite clock estimates, and station coordinates,
- monitor and assess GLONASS system performance,
- investigate the use of GLONASS to improve Earth orientation parameters,

  − improve atmospheric products of the IGS, and
  − fully integrate GLONASS into IGS products, operations, and programs.

To support these goals, three GLONASS satellites are tracked with high priority by Satellite Laser Ranging (see [8.5.1]).

The long-term success of an International GLONASS Service certainly depends on the reliability and maintenance of the GLONASS constellation. Many of the potential applications do not require a full constellation but take advantage of GLONASS as an augmentation to GPS. This may, however, become a critical issue if not enough new launches take place and the number of usable satellites further goes down.

7.7.2   GPS/GLONASS Augmentations

GPS and GLONASS are systems under military control and do not fulfill the requirements for safe navigation, in particular coming from the international aviation community. These requirements are, in particular,
  − accuracy,
  − integrity,
  − availability, and
  − continuity of service.

Accuracy requirements depend on the particular application, for example 4 m vertical position accuracy in Category I aircraft approach landing (FRNP, 2001). Differential techniques are required to meet these demands.

Integrity is the ability of a system to provide timely warnings to its users when it should not be used for navigation (see [7.4.6] and Langley (1999b)). This service requires a network of control stations and channels to transmit the warnings in due time to the user.

Availability means the ability of the system to provide usable service within the specified coverage area, and *continuity of service* means the availability of service without interruptions for the intended operations (Hein, 2000).

In order to meet these requirements, augmentation systems to the existing satellite navigation systems GPS (and GLONASS) have been established or are under development. These are the
  − *Wide Area Augmentation System* (WAAS) in the U.S.A. [7.5.3.1],
  − *European Geostationary Navigation Overlay System* (EGNOS) in Europe,
  − *Multi-functional Satellite-based Augmentation Service* (MSAS) in Japan, and
  − *Satellite Navigation Augmentation System* (SNAS) in China.

All are contributions to a first generation of a Global Navigation Satellite System (GNSS-1) and intend to provide seamless coverage of the whole globe. They are also known as *Satellite-Based Augmentation Systems* (SBAS). An alternative solution are *Ground-Based Regional Augmentation Systems* (GBRAS), broadcasting corrections on VHF.

The generic architecture of a satellite-based augmentation system is as follows. A network of GPS (GLONASS) stations at surveyed locations collects dual frequency

measurements of pseudorange and pseudorange rate for all spacecraft in view, along with local meteorological data. The data are processed, and generate precise corrections to the broadcast ephemerides and clock offsets. These corrections, together with system integrity messages, are transmitted to the users via a dedicated package on geostationary satellites. This technique also supports an additional GPS-like ranging signal between GEO and user. Hence, in total three additional signals are provided, a ranging, integrity, and WAD (wide area differential) signal.

The European contribution to GNSS-1, EGNOS (Benedicto et al., 1999), includes augmentations to GPS and GLONASS. It is described in more detail as an example. EGNOS is part of the *European Satellite Navigation Program* (ESNP) and is initiated by the *European Tripartite Group* (European Commission (EC), European Space Agency (ESA), EUROCONTROL) since about 1993. The current EGNOS space segment is composed of transponders on two geostationary INMARSAT-3 satellites, positioned over the Atlantic Ocean Region East (AOR-E) and the Indian Ocean Region (IOR). These satellites provide extra ranging signals over Europe. For the full operational capability (FOC), expected for 2004/2005, additional GEO transponders are required (Soddu, Razumovsky, 2001).

The EGNOS ground segment consists of about 40 "Ranging and Integrity Monitoring Stations" (RIMS), mostly in Europe. These RIMS collect ranging measurements from the GPS, GLONASS and GEO navigation signals on L1 and L2 frequencies. The collected data are transmitted to a set of redundant "Mission Control Centers" (MCC) where the integrity information, differential corrections, GEO satellite ephemerides and ionospheric delays are estimated. These data, together with the GEO ranging signal, are uplinked to the GEO satellites from where they are transmitted on the GPS L1 frequency, as GPS like navigation signals, to the users. It will be possible to receive EGNOS navigation data over Europe, South America, Africa, Western Australia, and a large part of Asia.

The U.S. WAAS architecture is very similar to EGNOS. For details see FRNP (2001). WAAS, however, does not include GLONASS satellites. The system is projected to be fully operational by the end of 2003.

Augmentation systems like WAAS, EGNOS, MSAS, or others, will make it possible for many applications to obtain DGPS accuracy without the cost of extra reference stations or radio data links, and they offer continent-wide coverage. In the long term, augmentation systems are likely to replace the conventional DGPS services.

7.7.3 GALILEO

The European Commission, together with the European Space Agency (ESA) and European industry, is building up a European Satellite Navigation System under the name *GALILEO* as Europe's contribution to GNSS-2. The system will be controlled by civil authorities and be inter-operable with GPS and GLONASS. It offers dual-frequency as standard and will provide real-time positioning and timing services at different levels of accuracy, integrity, and availability. Other than the existing satellite navigation systems, GALILEO is a suitable system for safety critical applications,

such as landing aircraft, guiding cars, tracking hazardous materials, and controlling rail traffic.

The GALILEO schedule comprises several phases. The *definition phase*, from 1999 to 2001, included the initial definition of requirements and system architecture. Two major studies took place, the EC study *GALA* on the system architecture, and the ESA study *GalileoSat* on the space segment. Based on the results of these studies, a 4 years *design and validation phase* from 2002 to 2005 was initiated by the European Council (EC) on March 26, 2002. This phase includes a consolidation of the requirements, the development of satellites and ground based components, and the in-orbit validation. A first experimental satellite will be launched by the end of 2004. Up to four operational satellites will be launched thereafter in 2005 and 2006 for final validation of the space and ground segment. The remaining operational satellites will be launched in the *deployment phase*, from 2006 to 2007, to reach the full *operational phase* in 2008.

The information within this section is mainly taken from ESA documents and the cited literature, e.g. (Forrest, 2002; Eisfeller, 2002). Details are subject to changes during the design and validation phase. For updated information see the ESA homepage, the journal *Galileo's World*, and conference proceedings like ION-GPS.

(a) *Space segment*

The GALILEO space segment, when fully deployed, consists of 30 satellites (27 operational + 3 active spares) in three circular Medium Earth Orbits (MEO) (Fig. 7.98). This configuration is also called *Walker constellation 27/3/1*. The inclination angle is 56 degrees, and the orbital altitude 23 616 km. The orbital period is 14 hours 4 minutes, and the ground tracks repeat after about 10 days. The constellation is optimized for Europe and provides a good coverage up to a northern latitude of 75 degrees.

The GALILEO satellite (Fig. 7.99) has a mass of 625 kg and measures $2.7 \times 1.2 \times 1.1$ m^3; it hence belongs to the class of minisatellites. The navigation payload includes 2 rubidium standards and 2 hydrogen masers. Other
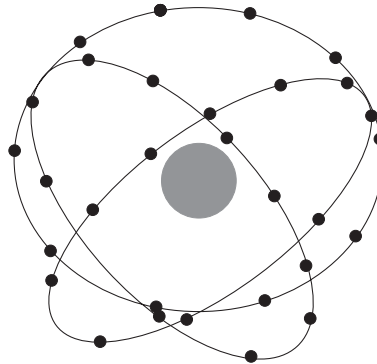
Figure 7.98. Probable GALILEO constellation

than GPS, each satellite carries laser reflectors for independent orbit determination. Several deployment strategies are possible, for example up to 8 GALILEO spacecraft simultaneously launched with ARIANE 5, or up to 6 spacecraft simultaneously with the PROTON launcher. The injection is directly into the MEO orbit.

(b) *Ground segment*

The GALILEO ground segment consists of two *GALILEO Control Centers* (GCC). One is responsible for the control of satellites and the generation of navigation and time
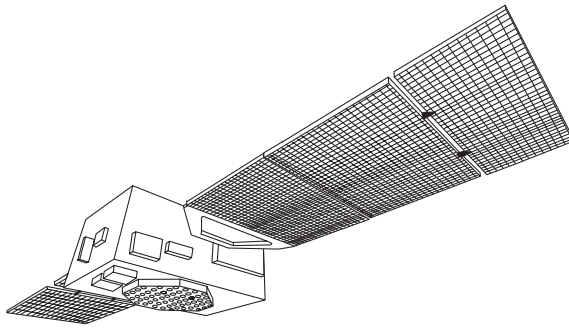
Figure 7.99. GALILEO satellite, possible schematic view

data; the other is responsible for the control of integrity. About 30 globally distributed monitor stations, the *GALILEO Sensor Stations* (GSS), provide data for the GCCs. Data transmission to the satellite is realized via 10 upload stations with S-band and/or C-band antennas.

A further feature is the global *Search and Rescue* (SAR) function. Each satellite is equipped with a transponder which is able to transmit emergency signals to a *rescue center*. A particular link gives a feed back to the user. In Europe the integrity service is closely related to the EGNOS system [7.7.2].

(c) *Services*

Several particular services will be offered within the GALILEO service framework. (Forrest, 2002):
- satellite navigation signals only services,
- combined services with other GNSS or with non GNSS, and
- local services.

Among the satellite only services, besides the "search and rescue service", are four "position, velocity and time services":
- *Open Service*, providing positioning, navigation and time for a mass market, free of charge;
- *Commercial Service*, with added value over the open service, for professional use with service guarantee and user fees;
- *Safety of Life Service*, includes integrity, in particular for landing approach and vehicle guidance;
- *Public Regulated Service*, for applications devoted to European/National security.

For geodesy, surveying and GIS, the open service and the commercial service are of particular interest. The accuracy with a single dual frequency receiver is estimated to be 4 m (horizontal), 8 m (vertical), 50 nsec (time) at the 95% level. The commercial service will have some additional features, such as augmentation with local elements like multiple reference stations [7.5.3].

(d) *User segment*

The receiver architecture will be similar to those used with GPS, but with modern elements in the digital signal processing and reference to the particular GALILEO signal structure. Combined GPS/GALILEO receivers will be designed at least for 4 frequencies. The user market is predicted for 2005 as follows (Eisfeller, 2002):

73% mobile phones,
23% car navigation,
 1% aviation,
 1% fleet management,
 1% leisure, and
 1% surveying.

(e) *Signal structure*

The GALILEO signal structure is not yet definitely decided. Probably the carrier frequencies, shown in Table 7.29, will be used in the lower, middle and upper L-band.

Table 7.29. GALILEO carrier frequencies, status August 2002

Carrier	Middle frequency (MHz)
E5a (L5)	1176.45
E5b	1207.14
E6	1278.750
E2 - (L1) - E1	1575.42

There are potential interferences with the GPS signals L1 and L5, mitigation of which will require particular modulation techniques. On the other hand, such signal overlap facilitates antenna design for hybrid receivers, and guarantees maximum inter-operability. Difficulties also exist with the bandwidths. For E1 and E5 the bandwidth is just 4 MHZ and rather small for a robust signal. As shown, E6 is not a protected band, and hence its use is questionable. In the future at least 5 civil signals will be available to combined GPS/GALILEO receivers, namely code pseudorange and carrier phase:

Modernized GPS: L1, L2, L5
GALILEO: E1-L1-E2, E5a + b.

Simulations show (Eisfeller, 2002) that a combined evaluation of GPS/GALILEO data for geodetic purposes has several advantages:

− increased number of satellites (> 15),
− smaller PDOP (< 1.6),
− increased success rate of the ambiguity fixing, and
− increased positioning accuracy, by a factor of 2 for the horizontal and a factor of 3 for the vertical component.

An interoperable GNSS will enhance the use of satellite-based positioning in difficult environments like mountainous terrain, urban canyons, and around large structures like dams or industrial complexes.

(f) *Applications*

The expected field of possible applications is manyfold, as with GPS [7.6.2]. An overview is given in Table 7.30.

Table 7.30. Possible application markets for GALILEO (Forrest, 2002)

Professional	Mass market	Safety of life
geodesy	personal communication	aviation
precision survey	and navigation	maritime
land survey, GIS	cars	rail
photogrammetry	buses, trucks	police
remote sensing	commercial vehicles	fire
timing	inland waterways	emergency
mining	coastal waters	ambulance
oil and gas	outdoor recreation	search and rescue
environment		personal protection
fleet management		traffic surveillance
precision agriculture		
EEZ delimitation		
fisheries		
vehicle control		
robotics		
construction		
engineering		
meteorology		
space application		

7.8 Services and Organizations Related to GPS

7.8.1 The International GPS Service (IGS)

The IGS was established by the International Association of Geodesy (IAG) and started its activities formally on January 1, 1994, after a pilot phase of about 1 year. The IGS is a member of the Federation of Astronomical and Geophysical Data Analysis Services (FAGS) and it works in close cooperation with the International Earth Rotation Service (IERS). On January 1, 1999, the name of the service was changed from the original "International Global Positioning System (GPS) Service for Geodynamics (IGS)" to *International GPS Service* (IGS). The information in this section is mainly taken from IGS documents, like the "IGS Annual Reports" and the "IGS Directory". These documents are available from the IGS Central Bureau.

Following the *IGS Terms of Reference* (IGS, 2002a), "the primary objective of the IGS is to provide a service to support, through GPS data and data products, geodetic

and geophysical research activities."

The IGS collects, archives and distributes GPS observation data at a number of tracking stations. These data sets are used by the IGS to generate *data products*, namely

− high accuracy GPS satellite ephemerides,
− Earth rotation parameters,
− IGS tracking station coordinates and velocities,
− GPS satellite and tracking station clock information,
− ionospheric information, and
− tropospheric information.

In order to fulfill its tasks, the IGS has a certain structure (Fig. 7.100) with several components:

− Networks of Tracking Stations,
− Data Centers,
− Analysis and Associate Analysis Centers–Analysis Coordinator,
− Working Groups and Pilot Projects,
− Central Bureau, and
− Governing Board.
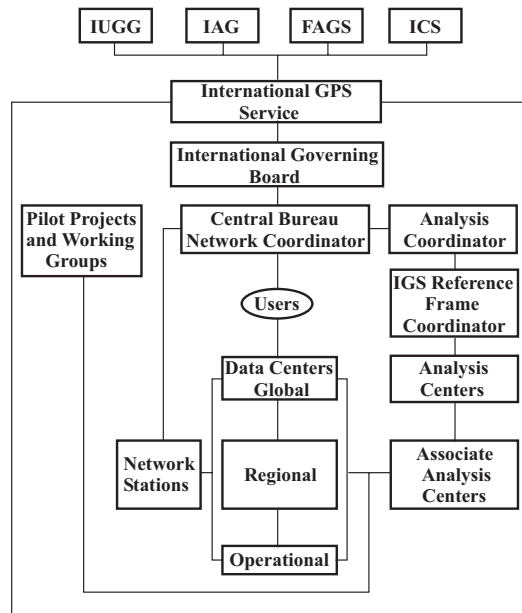
Figure 7.100. Structure of the IGS

The products of the IGS support scientific activities, such as
  − improving and extending the International Terrestrial Reference Frame (ITRF),
  − monitoring deformations of the solid Earth,
  − monitoring Earth rotation,
  − monitoring variations in the liquid Earth (sea level, ice sheets),
  − determining orbits of scientific satellites,
  − monitoring the high atmosphere (ionospheric tomography), and
  − climatological research, contributions to weather prediction.

(a) *Network of Tracking Stations*
The global IGS network of permanent dual-frequency GPS tracking stations included more than 300 stations in 2002, representing some 200 agencies around the world. The number is still growing. Fig. 7.101 shows the global station distribution. The stations have to meet certain requirements, in particular they need data transmission facilities for a rapid (at least daily) data transfer to the data centers. The tracking data are analyzed by at least one Analysis Center or Associate Analysis Center. IGS stations which are analyzed by at least three IGS Analysis Centers, for the purpose of orbit generation, are called *IGS Global Stations* (numbering about 120 early in 2002). All IGS stations can be taken as reference stations for regional GPS analyses. The tracking data are available in RINEX format from the Data Centers. Approximately 90 IGS stations are producing hourly 30-seconds RINEX files, and about 35 stations are providing data in near real-time, delivering 15 minute, 1 Hz data files (Schmidt, Moore, 2002). The ensemble of IGS stations form the IGS network or *polyhedron*. In 2002, at about 50 stations GLONASS satellites were also tracked.
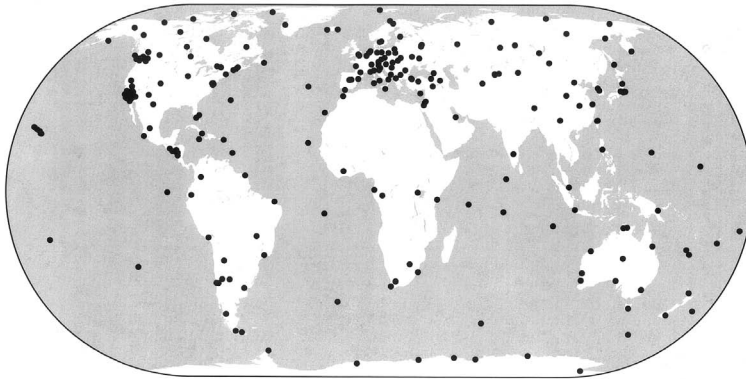
Figure 7.101. IGS Network 2002, source: IGS

(b) *Data Centers*
There are three categories of data centers: Operational, Regional, and Global Data Centers. The *Operational Data Centers* (ODCs), in total 25 at the end of 2000, are in direct contact to the tracking stations. They maintain the data in local archives, validate

and reformat the data (conversion to RINEX and data compression), and transmit them to a regional or global data center. *Regional Data Centers* (RDCs), in total five, collect reformatted tracking data from several data centers, maintain a local archive for users interested in stations of a particular region, and transmit the data to the Global Data Centers.

The *Global Data Centers* (GDCs) are the main interfaces to the Analysis Centers and to the general user community. Among other tasks they archive and provide on-line access to tracking data and IGS products. GDCs provide an on-line archive of at least 100 days of GPS data in the RINEX format, including the data from all global IGS sites. The GDCs also provide an on-line archive of derived products, generated by the IGS analysis or associate analysis centers. There is also on-line access to IGS products generated since the start of the IGS test campaign in 1992. The three Global Data Centers are

CDDIS   Crustal Dynamics Data Information System, NASA
        Goddard Space Flight Center, USA
IGN   Institut Géographique National, France, and
SIO   Scripps Institution of Oceanography, USA

(c) *Analysis Centers*

There are two categories of analysis centers: Analysis Centers (ACs) and Associate Analysis Centers (AACs). The *Analysis Centers* receive and process tracking data from one or more data centers and generate IGS products, as a minimum ephemerides, Earth rotation parameters, station coordinates, and clock information as well as other recommended products. The products are delivered to the Global Data Centers and to the IERS, using designated standards. The Analysis Centers are

COD   Center for Orbit Determination in Europe, University of Berne,
      Switzerland,
EMR   Geodetic Resources Division, Natural Resources Canada,
      Ottawa, Canada,
ESA   European Space Operations Center, European Space Agency,
      Darmstadt, Germany,
GFZ   GeoForschungsZentrum, Potsdam, Germany,
JPL   Jet Propulsion Laboratory, California Institute of Technology,
      Pasadena, California, USA,
NGS   National Oceanic and Atmospheric Administration/National
      Geodetic Survey, Silver Spring, Maryland, USA, and
SIO   Scripps Institution of Oceanography, University of California,
      San Diego, California, USA.

The *Associate Analysis Centers* produce specialized products, for example ionospheric maps or station coordinates and velocities for a global or regional sub-network. Of particular importance are the "Global or Regional Network Associate Analysis Centers" (GNAACs or RNAACs), producing weekly solutions of the global polyhedron or regional subsets thereof. Examples for RNAACs are the EUREF, NAREF or SIRGAS networks for Europe, North- and South America (Blewitt, 1998).

The *Analysis Coordinator* monitors the activities of the Analysis Centers. This person is also responsible for the appropriate combination of the Analysis Center products into a single set of products, in particular a single IGS ephemeris for each GPS satellite.

(d) *Working Groups and Pilot Projects*

A *Working Group* works on a particular topic. A *Pilot Project* has the objective to develop a particular IGS product or service relying on the IGS infrastructure. Active Working Groups and Pilot Projects by the end of 2002 were, for example,

- Reference Frame Densification Working Group,
- IGS/BIPM Time and Frequency Pilot Project,
- Ionosphere Working Group,
- Troposphere Working Group,
- International GLONASS Service Pilot Project,
- Low Earth Orbiter Pilot Project
- Real-time Working Group,
- Tide Gauge Pilot Project, and
- African Reference System (AFREF) Pilot Project

(e) *Central Bureau and Governing Board*

The *Central Bureau* (CB) is responsible for the general management of IGS and coordinates all IGS activities. The *Governing Board* (GB) sets the IGS policy and exercises a broad oversight of all IGS functions and components. The Central Bureau, for the time being, is located at the Jet Propulsion Laboratory in Pasadena, California and maintains an *IGS Information System* (CBIS), accessible at http://igscb.jpl.nasa.gov (World Wide Web). The CBIS contains information on the availability of and access to tracking data and IGS products, IGS orbits, Earth rotation parameters and other data. CBIS also gives access to IGS publications like annual reports and workshop proceedings. An overview of current IGS products is given in Table 7.31.

7.8.2 Other Services

A large number of international and national services provide information on GPS and other GNSS systems. Here only some indications are given. Note that web-addresses may change. For updated information see the regular "Almanac" pages in the August and December editions of *GPS World*.

Canadian Space Geodesy Forum, `http://gauss.gge.unb.ca/CANSPACE.html`
A service of the University of New Bruswick, Canada. Presents daily GPS constellation status reports and ionospheric disturbance warnings. News and discussion about GPS and other space-based positioning systems.

U.S. Coast Guard Navigation Center, `http://www.navcen.uscg.gov/gps`
GPS constellation status, almanac data, information on DGPS, Loran C.

U.S. National Geodetic Survey, `http://www.ngs.noaa.gov/GPS/GPS.html`
Precise and rapid orbits. General information on GPS.

Table 7.31. IGS products, status August 2002, source: IGS Central Bureau

	Accuracy	Latency	Updates	Sample Interval
GPS Orbits & Clocks				
Broadcast *)	$\sim$260 cm / $\sim$7 ns	real time	–	daily
Ultra-Rapid	$\sim$25 cm /$\sim$5 ns	real time	twice daily	$15^m/15^m$
Rapid	5 cm / 0.2 ns	17 hours	daily	$15^m/5^m$
Final	<5 cm / 0.1 ns	$\sim$ 13 days	weekly	$15^m/5^m$
GLONASS Orbits				
Final	30 cm	$\sim$ 4 weeks	weekly	15^m
Geocentric Coordinates				
IGS Tracking Stations				
Final horiz./vert. position	3 mm / 6mm	12 days	weekly	weekly
Final horiz./vert. velocity	2 mm / 3 mm per yr	12 days	weekly	weekly
Earth Rotation				
Rapid polar motion/ polar motion rates/ LOD	0.2 mas / 0.4 mas/d 0.029 ms	17 hours	daily	daily
Final polar motion/ polar motion rates/ LOD	0.1 mas / 0.2 mas/d 0.020 ms	$\sim$ 13 days	weekly	daily
Atmospheric Parameters				
Final tropospheric	4 mm zenith path delay	< 4 weeks	weekly	2 hours
Ionospheric TEC grid	under development			
*) for comparison				

Geodetic Survey Division, Nat.Res. Canada; `http://www.geod.nrcan.gc.ca`
Provides access to data of the Canadian Spatial Reference System (CSRS) and Canadian Active Control System (CACS).

Scripps Orbit and Permanent Array Center (SOPAC); `http://sopac.ucsd.edu`
A service of the Scripps Institution of Oceanography, University of California. SOPAC provides precise, rapid, ultra rapid, hourly and predicted orbits for the IGS in SP3 format. Further SOPAC archives daily RINEX data from about 800 continuous GPS sites from various networks and arrays (IGS, SCIGN, CORS, EUREF and others). SOPAC is also the operational Center for the California Spatial Reference Center.

National Imagery and Mapping Agency (NIMA); `http://164.214.2.59/GandG/sathtml/`
Offers precise GPS orbit information based on tracking data from U.S. Air Force, NIMA, and IGS sites. Daily precise ephemerides and satellite clock estimates in SP3 format. Earth Orientation Parameter predictions.

National Mapping Division, Australia (AUSLIG); `http://www.auslig.gov.au/`
Comprehensive www-site with information on various GPS-related topics. Data from

the Australian Regional GPS Network. On-line GPS processing service for RINEX data.

    Federal Agency for Cartography and Geodesy (BKG), Germany; `http://gibs.leipzig.ifag.de`
The information site "GIBS" is a service of the "Bundesamt für Kartographie und Geodäsie", Fankfurt/Leipzig. GIBS (GPS Informations- und Beobachtungssystem) provides almanac data, information on GPS status, datum transformations, satellite visibility, precise ephemerides, DGPS, GLONASS status and almanac data. The site also contains comprehensive information material on GPS and services like SAPOS. For information on Galileo see

    Galileo Homepage; `http://www.galileo-pgm.org`

    Genesis Office; `http://www.genesis-office.org/`.
Genesis is a project providing support to the European Commission on its GALILEO activities. GENESIS communicates and disseminates information related to GALILEO and also distributes a "Galileo Newsletter".

8 Laser Ranging

8.1 Introduction

In laser distance measurements to satellites (*Satellite Laser Ranging*, SLR) the time of flight of a laser pulse as it travels between a ground station and a satellite is observed. A short laser pulse is generated in the ground station, and is transmitted through an optical system to the satellite. A part of the outgoing laser pulse is used to start an electronic time interval counter (user clock). The target satellite carries appropriate retro-reflectors. The reflected pulse is received at the ground station, detected, amplified, analyzed, and used to stop the electronic counter (Fig. 8.1). The two-way travel
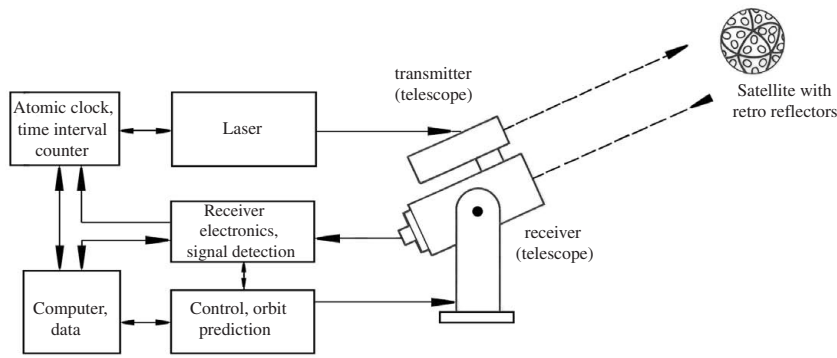
Figure 8.1. The principle of satellite laser ranging

time of the signal is derived from the two readings of the user clock, and is scaled into the distance, d, with the signal propagation velocity, c (cf. (4.8). The basic observation equation is hence very simple:

$$d = \frac{\Delta t}{2} c. \tag{8.1}$$

It is evident that satellite laser ranging is a *two-way ranging* method [4.2.2]. The main components of the ground equipment are
- a generator and transmitter of the laser pulses, including the optical system and mounting,
- a detector and analyser for the return pulses, including the receiver telescope, and
- a time-of-flight measurement unit.

In addition, some sub-components are required for pointing and control of the complete laser-tracking system, and for relation of the observation epochs to universal time

(computer, software, atomic clocks). The space segment consists of suitable satellites equipped with retro-reflectors.

The development of pulsed laser-systems for the tracking of artificial satellites started in the USA as early as 1961/62. The first satellite to carry laser-reflectors, BEACON EXPLORER-B (BE-B), was launched into an orbit of about 1000 km altitude and 80° inclination on October 9, 1964. The first successful signal returns were obtained in 1965 and yielded an accuracy of a few meters (Vonbun, 1977a).

In subsequent years progress has been very fast, the accuracy of range measurements being improved from several meters to a few centimeters (see Fig. 8.2). Satellite laser ranging systems have been developed, configured, and deployed at many places around the world, in some cases as in-house developments from working groups in the observatories. In 2002, about 40 systems were operating worldwide (ILRS, 2002).

The achievable range accuracy is strongly correlated with the length and resolution of the laser pulses. The simple relation is (cf. Fig. 8.5, p. 413) that

$$1 \text{ nanosecond (ns)} \stackrel{\wedge}{=} 15 \text{ cm.} \qquad (8.2)$$
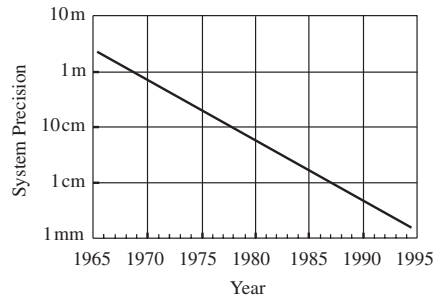
Figure 8.2. Improvement in SLR precision over 30 years

Usually, the laser systems were assigned to one of the following groups (generations), according to their concept and accuracy level:

First generation, pulse lengths of 10 to 40 ns, corresponding to 1 to 6 m in range accuracy; mostly ruby laser with Q-switch [8.3].

Second generation, pulse lengths of 2 to 5 ns, corresponding to 30 to 100 cm; application of sophisticated pulse analysis methods [8.4].

Third generation, pulse lengths of 0.1 to 0.2 ns (100 to 200 picoseconds), corresponding to 1 to 3 cm; mostly mode-locked Nd:YAG laser [8.3]; single photon detection capability.

A new generation of laser-systems is currently being developed with the capability of 1 to 3 mm range accuracy, low eye-safe energy and a high grade of autonomous tracking (Degnan, 2000), see [8.3.4]. Systems of the first and second generation have nearly disappeared from the field of scientific applications. The increase of accuracy over three decades was about three orders of magnitude (Fig. 8.2).

New and broad fields of application have been and are evolving with the increasing accuracy of the measuring systems. With an accuracy range of ±1 cm or better, considerable contributions can be made to the establishment of reference frames, to geodynamics, to the determination of precise satellite orbits and to the modeling of Earth's gravity field [8.5].

Laser distance measurements are among the most accurate observation techniques in satellite geodesy, which is why they will be continuously used in the long-term

solution of important tasks in geoscience. This remains true in spite of the increasing efficiency of microwave techniques like GPS and DORIS. The eminent advantages of the satellite laser ranging technique are, among others, the

- very high accuracy potential, in particular because of the favorable propagation properties of light,
- longevity of satellites without active elements,
- long time series of observations and derived parameters,
- determination of absolute (geocentric) coordinates, in particular absolute heights,
- independent control of other geodetic space techniques, and
- backup for active orbit determination systems like PRARE, DORIS, GPS.

Possible disadvantages are:

- strong dependence on suitable weather conditions,
- high costs in building and maintaining the ground segment,
- inhomogeneous data distribution compared to GPS, DORIS, or VLBI,
- no or limited mobility of the ground segment, and hence only limited operational capability.

For further reading on technical questions see the proceedings volumes of the *International Workshop on Laser Ranging Instrumentation* (e.g. Schlüter et al. (1999); ILRS (2001)). The application of SLR in geodesy and geodynamics is widely discussed in the geodetic literature, in proceedings of the IAG (e.g. Schwarz (ed.) (2000)), in scientific journals like the *Journal of Geophysical Research* or the *Journal of Geodesy*, and in the reports of the *International Laser Ranging Service* (ILRS), see [8.5.1]. A short introduction is also given in Degnan, Pavlis (1994).

8.2   Satellites Equipped with Laser Reflectors

Laser ranging is only possible to satellites equipped with appropriate reflectors. The incoming laser light must be sent back in exactly the same direction from which it comes. Such types of reflectors are also called *retro-reflectors*; they are mostly made from glass prisms. A retro-reflector can be created by cutting an evenly sized pyramid from the corner of a cube. This is why they are often named *corner cube reflectors* (Henriksen, 1977).

In order to attain the desired accuracy, reflectors have to be carefully designed for the particular satellite geometry and orbital height; in particular, the energy balance has to be adjusted. The reflector must be sufficiently large to reflect enough energy. In most cases several single reflectors with a diameter of 2 to 4 cm are assembled in certain arrays, to achieve the necessary energy level. The alignment of the individual reflector requires extreme care in order to avoid pulse deformations caused by signal superposition.

The signal path within the cube corner must be known. If the reflectors cannot be arranged symmetrically with respect to the spacecraft's center of mass, for instance for multiple-purpose satellites, the geometrical relationship between the individual reflector and the satellite's center of mass is required [8.4].

Reflectors are passive devices and can be fitted easily enough as additional components on a given satellite. This is why a fairly large number of space vehicles carry an array of laser-reflectors. Table 8.1 gives an overview of a selection of satellites carrying laser ranging targets. The total number by 2002 amounts to about 70. In

Table 8.1. Satellites carrying laser reflectors (selection)

Satellite Name	Launch [year]	Altitude [km]	Inclination [degrees]
BEACON-B	1964	890	80
BEACON-C	1965	930	41
GEOS-1	1965	1120	40
DIADEME-1C	1967	540	40
DIADEME-1D	1967	580	40
GEOS-2	1968	1080	106
STARLETTE	1975	810	50
GEOS-3	1975	840	115
LAGEOS-1	1976	5850	110
SEASAT	1978	800	108
ASIJAI	1986	1480	50
ETALON-1	1989	19100	65
ETALON-2	1989	19100	65
GLONASS-40 . . . GLONASS-88	1989–2001	19140	65
ERS-1	1991	780	99
TOPEX/POSEIDON	1992	1350	66
LAGEOS-2	1992	5630	53
STELLA	1993	810	99
GPS 35	1993	20100	54
GPS 36	1994	20100	55
ERS-2	1995	800	99
GFZ-1	1995	400	52
TIPS	1996	1020	63
GFO-1	1998	800	108
WESTPAC	1998	830	98
SUNSAT	1999	400	93
CHAMP	2000	430–470	87
STARSHINE-3	2001	470	67
JASON	2001	1340	66
METEOR 3M	2001	1000	100
REFLECTOR	2001	1020	100
ENVISAT	2002	800	98
GRACE A	2002	480–500	89
GRACE B	2002	480–500	89
ICESAT	2003	600	94

most cases, the SLR technique is applied to provide precise orbital information for the particular satellite mission (e.g. for altimeter satellites [9] or for gravity field missions [10]). Today, in most cases where precise orbits are required, space vehicles are fitted with a reflector array as a back-up system.

Some satellites have been launched with the sole objective of serving as precise targets in their orbits. These space vehicles have been optimized in design and orbital parameters. Dedicated laser satellites of this type are STARLETTE, STELLA, LAGEOS-1/2, AJISAI, ETALON-1/2, GFZ-1 and WESTPAC. They are described below in more detail.

STARLETTE was launched by the French Space Agency CNES (Centre National d'Etudes Spatiales) on February 6, 1975 with the following characteristic data (CNES, 1975):

perigee height	805 km,
apogee height	1 108 km.
orbit inclination	49.8 degrees,
period of perigee	∼ 110 days,
nodal period	∼ 91 days,
diameter	24 cm,
mass	47.295 kg, and
retro-reflectors	60, diameter 33 mm.

STARLETTE was the first satellite to be designed with minimized surface forces in order to allow highly precise laser ranging. The core consists of Uranium 238 with a density of 18.7 [g/cm^3], formed as an icosahedron with 20 triangular planes. Each triangle carries a spherical aluminium cap with three integrated retro-reflectors.

Due to the extremely favorable area/mass ratio the disturbing forces (drag and solar radiation pressure [3.2.3]) are minimized, and can be precisely modeled. This is why gravitational forces, acting on low orbiting satellites, can be separated and well analyzed. The main fields of application are the determination and analysis of [8.5]

— ocean tides and body tides (main purpose),
— Earth's gravity field,
— geocentric station coordinates,
— polar motion, Earth rotation, and
— tidal friction.

Because of its rather low orbit, STARLETTE is particularly suitable for the study of solid Earth tides and related elasticity models of the Earth [8.5.6].

A virtually identical twin satellite, named *STELLA*, was launched into a sun-synchronous orbit on September 26, 1993. The orbital parameters are:

— inclination 98.6 degrees,
— height 800 km,
— quasi circular orbit.

As with STARLETTE, the main objectives are: a contribution to the gravity field, in particular tuning the field for sun-synchronous Earth observing satellites such as SPOT, ERS and others; a contribution to the modeling of non-gravitational forces; and

for modeling the Earth and ocean tides. The anticipated lifetime of STARLETTE and STELLA is several centuries.

   LAGEOS-1 was launched by the American Space Agency NASA on May 4, 1976; and *LAGEOS-2* as a joint U.S.-Italian project on October 22, 1992. The orbital characteristics are:

	LAGEOS-1	LAGEOS-2
Perigee height	5860 km	5620 km
Orbit inclination	109.84 degrees	52.64 degrees
Eccentricity	0.0045	0.0135
Period	225 minutes	223 minutes
Diameter	60 cm	60 cm
Shape	sphere	sphere
Mass	411 kg	405 kg
Reflectors	426 corner cubes	426 corner cubes

The design goals of LAGEOS were, as for STARLETTE, to minimize surface forces, and to create a precise relationship between the satellite's center of gravity and the individual reflectors. Due to its greater altitude, the LAGEOS orbit is less sensitive to atmospheric drag and short wavelength terms of Earth's gravity field than is the STARLETTE orbit. The retro-reflectors are incorporated into an aluminum sphere surrounding a cylindric brass core (Fig. 8.3). 422 silicon reflectors serve for the pulse range measurements. Four additional germanium reflectors were designed for range rate observations with optical Doppler measurements. The name LAGEOS stands for *Laser Geodynamics Satellite* (originally "Laser Geodetic Satellite"), and thus indicates the main fields of application [8.5]:

Figure 8.3.   LAGEOS (Laser Geodynamics Satellite) structure

   − installation and maintenance of a precise geodetic reference frame,
   − determination of tectonic plate motion and regional crustal movements,
   − determination of Earth orientation parameters (polar motion, Earth rotation),
   − study of Earth's gravity field.

The lifetime of the LAGEOS satellites is estimated to be several millions of years. This is why a steel plaque, indicating continental drift, was added to the first spacecraft as a "message to the future".

   The Japanese *Experimental Geodetic Satellite* (EGS), also named *AJISAI*, was launched on August 12, 1986 into a circular orbit of 1 500 km altitude and 50° inclination (cf. [4.3.2]). The orbital period is 1.93 hours, the rotation period of perigee 142.53 days, and the nodal period 117.53 days. The spherical satellite has a diameter

of 214 cm, a mass of 685 kg, and carries 120 laser reflector assemblies. The area/mass ratio is hence less favorable than for STARLETTE and LAGEOS. The satellite can be used for laser range and photographic direction measurements. The original goal was to determine the location of isolated islands and to adjust the geodetic network of Japan (Komaki et al., 1985). In the meantime, tracking of AJISAI has contributed considerably to the improvement of gravity field models and the geodetic reference frame (Torrence, 1999).

In January and May 1989 the former Soviet Union launched two spherical satellites, named *ETALON*-1 and *ETALON*-2, each time together with two GLONASS satellites [7.7.1] into rather high orbits. The characteristic parameters are:

altitude	19 120 km,
eccentricity	0.00068,
orbit inclination	65 degrees,
diameter of sphere	1.294 m,
mass	1 415 kg,
period	675 minutes, and
reflector arrays	306, each with 14 corner cubes.

Of the reflectors, six are made of germanium for possible future infrared interferometric measurements, and they are placed symmetrically.

The original objective of the ETALON mission was to determine solar radiation pressure for the orbit control of GLONASS satellites (Anodina, Prilepin, 1989). Because of the high orbital altitude, the ETALON satellites, together with LAGEOS, form the basis of a high-accuracy global reference coordinate frame [8.5.4]. Further significant contributions are expected to the modeling of the low order gravitational field parameters, to the determination of the geocentric gravitational constant (GM) and station positions, and to the estimation of Earth orientation parameters.

Two particular dedicated laser tracking satellite are *GFZ-1* and *WESTPAC-1*. GFZ stands for *GeoForschungsZentrum Potsdam*. The small spherical satellite with a diameter of 21.5 cm and a mass of 20.6 kg (Fig. 8.4) carried 60 retro-reflectors and was jettisoned from the Russian MIR space station on April 19, 1995, into a low (400 km) nearly circular Earth orbit of 51.6 degrees inclination. From this initial altitude it was to decay naturally with a predicted lifetime of 3.5 to 5 years. On June 23rd 1999 it burned up after nearly 24 000 orbits. The last observation placed GFZ-1 at an altitude of 230 km.

Figure 8.4. GFZ-1; courtesy GFZ Potsdam, photo: L. Grunwaldt

The mission objectives of GFZ-1 were to determine variations in rotational characteristics of the Earth and to recover high resolution parameters of the gravity field.

Through the changing orbital height during the satellite's lifetime it was possible to estimate a wide variety of higher order gravity coefficients [8.5.3].

WESTPAC-1 stands for *Western Pacific Laser Tracking Network Satellite*. The satellite is of a similar size to GFZ-1, and was launched on July 10, 1998, into a sun-synchronous circular orbit of 835 km altitude and 98 degrees inclination. The satellite has a diameter of 24 cm, a mass of 23 kg and carries 60 corner cube reflectors. WESTPAC-1 was designed in particular to provide a high ranging accuracy. The specific features are a center-of-mass correction within 0.5 mm accuracy and that only a single corner-cube will reflect at any shot.

In total, at the end of the year 2002, eight dedicated laser-satellites are in orbit. The tracking characteristics are quite different allowing multi-satellite ranging for appropriate tracking systems [8.3]. Table 8.2 shows some characteristic features of the most important dedicated targets.

Table 8.2. Tracking characteristics of dedicated satellite laser targets

Satellite	Mean altitude	Maximum pass duration	Signal flight time Min./Max.
STARLETTE/STELLA	810 km	10 min.	6/14 ms
AJISAI	1 490 km	15 min.	10/20 ms
LAGEOS-1, -2	5 999 km	50 min.	40/57 ms
ETALON-1, -2	19 100 km	330 min.	127/150 ms

8.3 Laser Ranging Systems and Components

8.3.1 Laser Oscillators

The most important component of a laser ranging system is the laser-oscillator. The artificial word *LASER* (*L*ight *A*mplification by *S*timulated *E*mission of *R*adiation) denotes a configuration for the coherent amplification of electromagnetic oscillations in the (optical) spectral domain through induced emission. In an optical resonator, the electromagnetic wave interacts with excited material.

Besides the coherence, (i.e. the fixed phase coupling between the individual beams providing monochromatic light), two more properties of the laser are exploited in satellite geodesy. These are the high degree of collimation of the beam, and the high power density. Hence, very high-energy, sharply defined, pulses can be transported over large distances.

In satellite geodesy two types of solid state pulsed lasers are widely used, the ruby laser and the Neodymium-YAG laser. The SLR systems of the first and second generation are almost exclusively equipped with ruby lasers, whereas the third generation systems mostly use the Nd:YAG laser.

Ruby is the classic material of *solid state lasers*. Ruby is a crystal, absorbing green and blue-violet light, and emitting sharp red spectral lines at 694.3 nm. By changing the resonator quality and opening the resonator at the predefined maximum of energy absorption, single laser pulses can be generated with a pulse width of about 10 to 50 nanoseconds and a peak power of 1 GigaWatt. The process is controlled by the so-called Q-switch (Q stands for quality). With a special arrangement of the Q-switch inside the resonator it becomes possible to reduce the pulse width to 2–5 ns. This is, however, the upper limit of performance for a ruby laser.

Another way of generating short pulses is the coupling of longitudinal resonator oscillations, the so-called *modes*, through active modulators, producing a defined sequence of short, high energy pulses. In particular, the *Neodymium-YAG laser* (YAG = *Y*ttrium-*A*luminium-*G*arnet) is suited for the mode-coupling. This technique makes a reduction of the pulse width to 100 to 200 picoseconds possible. It also requires less pumping energy, and hence provides a better system performance and a higher pulse repetition rate. Finally, the frequency is doubled and, with a wavelength of 530 nm (green) instead of 1060 nm (infrared), produces better conditions for the reception of return pulses.

Modern developments are directed toward eyesafe lasers, i.e. lasers with low power and high repetition rate (see [8.3.4]). For an overview of the current status in satellite laser ranging technology see the proceedings of the biannual *International Workshop on Laser Ranging*, e.g. Schlüter et al. (1999); ILRS (2001).

8.3.2 Other System Components

(a) *Telescope Mount*
The transmitter component must be able to follow moving targets. This is possible with a mounting that permits changes in azimuth and elevation. It is advisable to fit the receiver to the same mounting, or to integrate the transmitter and receiver telescopes.

For first generation systems the laser apparatus was usually also mounted on the pointing assembly. Third generation lasers are rather sensitive and hence need a well controlled environment. Stationary systems are usually kept on a rigid bench in a particular clean-room near the pointing assembly. The laser pulses are directed via a series of prisms or optical conductors to the transmission telescope. It is necessary to point the telescope with sufficient accuracy to the satellite. For first generation systems, tracking was often controlled visually with the help of a guidance telescope. The pointing control of third generation systems usually works automatically, with computer control, based on pre-computed ephemerides, so-called IRVs (*Inter-Range Vector*, see [8.5.1]). This is also required because of the ability to make daytime observations.

During the satellite pass, corrections are derived from a comparison between the pre-determined and actual satellite positions. In order to achieve a high return rate, even for distant satellites, a pointing accuracy of $\pm 1''$ or better is aimed for, which is quite a demanding requirement for guidance and control. The divergence of the outgoing laser beam can usually be adapted to different satellite ranges, and for tracing the

target. A fast switch between different satellites, following a priority list, is essential for participation in international projects [8.5.1].

(b) *Receiver elements, signal detection package*

The laser pulse energy per unit area decreases with the square of the distance, and in addition the signal is attenuated in the atmosphere. This is the reason why very little energy comes back to the ground system after travelling to and from the satellite, in spite of the high initial energy and the strong beam focussing. Very powerful receiver systems are hence required, especially for distant satellites.

The receiver unit consists of optical and electronic components. Reflector telescopes (mirrors) or refractor telescopes (lenses) can be used as optical receivers. In most cases preference is given to a reflector because of its better capacity for weak luminosities. The geometrical quality of the signals is of minor interest. A filter of very low bandwidth is used ($\Delta\lambda \sim 1$ nm) in the frequency domain of the laser light, in order to minimize the influence of disturbing light.

The signal detection package usually contains a *photomultiplier* with an extremely short rise time and high resolution. To avoid the reception of disturbing signals, the element is only activated for a very short pre-determined time interval Δt (0.1 to 10 μs). The rise time should not be larger than 100 to 300 ps, and the necessary amplification is about 10^5. Third generation laser systems work on the basis of single photon detection. Some new developments use *microchannel plate photomultipliers* (MCP) for amplification and a so-called *streak-camera* to collect the echo photons (Riepl, 1998).

(c) *Impulse analysis*

The transmitted laser pulse has a well defined form as is shown in Fig. 8.5 (a), in which the energy distribution is traced along the direction of the signal propagation. H_1 and H_2 denote the level where the impulse reaches half of the amplitude. The distance between H_1 and H_2 is called pulse length or *pulse half width*. The epoch of transmission, t_a, referred to the pulse center, can be easily determined by triggering techniques if the half-length of the known or measured pulse length, l_i, is added to the trigger signal, H_1.
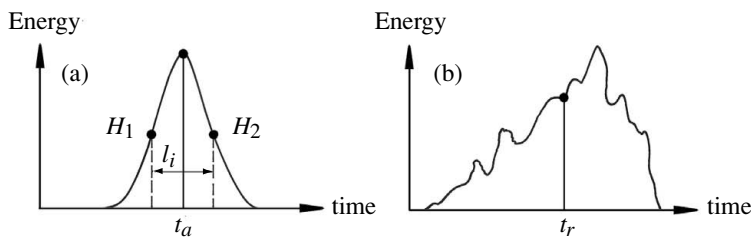
Figure 8.5. Shape of the transmitted (a) and received (b) pulse signal in satellite laser ranging

The shape of the return signal is deformed because of several disturbing influences (Fig. 8.5 (b)). Amongst these are atmospheric disturbances, superposition caused by

signal reflection at different retro-cubes, and relative motion between transmitter and reflector. A careful pulse-analysis is required to determine the pulse center.

For modern systems, working on the basis of single photon-electron detection, no pulse analysis is possible. In these cases the single photon-electrons have to be identified and analyzed with very fast detectors. Current techniques use, for example, SPAD (*Single Photon Avalanche Diode*) techniques (Prochazka et al., 1999).

Instead of using a single pulse, techniques have been developed to use a train of 5 to 10 short pulses (about 50 ps length) at a fixed interval of a few nanoseconds. From this train an electro-optical shutter passes about half the pulses, the so-called *semitrain*, containing 3 to 5 pulses (Paunonen, 1999), see also Hamal, Prochazka (1989). This method increases the precision and decreases the single pulse energy.

(d) *Time Base*

The signal travel time is measured by a propagation timer, which is controlled by an extremely accurate clock. Electronic counters are used with a resolution of about 10 ps. The counters are controlled by atomic clocks with high long-term and short-term stability, in particular rubidium and cesium standards, or hydrogen masers [2.2.5]. Atomic clocks also define the station system time which is needed for determination of the observation epochs. Regular comparisons with international time scales (UTC) are required, and can be easily realized via an appropriate GPS receiver with an accuracy level of better than 20 ns [7.6.2.9].

(e) *System Computer*

A suitable computer is required, with the related software, for the pre-calculation of satellite ephemerides and pointing elements, the guidance and control of the instrument mounting, the control of the whole system, calibration and control of system parameters, data analysis, data control and data transfer. In modern systems, multitasking and network processors with real-time capability, as well as remote control, are required.

(f) *Aircraft Detector*

In some areas with dense air traffic, and near airports, it may be required to make provisions against an airplane passing through the laser beam. An optical or radar airplane detection system can be deployed that automatically interrupts the laser operation. Because of the low energy of modern laser ranging systems (eye safe operation), the requirement for installing airplane detection devices is now less stringent.

8.3.3   Currently Available Fixed and Transportable Laser Systems

In 2002 about 40 systems were used worldwide for laser ranging to satellites. Most of them now belong to the third generation or are new developments. The majority has the capability of ranging to high satellites such as ETALON, GLONASS and GPS, while only three or four systems can reach the Moon. Most systems installed are stationary, although the number of transportable systems is increasing. A current overview of systems contributing to global geodesy and geodynamics [8.5] is given in

the documents of the *International Laser Ranging Service* (ILRS) [8.5.1]. Many of
the older systems have been replaced or upgraded in recent years (Husson, 1999).

Table 8.3 gives an overview on the system data of two modern laser ranging systems,
the *Wettzell Laser Ranging System* (WLRS), operating in the fundamental station
Wettzell, Germany [12.5], and the MOBLAS-7, operating at the *Goddard Geophysical
Astronomical Observatory* (GGAO) in Greenbelt, Md. USA.

Table 8.3. System data of two laser ranging systems

System	WLRS	MOBLAS-7
Telescope	75 cm mirror	76 cm mirror
Mount	Alt/Az	Alt/Az
Lasertype	Nd:YAG	Nd:YAG
Laserfrequency	532 nm	532 nm
Operating mode	single pulse 532nm (100ps, 180 mJ)	single pulse
	single pulse 1064 nm (100 ps, 360 mJ)	(200ps,100 mJ)
	pulse semitrain (4-8 pulses, 300 mJ)	
Pulse repetition rate	1 to 10 Hz	1, 5, 10 Hz
Receiver	Photomultiplier, Avalanche diode,	Photomultiplier
	Microchannel plate photomultiplier	Microchannel plate
	Streak camera	photomultiplier
Observation range	satellites and Moon	high satellites

The global geographical distribution of laser systems (see Fig. 8.13, p. 432) reflects
national capabilities and interests, and is often not very suitable for the analysis of
regional and local geodynamical phenomena. To allow more flexible applications, in
particular in the determination of crustal motion [8.5.1] [12.4.1], transportable systems
of the newest laser technology are being developed. Some of them have already been
widely used in recent years, for example in the Mediterranean area (MEDLAS project
[8.5.4]) and in the NASA *Crustal Dynamics Program* [12.4.1]. The systems have
a modular construction and can be transported with containers in regular airplanes.
They work with quite low energy and with single photon detection. Typically, mobile
systems occupy sites for periods of 2–3 months, and then require several days for
relocation.   Examples of current transportable systems are:

    FTRLS-1   (France),
    TLRS   (Germany),
    TROS   (China),
    MTLRS-1   (Germany), and
    MTLRS-2   (Netherlands).

Table 8.4 gives some system data. All systems work with the single photon technique
and allow daylight operation.

Table 8.4. System data of transportable laser ranging systems

System	MTLRS-1/2	FTLRS	TLRS
Aperture	40 cm	13 cm	50 cm
Weight	500 kg	300 kg	1700 kg incl. cart
Laser	Nd:YAP 539 nm	Nd:YAG 1064 nm	Titan Sapphire 427/ 854 nm
Pulse energy	10 mJ	100 mJ	30 mJ
Pulse length	200 ps	100 ps	80 ps
Repetition rate	10 Hz	10 Hz	10 Hz
Time base	cesium	rubidium, GPS controlled	2 cesium 2 hydrogen maser
Range	6 000 km	6 000 km	36 000 km

The *Modular Transportable Laser Ranging Systems*, MTLRS-1 and MTLRS-2, have been operated successfully since 1984 for about 15 years, mainly in the Mediterranean for international geodynamic projects (e.g. WEGENER/MEDLAS). The slight difference of wavelength, as compared with the Nd:YAG laser, comes from the use of a different laser active material, named Nd:YAP (YAP: *Y*ttrium Ortho*A*luminate, a crystal of the mineral type *P*erovskite). The normal point [8.4.2] accuracy is about 1 to 2 cm.

The *French Transportable Laser Ranging Station*, FTRLS (Pierron et al., 1999), was developed by French organizations and entered its operational phase in 1996. The system is highly mobile, weighing only 300 kg, in 8 containers. The system can reach satellites at LAGEOS height. Its main objective is the installation of low cost laser stations in remote areas for research in geodynamics and the calibration of altimeter satellites [9.3.3].

The *TIGO Laser Ranging System*, TLRS, is designed to measure ranges to satellites with an accuracy better than 1 cm simultaneously at two wavelengths. The tracking range is from low orbit satellites at about 300 km altitude up to geostationary satellites (36 000 km). TIGO stands for *Transportable Integrated Geodetic Observatory*. Besides the SLR module, TIGO includes a VLBI module [11.1], GPS, GLONASS and DORIS receivers, a time-keeping laboratory and a superconducting gravity meter (Schlüter et al., 2000). The installation of TIGO at a particular site is always anticipated for a duration of several years. Since 2002 TIGO has been operating in Concepción, Chile [12.5.2].

8.3.4   Trends in SLR System Developments

Some of the main development trends are toward the following characteristics:
   − short pulse width, down to 50 picoseconds or even less; pulse trains,
   − high repetition rate (10 Hz); low output signal strength,
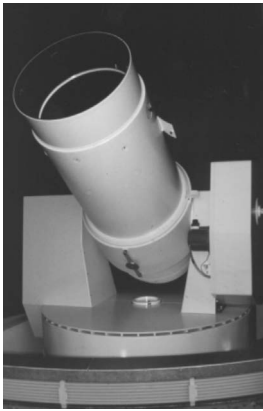
Figure 8.6. Wettzell Laser Ranging System; courtesy BKG

Figure 8.7. TIGO with SLR module (Concepción, Chile); courtesy BKG

- single photon detection; exploitation of quantum-statistic properties; faster electronics; improved photodetectors such as *single photon avalanche diode* or *microchannel plate*; streak cameras,
- eyesafe systems, i.e. low energy (150 μJ) and high repetition rate (2 KHz),
- fully automatic operation, remote control, 24-hour operation,
- reduced station construction, operating and maintenance costs,
- higher mobility through low-weight optics,
- real-time data processing and data transfer,
- software-oriented systems, hence higher flexibility,
- multiple satellite tracking capability,
- low error budget ($\sim$1 mm),
- two-color ranging, and
- epoch synchronization down to 10 ns (corresponding to 0.1 mm).

NASA is developing a modern SLR system meeting most of these requirements under the name *SLR2000*. SLR2000 is an autonomous, eyesafe, single photon-counting satellite laser ranging station with an expected single shot range precision of about one centimeter and a normal point precision better than 3 mm (Degnan, 2000). The system is designed to provide 24-hour tracking coverage. It is planned to build more than 10 systems with a replication cost of $1M per system. The main features are:

- Q-switched Nd:YAG microlaser, frequency-doubled, 532 nm,
- 130 μJ of energy, 2 KHz repetition rate,
- high-speed quadrant microchannel plate photomultiplier,
- high-speed high resolution event timer,
- arcsecond precision tracking mount,
- shelter and protective dome,
- CCD camera for guidance control and focussing, and
- daylight tracking capability to GPS.

Potential upgrades are to a two-color system as well as adapting the SLR2000 for interplanetary ranging through the use of transponders.

The ILRS [8.5.1] has established *System Performance Standards* in order to evaluate SLR systems. The performance guidelines are divided into three categories (data quantity, data quality and operational compliance). A high performance SLR system should fulfill the following criteria for

yearly data quantity:
  – 1000 Low Earth Satellite (LEO) passes,
  – 400 LAGEOS-1,2 passes, and
  – 100 high satellite passes;

data quality:
  – 1 cm LAGEOS normal point precision,
  – 2 cm short term range bias stability (standard deviation of the pass by pass biases), and
  – 1 cm long term bias stability (standard deviation of the monthly biases for 8 of the last 12 months);

operational compliance:
  – data delivery within 24 hours,
  – specific ILRS normal point data format, and
  – site and system information form.

About one third of the approximately 40 SLR stations operating worldwide at the end of 2002 fulfill the System Performance Standards. For current details see the ILRS Annual Reports.

8.4   Corrections, Data Processing and Accuracy

8.4.1   Extended Ranging Equation

In order to describe the measuring process it is necessary to introduce additional parameters and corrections into the simple basic observation equation (8.1). Following Fig. 8.8 (Aardoom et al., 1982) we find for the most important components that

$$d = \frac{1}{2}c\Delta t + \Delta d_0 + \Delta d_S + \Delta d_b + \Delta d_r + \eta, \qquad (8.3)$$

with
Δd_0 eccentricity correction on the ground,
Δd_S eccentricity correction at the satellite,
Δd_b signal delay in the ground system,
 Δt measured flight-time of the laser pulse between the start and the stop signal,
Δd_r refraction correction, and
  η remaining systematic and random observation errors.
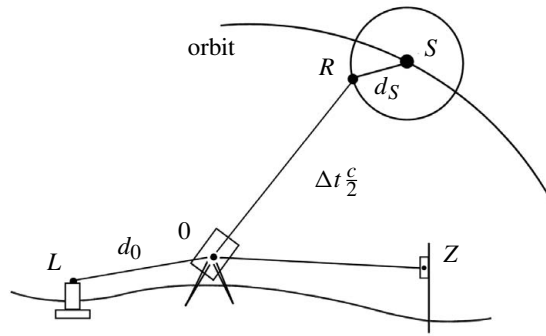
Figure 8.8. Geometrical relationship in satellite laser ranging

As a general rule, geometrical and physical corrections should be applied with one order of magnitude accuracy higher than the corresponding resolution of the observables. This means, for third generation satellite laser ranging systems, that the corrections are required with an accuracy of 2 to 3 mm. For the current development of laser systems (toward the 1 mm level of observation accuracy), corrections and biases should accordingly be modeled with submillimeter accuracy.

(a) *Time measurement,* Δt

Two aspects have to be distinguished. Firstly, the measured ranges must be tied to universal time, UTC, because of satellite motion relative to Earth. An accuracy of ± 100 ns (corresponding to a satellite motion of 1 mm) is completely sufficient for most purposes and does not constitute any problem for modern techniques of time keeping or time transfer [2.2.5]. Secondly, the flight time of pulses, Δt, has to be measured. The time-tagging of the start and stop events is affected by uncertainties in the signal identification. The quality of these measurements constitutes one of the most critical accuracy limitations in the whole error budget. The desired resolution is a few picoseconds.

(b) *Eccentricity corrections,* Δd_0, Δd_S

Generally the intersection, 0, of the vertical axis with the horizontal axis is used as the reference point in the ranging system. The position of 0 has to be connected with 1 mm-accuracy or better to the station marker, L. The stability of 0 must be controlled. The eccentricities are also needed for local ties between collocated sites of differing space techniques like SLR, VLBI and GPS.

The geometrical relationship between the center of mass, S, of a satellite and the optical center, R, of a single cube corner reflector, the so-called *center-of-mass correction* (CoM), has to be established for all usable satellites. This can be done with high accuracy for the spherical satellites such as LAGEOS, STARLETTE and AJISAI. The situation is more difficult for irregularly shaped satellites (e.g. altimeter satellites). A careful pre-launch calibration is of eminent importance. If return signals come from

several cube corner reflectors it is necessary to analyze the impulse response functions. The CoM corrections may vary by several millimeters between single photon systems and microchannel plate photomultipliers (MCP) (ILRS, 2000). A careful design of the retro-reflector array (e.g. for WESTPAC-1) ensures that only a single corner-cube will reflect any shot.

(c) *Propagation correction, Δd_r*

Laser impulses experience a delay in the atmosphere. It is not possible to measure the atmospheric state parameters along the total path; therefore atmospheric models are used which are supported by measured atmospheric data at the laser site. In the frequency domain, applied for laser light, the atmospheric refraction can be modeled reliably for elevations above 10°. The correction is fairly insensitive to water vapor content. The total refraction delay (see [2.3.3.2]) for some elevations is:

zenith direction ~ 2.5 m,
20° elevation ~ 7.3 m,
10° elevation ~ 14 m.

For satellite laser ranging the formulation of *Marini* and *Murray* is commonly used and recommended in the IERS conventions (McCarthy, 2000). The formula has been tested by comparisons with ray-tracing radio-sonde profiles. The correction to the one-way range is

$$\Delta d_r = \frac{f(\lambda)}{f(\varphi, H)} \cdot \frac{A + B}{\sin E + \frac{B/(A+B)}{\sin E + 0.01}}, \tag{8.4}$$

where

$$A = 0.002357 P_0 + 0.000141 e_0,$$

$$B = (1.084 \times 10^{-8}) P_0 T_0 K + (4.734 \times 10^{-8}) \frac{P_0{}^2}{T_0} \frac{2}{(3 - 1/K)},$$

$$K = 1.163 - 0.00968 \cos 2\varphi - 0.00104 T_0 + 0.00001435 P_0,$$

and

Δd_r range correction (meters),
E true elevation of satellite (degrees),
P_0 atmospheric pressure at the laser site
    (in 10^{-1} kPa, equivalent to millibars),
T_0 atmospheric temperature at the laser site (degrees Kelvin),
e_0 water vapor pressure at the laser site
    (in 10^{-1} kPa, equivalent to millibars),
f(λ)   laser frequency parameter ($\lambda =$ wavelength in micrometers), and
f(φ, H)   laser site function.

The *laser frequency parameter* is

$$f(\lambda) = 0.9650 + \frac{0.0164}{\lambda^2} + \frac{0.000228}{\lambda^4}.$$

The *laser site function* is

$$f(\varphi, H) = 1 - 0.0026 \cos 2\varphi - 0.00031H,$$

where φ is the latitude and H is the station height in kilometers.

The currently recommended model of Marini and Murray is believed to be uncertain by up to 1 cm in the zenith delay term and up to several cm at low elevations. Improved models and mapping functions are under discussion, see e.g. Mendes et al. (2001). Modeling of the propagation correction can be improved with two-color ranging because the troposphere is a dispersive medium for wavelengths in the optical domain [2.3.1.2]. Two pulses at different wavelengths have to be sent out, and the differential times of arrival have to be measured with picosecond precision. Developments of this challenging method are underway. Successful test measurements with streak-camera based systems have been reported (Bianco et al., 1999).

(d) *Delay correction,* Δd_b
The geometric reference point, 0, within the laser ranging system does not necessarily correspond to the electrical zero point of the measurements. This can be interpreted as a systematic *time delay* with a superposed uncertainty, the *laser jitter* (Fig. 8.9). The laser delay is determined by calibration. Older systems are calibrated with respect to a known terrestrial target, Z. For modern instruments the calibration is performed inside the laser system.
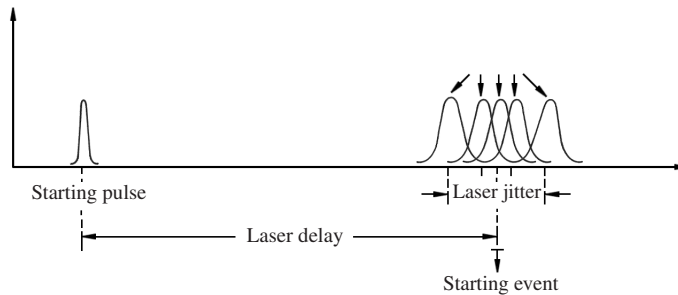
Figure 8.9. Laser delay and laser jitter

A considerable contribution to a delayed range measurement comes from internal detector properties and is related to the intensity of the detected light pulse. For single photon operation the delay is constant and can be calibrated. Single photon sensors are hence capable of millimeter range precision. However, when the return energy exceeds the single photon level, time walk effects are introduced and have to be compensated for (Kirchner et al., 1999).

The term η contains unmodeled residual effects, such as instabilities in the ranging system. For more detailed discussions on accuracy and corrections see e.g. the

proceedings of the biannual *International Workshop on Laser Ranging*, e.g. Schlüter et al. (1999); ILRS (2001).

The most effective method of testing a satellite laser ranging system is by parallel observation with another system at the same site (*collocation test*).

8.4.2   Data Control, Data Compression, and Normal Points

The observed raw data are controlled in a filtering and data compression process in order to
  − detect and eliminate gross errors (blunders),
  − evaluate the accuracy of the observations, and
  − reduce the amount of data for subsequent processing.
Gross errors may arise, in particular, during day-time observations, if spurious return signals are acquired. The quality of the single observations can be assessed through comparison of the individual measurements with a curve smoothed through all observations. Data compression is necessary, because modern SLR systems with pulse repetition rates of 10 Hz may produce several thousand data points per satellite pass. These data are highly correlated because of instrumental and meteorological effects. For subsequent investigations, only one representative range mean is required for each time interval of about one or a few minutes.

Data control and data compression can be achieved operationally within one multi-step process. Several methods have been proposed; in the interest of international cooperation, however, the procedure recommended at the "Herstmonceux Laser Workshop" in 1984 (Gaignebet (ed.), 1985) is mostly used, and has since been discussed and updated (e.g. Sinclair, 1999).

In the first step, the observed ranges, d_0, are compared with computed reference ranges, d_p (predictions), and a series of residuals, d_r, is generated (Fig. 8.10):

$$d_r = d_0 - d_p. \tag{8.5}$$

The reference ranges can be obtained from all available observations, be they short arc or long arc approximations of the observed orbit [3.3.3.3]. This procedure demands a rather large computational effort; high precision predictions are required with a best available estimate for a time bias. The predicted range must include the refraction delay. Data with gross differences (outliers) are eliminated using an adequate range window.

In the second step a suitable trend function, $f(p)$, is fitted to the residuals, d_r, either using a set of orbital parameters (preferable) or a polynomial, e.g. Chebyshev polynomials [3.3.3.2]. Care has to be taken not to introduce spurious high frequency signals by fitting a high order polynomial. The deviations after the fit,

$$f_r = d_r - f(p), \tag{8.6}$$

are analyzed for any remaining outliers, using a $3\,\sigma$-criterion. This approximation procedure can be repeated iteratively. For systems that detect the first photo-electron a $2.5\,\sigma$ criterion can be of advantage (Sinclair, 1999).
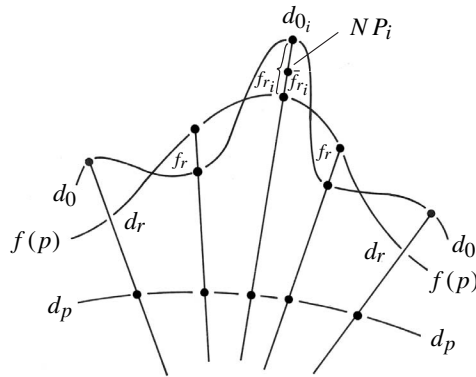
Figure 8.10. Formation of normal points

In the third step, the observed trajectory is divided into fixed intervals, so-called *bins*, starting from 0^h UTC. The proposed interval sizes for various satellites are for example:

GPS, GLONASS	300 seconds,
LAGEOS-1,2	120 seconds,
STARLETTE, STELLA	30 seconds,
ERS-1/2	15 seconds, and
GRACE	5 seconds.

In each interval, i, a mean value of all deviations, $\bar{f}_{r_i}$, is formed and added to the trend function at the center of the interval. This point, NP_i (Fig. 8.10), is called the *normal point*, and represents all single observations of the particular interval. The observation, d_{0_i}, with the fit residual, f_{r_i}, nearest to the mean epoch of the accepted fit residuals in bin i, leads to the normal point range, d_{NP_i}:

$$d_{NP_i} = d_{0_i} - (f_{r_i} - \bar{f}_{r_i}). \tag{8.7}$$

The discrepancies between the single residuals, f_r, with respect to the mean, $\bar{f}_r$, are used to determine the observation noise of the single distance measurement. The precision of the mean laser range in the normal point (8.7) is used as the characteristic measure of the internal accuracy of the laser ranging equipment. It is called the *normal point precision*, and is about ± 1 to 2 cm for the third generation SLR configurations. For modern systems, like the SLR 2000, a normal point precision better than 3 mm is expected (Degnan, 2000). Systematic effects are not included; they have to be estimated in the subsequent adjustment model.

Summarizing, the following aspects have to be emphasized when forming normal points:

 − the essential information of the raw measurements is maintained,
 − outliers are eliminated from the data,

  − the remaining correlation between normal points is insignificant, and
  − the observation noise is removed.

Normal points are also referred-to as "quick-look data" because they are generated very shortly after the satellite pass and can be used, together with equivalent data from other stations, for rapid orbit prediction. Today, normal point data are the primary product of SLR stations. They have, in most cases, replaced the full-rate data.

8.5 Applications of Satellite Laser Ranging

Due to the very high accuracy potential of laser range observations to satellites a broad field of applications in geodesy and geodynamics opened early on. Fig. 8.11 gives an
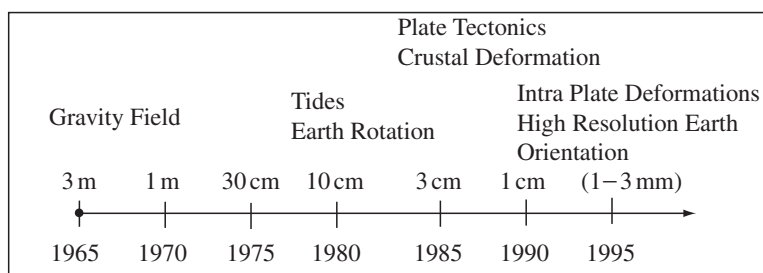
Figure 8.11. SLR accuracy development and application fields

overview of the development related to the achievable accuracy. The main fields of application are in:

Gravity field and satellite orbits [8.5.3]:	precise determination of low degree and order coefficients; tailored Earth models for particular satellite orbits; precise orbit determination (POD);
Positions, position changes and reference frames [8.5.4]:	absolute geocentric coordinates; absolute heights; contribution to ITRF, crustal deformations;
Earth Orientation Parameters (EOP) [8.5.5]:	polar motion, variation of Earth rotation, LOD; and
Particular applications [8.5.6]:	tides, precise time transfer, relativity.

8.5.1   Realization of Observation Programs, International Laser Ranging Service (ILRS)

Progress in the different tasks listed above is only possible through international cooperation and by the use of data from globally distributed stations. This is why, from the beginning of SLR activities, a close cooperation developed between the agencies and

groups responsible for SLR stations. About 20 fixed stations, and several mobile systems, contributed permanently to the NASA Crustal Dynamics Project [12.4.1]. About 30 stations participated with SLR equipment within the framework of the MERIT campaign (*M*onitor *E*arth *R*otation and *I*ntercompare the *T*echniques) (Moritz, Mueller, 1987). The International Earth Rotation Service (IERS), during the first years after its establishment in 1988, was primarily based on continuous input from about 25 laser-sites and still uses SLR data [8.5.5], [12.4.2]. Several regional groups have worked together with dedicated objectives, for example the MEDLAS (*Med*iterranean *Las*er ranging project) group within the WEGENER (*W*orking Group of *E*uropean *Geo*-scientists for the *E*stablishment of *N*etworks for *E*arthquake *R*esearch) framework.

The meaningful use of the observation results is only possible if international standards are agreed upon for data production, data reduction, and data analysis. Such standards were formulated in 1983 with the *MERIT Standards* (Melbourne et al., 1983), and they are maintained and updated as necessary with the *IERS Standards* and now the *IERS Conventions* (McCarthy, 2000).

Today, the role of laser ranging for some products and applications has decreased due to the strength of other technologies. This holds in particular for the determination of recent crustal motion in regional projects, where GPS is much more efficient, or for the analysis of high resolution Earth orientation parameters, where VLBI and GPS are of increasing importance. On the other hand, SLR data are mandatory for the determination of absolute coordinates; they still form an essential part in gravity field models; they are a backup system and sometimes the only means for precise orbit determination; and they play an increasing role in various scientific space experiments.

With the objective to concentrate international efforts in the field of satellite and lunar laser ranging, the *International Laser Ranging Service* (ILRS) was established in 1998 as a service of the IAG. The objectives and organization of the ILRS are similar to the IGS [7.8.1]. Following the *Terms of Reference* (ILRS, 2000), the "ILRS provides global satellite and lunar laser ranging data and their related products to support geodetic and geophysical research activities as well as IERS products important to the maintenance of an accurate ITRF".

The ILRS collects, archives and distributes SLR and LLR observation data sets of sufficient accuracy and uses the data to generate data products, including
- Earth orientation parameters,
- station coordinates and velocities,
- time-varying geocenter coordinates,
- static and time-varying coefficients of Earth's gravity field,
- centimeter accuracy satellite ephemerides,
- fundamental physical constants,
- lunar orientation parameters, and
- lunar ephemerides and librations.

The organizational components of the ILRS are, besides the Governing Board and the Central Bureau:
- Tracking Stations and Subnetworks,

- Operations Centers,
- Global and Regional Data Centers,
- Analysis and Associate Analysis Centers, and
- Permanent and Temporary Working Groups.

Detailed information on the ILRS can be found in the annual reports and the ILRS website. The global SLR network (about 40 stations in 2002) is reflected in Fig. 8.13. At the moment, we can distinguish three regional subnetworks:

- the European Laser Network (EUROLAS) incorporating the European stations,
- the NASA network in North America, with some stations in South America, South Africa and the Pacific,
- the Western Pacific Laser Tracking Network (WPLTN) encompassing Japan, China, Eastern Russia and Australia.

According to the "System Performance Standards" [8.3.4], the ILRS tracking stations are divided into three categories:

- *Core Stations*, meeting the highest standards of performance,
- *Contributing Stations*, contributing significantly to ILRS goals, and
- *Associate Stations*, presently providing intermittent, variable quality data.

By the end of 2002 about two thirds of the total number of ILRS stations belong to the first two categories.

An essential prerequisite for sufficient data points and high quality data are good predictions of satellite passes. Some of the Associate Analysis Centers serve as Prediction Centers and provide so-called *Inter-Range Vectors* (IRV) or *Tuned Inter-Range Vectors* (TIV) to the stations. Prediction centers compute precise orbits and extrapolate them forward. An IRV file is derived from the predicted orbit and contains position and velocity of the satellite for a given epoch, say 00:00 UT each day. The IRV are tuned such that a simple orbit integrator is capable to generate a prediction file at the particular tracking station. The *prediction file* contains altitudes, azimuths, ranges, and velocities at close intervals (e.g. every minute) and serves to control the observation process (telescope motion and detector gating). Along-track errors can be easily detected and modeled on-site as a time bias. For low orbiting satellites, like CHAMP or GRACE, more sophisticated force models and shorter tuning intervals (e.g. 6 hours) may be necessary (Wood, 1999).

Satellites are tracked following an ILRS *Tracking Priority List*. The general rules are that priorities decrease with

- increasing orbital altitude, and
- increasing orbital inclination (at a given altitude).

Particular satellites can be supported by higher priority, namely

- active missions (such as altimetry),
- special campaigns (such as the tandem mission ERS-2–ENVISAT), and
- post-launch intensive tracking phases.

Current lists are available from the ILRS website. As of January 2003, a total of 22 satellites were included in the tracking priority list.

8.5.2 Parameter Estimation

In principle, two different concepts can be used, namely geometrical and dynamical methods [1.2]. The *geometrical method* can only be applied for the determination of positions and baselines. Basically, simultaneous range measurements from at least four ground stations to a target satellite have to be carried out at identical epochs. The distance between the participating ground stations can be derived in the concept of spatial trilateration (cf. Fig. 1.2, p. 3), or new stations can be related to a network of existing control points.

The method corresponds to the classical SECOR technique [4.4.1]. It has conceptual advantages, because no assumptions are needed, for example in orbit modeling. However, from the practical point of view it cannot be applied, because weather conditions do not allow rigorous simultaneous observations at four or more stations. The larger the station separation, the smaller the probability of meeting favorable weather conditions at the same time. Experiences from the geometrical BC4-network [5.1.5] demonstrate that common observations are very rare at three stations, and nearly impossible at four. Consequently, the pure geometric method of laser ranging is more of theoretical interest, and has never been applied in practice. For model calculations see Campbell et al. (1973).

In the *dynamical method* all observed ranges can be used. The motion of the satellite is described with an adequate orbital model and relates all observations to each other. To exploit the high accuracy level of the observations, all forces acting on the satellite have to be carefully modeled, and the rotational behavior of Earth, with respect to the orbital plane, has to be known. The satellite motion refers to Earth's center of mass, hence geocentric coordinates are determined.

It is clear, for the dynamical method, that the determination of station coordinates is not an isolated problem. Station coordinates have to be estimated together with other quantities in the course of a general parameter estimation process [4.1]. Possible parameters are:
   − geocentric station coordinates,
   − gravity field coefficients,
   − pole components,
   − Earth rotation and universal time (UT1),
   − model parameters of Earth and ocean tides, and
   − additional parameters for the description of the satellite orbit.

It is not generally possible to derive all parameters of interest from the same set of observations, because the solution system may become unstable (cf. [4.1]). Usually, the coefficients of Earth's gravity field will be held fixed in the estimation of station coordinates, or the station coordinates will be treated as known quantities in the determination of Earth rotation parameters. Hence we have two groups of parameters in dynamic modeling:
  (a) parameters contained in the solution, and
  (b) adopted parameters, not contained in the solution.

Today the dynamic approach is almost exclusively used, based on all available tracking data from the global SLR network.

In all parameter estimation processes the necessity arises to fit a precise trajectory to the observed data. Usually, the SLR analysis is performed in several steps, (e.g. Devoti et al., 2001). In the first step, satellite orbits are reduced piecemeal, solving for arc dependent parameters like the state vector, non-gravitational forces and measurement biases. The arc length is shorter for low orbiting satellites (e.g. five days for STARLETTE, STELLA, ERS-2) and longer for high orbiting satellites (e.g. between 1 week and 1 month for LAGEOS). In a second step, the arc solutions are combined in a multi-arc solution, and global parameters are estimated, such as coordinates, Earth orientation parameters and coefficients of the gravity field. In a final step, very long arcs, over many years, are analyzed to verify fundamental physical models or to solve for the secular drift of certain parameters like low order zonals.

Two particular effects were derived rather early from analysis of LAGEOS orbits over many years. These are a secular nodal acceleration and an unexpected decrease in the semi-major axis at the submillimeter/day level. The nodal acceleration is related to a secular change of J_2 and reflects a decrease of Earth's flattening. This effect can be explained by relaxation of the Earth since the last glaciation. The decrease of the semi-major axis is mostly explained by thermal effects on the corner-cube reflectors caused by Earth's infrared radiation (Rubicam, 1986).

The possibilities and techniques for orbit modeling have been continuously improved since the launch of the first laser satellites. A 1 month arc of the LAGEOS orbit can be modeled with about ± 1 to 2 cm accuracy. For lower, and hence more disturbed satellites, like STARLETTE or ERS-1/2, the accuracy in orbit modeling is about ± 5 cm or slightly better. Fig. 8.12 demonstrates the improvement over about 15 years in the modeling of 1 month LAGEOS arcs (Pavlis et al., 1991).
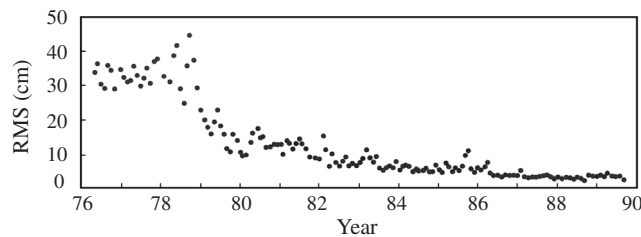
Figure 8.12. Accuracy improvement in the modeling of 1-month LAGEOS arcs

8.5.3   Earth Gravity Field, Precise Orbit Determination (POD)

Because of their high accuracy, laser distance measurements to satellites have been included in the computation of Earth models since the launch of the first satellites

equipped with retro-reflectors (cf. [12.2]). The last gravity field model in the pre-LAGEOS era, computed by the NASA Goddard Space Flight Center, was named GEM 9 and contained about 200 000 laser ranges to 9 satellites. The model was developed up to degree and order 20 (Lerch et al., 1979). Because of the decreasing sensitivity of satellite orbits to smaller gravity anomalies it is necessary to incorporate results from the direct mapping of the gravity field (satellite altimetry [9], satellite-to-satellite tracking, gradiometry [10], or surface gravity data) into the solutions with higher order coefficients. It can be stated, as a general rule, that the limit of resolution of gravity field structures by orbit analysis is within a wavelength of about 1000 km. The influence of high-frequency terms in the gravity field on the satellite orbits decreases with increasing height.

Because of this relationship between gravity field development and satellite orbital height, it is very important to know precisely the low frequency components of the gravity field for the exact orbit modeling of dedicated satellites like STARLETTE, STELLA, LAGEOS, or ETALON, in order to meet the requirements of geodynamical research and reference frame stability. Vice versa, LAGEOS orbit analysis permits the isolation of long wavelength geopotential signals within gravity field solutions, because LAGEOS is rather insensitive to the gravity field above degree 10, and is unaffected by terms above degree 20 (Klosko, 1999).

Dedicated gravity fields of this type are called *tailored gravity fields*. The GEM-L2 solution (Lerch et al., 1983) is an early such tailored field for LAGEOS orbits; the related long-wave geoid can be modeled to degree and order (4,4), with an accuracy level of 8 cm. An equivalent tailored gravity field has been designed for STARLETTE with the PGS-1331 model up to degree and order 36 (Marsh, Lerch, 1985). This model is also of value for other missions at a similar orbital height, such as STELLA, ERS-1 and ERS-2.

With the availability of new precise laser ranging data to LAGEOS and other satellites, and with the requirements for a precise modeling of the TOPEX/POSEIDON altimeter mission, a new GEM-series of satellite based long wavelength gravity field models was started in 1987. The model GEM-T1, was exclusively based upon direct satellite tracking observations. It is complete to degree and order 36 (Marsh et al., 1987), and contains about 440 000 laser observations. The follow-up model GEM-T2 (*Marsh* et al. 1990) was improved by additional laser observations to LAGEOS, STARLETTE, and AJISAI, as well as by older arcs of GEOS-1 and GEOS-2. GEM-T3 (Lerch, 1992) was complete to degree 50, using tracking data from 31 satellites, and in addition altimeter data from GEOS-3, SEASAT and GEOSAT.

The availability of laser targets in low altitudes, like GFZ-1, gave rise to the development of higher order satellite-only gravity models. About 74 000 laser data to GFZ-1 together with 2.8 million older tracking observations were used to estimate the gravity field model GRIM4-S4G complete to degree and order 60 with higher degree terms (up to 100) in zonal and resonant bands (König et al., 1999).

Many more gravity field models have been developed where the SLR data form a substantial part of the data base, in particular for the long-wavelength part. Frequently

used models are the *Joint Gravity Model 3*, JGM-3 (Tapley et al., 1996), a tailored model for TOPEX/POSEIDON and complete to degree 70, and the *NASA* and *NIMA Joint Geopotential Model* EGM96 (Lemoine et al., 1998), complete to degree 360. An excellent overview of historical and current models is given by Rapp (1998); see also [12.2]. The realization of the importance of SLR as a long-lived "passive" tracking technique for the estimation and continuous improvement of gravity models is enlightened by the fact that several old, long abandoned satellites like BE-C, D1-C, D1-D or GEOS-3 have been included in new international tracking programs (Klosko, 1999; ILRS, 2000).

Laser ranging to geodetic satellites with stable orbits (in particular LAGEOS-1 and LAGEOS-2) is useful to measure the evolution over time of the long wavelength part of the gravity field. Several authors have reported time derivatives of the zonal coefficients J_2 to J_6 (Devoti et al., 2001). So far, only the term $\dot{J}_2$ could be determined significantly:

$$\dot{J}_2 = -2.6\ldots 3.0 \pm 0.2 \cdots \pm 0.5 \cdot 10^{-11} / \text{yr}.$$

The effect can be related to post-glacial rebound, the ongoing mass redistribution following Pleistocene deglaciation in the northern hemisphere.

LAGEOS is particularly suitable for the determination of the geocentric gravitational constant GM (cf. [12.2.2]), because of its fairly undisturbed orbit and the high ranging accuracy. The value of GM is estimated each time as part of a global solution. The precision of the estimate has improved by an order of magnitude in each of the last two decades (Smith et al., 2000). The current value from recent LAGEOS estimations is

$$GM = 398\,600.44187 \pm 0.00020 \, [\text{km}^3/\text{sec}^2]. \tag{8.8}$$

The results are confirmed by estimates from SLR data from STELLA, STARLETTE, AJISAI and ETALON, however with a much higher scatter. The value of GM defines the scale in satellite orbit determination. The value from recent LAGEOS observations is about 1ppb higher than the previously adopted value from SLR observations (Ries et al., 1992). This difference corresponds to a difference in orbital height of about 3 mm (Smith et al., 2000).

For further discussion on gravity field determination from satellite data see [12.2], or Torge (2001).

Precision Orbit Determination (POD) is one of the most important applications of today's SLR technology. Based on a tailored gravity field for a particular satellite, and an appropriate dynamical model, all available SLR data from the ILRS tracking network are used to estimate a precise orbit. The unique feature of SLR data is that the orbits are absolute in the sense that they refer to Earth's center of mass. In many cases the orbit determination is based on SLR and additional tracking systems such as GPS, DORIS or PRARE. In some cases, SLR is the only tracking device because of a failure of the primary tracking system, such as PRARE for ERS-1, or GPS for GFO (GEOSAT Follow On). SLR is hence an excellent backup system with an extremely long lifetime that survives all other tracking systems.

In combined orbit determination, for example DORIS and SLR for TOPEX/POSEIDON, the DORIS data provide the main contribution to the overall orbit accuracy, and SLR contributes in the crucial centering of the orbit. Centering errors in the absolute height of altimeter satellites would introduce asymmetry in the estimated sea surface variations and hence corrupt the oceanographic interpretation. Recent studies indicate a POD accuracy for TOPEX/POSEIDON of 2 to 3 cm in the radial direction (ILRS, 2000).

The high accuracy of SLR determined orbits is of eminent importance for the absolute calibration of sensor errors in new missions. This is in particular true for the radial altimeter errors, such as in the JASON-1 or ENVISAT-1 mission [9.2]. SLR also contributes, in combination with GPS, to the precise orbit determination of gravity field missions like CHAMP and GRACE [10.2]. The high value of the SLR tracking data for orbit determination is illuminated in the long list of tracking priorities and SLR missions set up by the ILRS [8.5.1]. For the technique of precise orbit determination including SLR data see [3.3] and e.g. Rim, Schutz (1999); Montenbruck, Gill (2000).

8.5.4 Positions and Position Changes

The dynamical modeling of satellite laser range data offers the possibility of estimating geocentric three-dimensional positions. If gravity field parameters form part of the solution, the coordinates refer conceptually to the Earth's center of gravity. Today, in most cases, a tailored gravity field model is used, such as JGM-3. The scale is introduced through the velocity of light and the adopted GM value.

During the early years of satellite laser ranging the technique was mainly used for the determination of crustal motion along selected baselines or in regional networks. One example is the continuous monitoring of crustal deformation along the San Andreas Fault. For a 400 km baseline between Quincy and Monument Peak a significant deformation of about 6 cm ± 3 mm/year could be detected (Watkins et al., 1990). Another example is the WEGENER/MEDLAS project in the Mediterranean. Three transportable laser systems have provided accurate epoch positions for sites in Italy, Greece, and Turkey since 1985. The apparent motions, with respect to a coordinate system that is rigidly attached to the Eurasian tectonic plate, reach 20 to 40 mm/year (Ambrosius et al., 1991).

Today, particular SLR campaigns are no longer arranged; instead the continuously available tracking data from the global ILRS network are used to estimate coordinates and coordinate changes. Some of the ILRS analysis or associate analysis centers do this on a regular basis (ILRS, 2000). The usual procedure is firstly to compute weekly solutions (or from similar short intervals up to one month) for the total time span of analysis. These solutions are used to clean the data and they offer a valuable insight into the quality of the station data. The variations of the weekly coordinates are in the order of 2 cm for high performance laser stations. In a second step, improved weekly or monthly solutions for coordinates and velocities are generated which are combined in a final adjustment over the total time span of, say, several years. Fig. 8.13 shows as an example the results of a 10 years global solution for 40 global SLR stations based

on tracking data to LAGEOS-1/2. The estimated accuracy is 6 mm for the coordinates
and 2 mm/a for the velocities (Angermann et al., 2001).
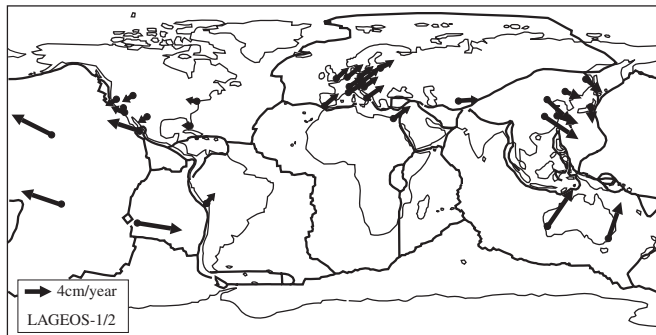
Figure 8.13. Station velocities of a combined LAGEOS-1/2 solution from 10 years of LAGEOS
data, cf. Angermann et al. (2001)

    One disadvantage of SLR solutions is the rather inhomogeneous distribution of data
quantity and data quality. For a given orbital arc, some stations only contribute 10% of
the data of other stations. This is why, in most cases, SLR solutions are combined with
solutions from other space techniques [12.1]. This holds in particular for the various
realizations of the ITRS [2.1.2.2]. Five different SLR solutions were included in the
ITRF97 alongside four VLBI, six GPS, and three DORIS solutions (Boucher et al.,
1999). In total, 10 different SLR solutions were included in the ITRF2000 solution.
The main contribution of SLR to ITRF2000 is seen as follows (IERS, 2001):

  − the origin and its rate are defined by a weighted average of the most consistent
    SLR solutions, and
  − the scale and its rate are defined by a weighted average of VLBI and the most
    consistent SLR solutions.

Absolute coordinates refer to Earth's center of mass, including the oceans and at-
mosphere. Analyses of SLR data have shown that the ensemble of tracking stations
on Earth's crust is always moving with respect to the center of mass. This motion,
regarded from the crust-fixed stations, is called *motion of the geocenter* (Rothacher,
2000b). The detected variation is below 1 cm and shows annual and semiannual
periods. It is mainly caused by mass movements in the atmosphere and the oceans.

8.5.5   Earth Rotation, Polar Motion

The stability of the LAGEOS orbits provides an excellent external reference frame for
Earth-based observations of Earth's orientation. *Earth Orientation Parameters* (EOP)
or *Earth Rotation Parameters* (ERP) are the

  pole coordinates x_p, y_p,
  Greenwich Apparent Sidereal Time   GAST = Θ.

The pole coordinates (cf. Fig. 2.6, p. 20) are defined as the difference between the actual orientation of Earth's rotational axis (instantaneous pole) and an agreed mean orientation (Conventional Terrestrial Pole CTP) [2.1.2.3]. The Greenwich sidereal time, GAST, is equivalent to the universal time UT1 [2.2.2]. Variations in the rate of Earth rotation can be described by the difference UT1−UTC. They are also expressed by the period of one complete Earth revolution about its axis (*length of day*, LOD).

Due to polar motion and daily fluctuations in the rotational velocity, any coordinate system fixed to the Earth (Earth-fixed system) experiences variations with respect to a space-fixed (inertial) reference frame. These variations can be described, cf. (2.24), as

$$X_{\text{CIS}_i} = R_3(-\Theta)R_1(y_p)R_2(x_p)X_{\text{CTS}},$$ (8.9)

in which X_{CIS_i} and X_{CTS} are the position vectors of arbitrary points in the instantaneous space-fixed or conventional Earth-fixed reference frame, respectively. The well described orbits of laser satellites can be regarded as a realization of the inertial reference system.

(a) epoch t (b) epoch $t + \Delta t$
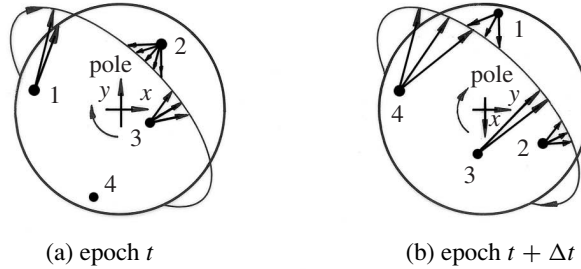
Figure 8.14. Determination of pole coordinates from satellite laser ranging

Fig. 8.14 explains how the instantaneous pole coordinates can be derived from distance measurements to satellites. The instantaneous pole position is identified from a several days long satellite arc as the point around which the observation stations rotate beneath the "stable" satellite orbit. The fluctuations of Earth rotation are accordingly analyzed from the range residuals. For the individual distance measurement we find the observation equation, e.g. (Montag, 1984):

$$ds = \frac{rR}{s}(\sin\Phi(\cos(\Theta - \lambda - \Omega)\cos u + \sin(\Theta + \lambda - \Omega)\sin u \cos i)$$
$$- \cos\Phi \sin u \sin i)d\Phi + \frac{rR}{s}\cos\Phi(\sin(\Theta + \lambda - \Omega)\cos u$$ (8.10)
$$- \cos(\Theta + \lambda - \Omega)\sin u \cos i)(d\Theta + d\lambda - d\Omega),$$

with

$$d\Phi \qquad = \cos\lambda\, dx_p - \sin\lambda\, dy_p, \text{ variation in the geographic latitude,}$$
$$d\lambda \qquad = (\sin\lambda\, dx_p + \cos\lambda\, dy_p)\tan\Phi, \text{ variation in the geographic longitude,}$$

s topocentric satellite range,

r, R geocentric ranges to satellite and ground station,

Θ sidereal time, and

Ω, i, u orbital elements, $u = \omega + v$.

Equation (8.10) describes the relationship between the variations of Earth orientation parameters, dx_p, dy_p, $d\Theta$, and the resulting range variations to the satellites.

Earth orientation parameters derived from satellite laser ranging have played an important role since about 1980 because they were at least one order of magnitude more precise than the classical astrometric and Doppler techniques. As a result of the MERIT campaign (Mueller, Wei, 1985), cf. [12.4.2], SLR data were routinely introduced into the EOP products of the BIH and, since 1988, of the IERS. Fig. 8.15 gives an impression on the data quality around 1980.

Under the framework of the ILRS, EOP parameters are estimated on a routine basis as one of the operational data products. The accuracy is of the order

0.1 mas [milliarcseconds]   for the pole coordinates x_p, y_p, and

0.05 ms [milliseconds]   for Earth rotation (UT1-UTC).

The contribution of SLR to the generation of Earth orientation parameters within the IERS has been substituted more and more by VLBI and GPS because of their weather independence and higher temporal resolution. As of 2000, the percentage of SLR contribution to the polar motion components was only 10, compared with 20 for VLBI and 70 for GPS. UT1-UTC is exclusively based on VLBI (IERS, 2001).

The unique advantage of SLR data, when compared with other space techniques, is the steadily increasing length of homogeneous data series since the launch of LAGEOS-1 in 1976. The long-term analysis of Earth's rotation and orientation reveals changes in the distribution of mass and exchange of angular momentum in the Earth system.
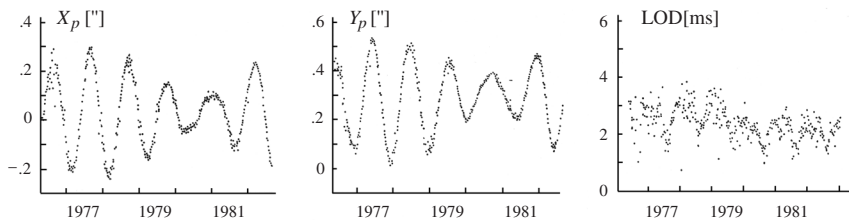
Figure 8.15. Earth rotation parameters from LAGEOS observations (1976–1982), (Smith, et al., 1985)

8.5.6 Other applications

Some additional uses of satellite laser range observations are summarized here.

Solid Earth and Ocean Tides
In precise satellite tracking and data analysis three tidal effects caused by lunar and solar gravitation have to be considered. These are the
   (1) direct perturbation of the satellite orbit relative to the adopted terrestrial reference frame,
   (2) tidal deformation of the oceans and of the solid Earth, and
   (3) gravitational effects of these deformations on the satellite orbit.

The direct orbital perturbation (1) is by far the largest effect, but it can be modeled very accurately [3.2.3.2].

Modeling of Earth's *tidal response* to lunar and solar forces is rather difficult because of the lack of detailed knowledge of Earth's internal structure. It is usually described by elasticity parameters (*Love's numbers, k*). The periodic deformation of the Earth (2) causes a purely geometrical periodic change of range between the observation station and a given satellite. The radial component (*tidal uplift*) reaches 30 to 40 cm, and has to be modeled in the parameter estimation process. The effect is similar for adjacent stations. For larger station separations the differential effect has to be considered.

The gravitational effect of Earth's tidal response on satellite orbits (3) can be modeled for near-Earth satellites along with the orbit analysis from precise laser ranging. STARLETTE and STELLA are particularly well suited for this purpose (Williamson, Marsh, 1985). The effect of the tidal induced gravitational potential on satellite orbits can reach several meters. The analysis supplies a better knowledge of the elasticity parameters and hence leads to a better model of Earth's tidal response. From STAR-LETTE (and STELLA) orbital data the Love's number, k_2, can be derived with an accuracy of 1% for the leading tidal terms, and the amplitude and phases of the main ocean tidal parameters can be determined (Williamson, Marsh, 1985).

Tidal parameters, derived from satellite orbits, are global in character. Terrestrial methods have a higher resolution, but they may be influenced by local effects.

Precise Time Transfer
Laser ranging from two or more stations to a properly equipped satellite can be used for time comparisons of remote atomic station clocks. The satellite carries an active on-board package, capable of detecting and dating a laser pulse. For time comparison, two laser stations fire to the satellite so that the two beams arrive very close in time. The on-board oscillator measures the interval between the arrival times of each laser pulse. Together with the ground measured round-trip times from the stations to the satellite, it is possible to compare the station time scales. The simple equation is

$$\Delta T = t_A - t_B + \tau_A - \tau_B + R, \qquad (8.11)$$

with

t_A epoch of range measurement at station A,

t_B epoch of range measurement at station B,

τ_A pulse travel time from station A to the satellite,

τ_B pulse travel time from station B to the satellite,

R interval between pulse arrival times at the satellite, and

ΔT epoch difference between time scales at both stations.

A dedicated experiment *LASSO* (*LA*ser *S*ynchronization from a *S*tationary *O*rbit) was proposed already in 1980 and tested by the end of 1992. The accuracy expectation was 100 ps, but could not be verified because of bad weather conditions (Lewandowski et al., 1999). A future generation of LASSO, designated *T2L2* (*T*ime *T*ransfer by *L*aser *L*ink), is expected to provide an uncertainty of 50 ps or better (Samain, Fridelance, 1998). Because of their sensitivity to weather conditions, LASSO and T2L2 are not suited for operational use; but are excellent tools for assessing the accuracy of GPS or GLONASS time transfers (cf. [7.6.2.9]).

Fundamental physics
SLR will support research in fundamental physics. As stated in [8.5.3] SLR measurements of LAGEOS have provided the most accurate values of GM, and have confirmed that GM does not change with time. SLR also contributes to tests of theories of gravitation. As soon as a third LAGEOS satellite is launched into an orbit with a supplementary inclination to that of either LAGEOS-1 or LAGEOS-2, a pair of satellites would be sensitive to the *Lense-Thirring Precession* or *frame-dragging* effect on a satellite orbit, that is the orbit plane is "dragged" in the direction of Earth's rotation. This important test of general relativity, as well as other tests of relativistic formulations, could be made with such a satellite pair to the 3 to 4 percent level (Beutler et al., 1997).

Precision tracking application
Satellites equipped with retro-reflectors benefit from SLR when active tracking systems fail (e.g. ERS-1). In addition, SLR provides an excellent opportunity to calibrate active tracking systems and to estimate biases like center-of-mass corrections or antenna offsets. This is possible for all GLONASS spacecraft, but for the time being only for two GPS satellites. The forthcoming GALILEO spacecraft will also carry retro-reflectors for independent orbit determination,

For scientific satellites without an active tracking system, like the Tether Physics and Survivability Experiment (TIPS), SLR forms an essential part of the experiment.

8.6 Lunar Laser Ranging

Since 1969 it has been possible to determine precise distances between Earth and the Moon by laser ranging techniques. In the course of the manned American space missions three reflector assemblies were installed on the lunar surface and pointed toward Earth (Fig. 8.16):
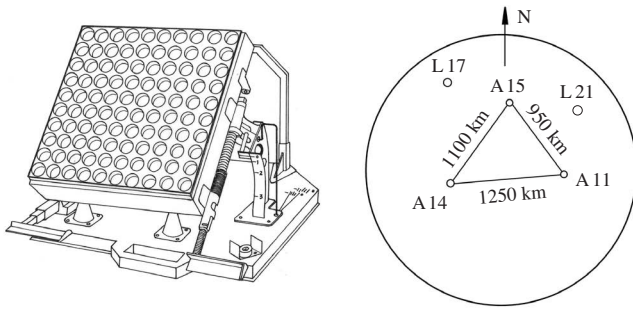
Figure 8.16. Laser reflectors on the Moon; Apollo 11 reflector and distribution of targets

July 1969	Apollo 11	100 single reflectors	(Sea of Tranquillity),
February 1971	Apollo 14	100 single reflectors	(Fra Mauro), and
July 1971	Apollo 15	300 single reflectors	(Hadley Rille).

The three reflector assemblies form a triangle with side lengths of 950, 1100, and 1250 km, and are well distributed in latitude and longitude. They are hence well suited for the separation of the lunar libration components. The assembly is completed by two French reflectors L17 (Sea of Rains) and L21 (Sea of Serenity) which were deployed by two Soviet automatic lunar missions in November 1970 and January 1973. The reflector L21 is regularly included in *Lunar Laser Ranging* (LLR) programs. The reflector L17 gives no return signals because it may have been covered with dust from the departing spacecraft. The reflector with highest priority is A15.

The Moon can be regarded as a highly stable satellite with a precisely modeled orbit and a long data series of more than 30 years. Valuable insights into the dynamics of Earth as well as the dynamics of the Earth-lunar-system can be derived from the analysis of continuous LLR observations.

Laser ranging to the Moon is technically much more challenging than satellite laser ranging. The energy balance is very weak. Even with a laser firing rate of 10 Hz less than a few tens of photoelectrons per minute out of the 10^{19} per second transmitted are routinely received. This corresponds to an overall signal loss of approximately 10^{-21} (Dickey et al., 1994; Shelus et al., 1996). To aim at the reflector on the Moon the required pointing accuracy is about $2''$. A very short time interval of $\Delta t = 200$ ns is necessary to filter the return signal from the disturbance background. The time interval of 200 ns corresponds to a necessary prediction accuracy of ± 15 m for the lunar distance.

Because of these extremely demanding requirements only very few observatories can successfully measure to the Moon. The only observatory with continuous ranging since 1970 is the *McDonald Observatory* in Western Texas. During the first 15 years after the deployment of the Apollo 11 reflector array, it was also the only facility worldwide that routinely ranged to the Moon. The 2.7 m reflector telescope, however,

was mainly used for other astronomical purposes. In the mid-1980s a transition was made to the dedicated 0.76 m *McDonald Laser Ranging System* (MLRS), capable of ranging to artificial satellites as well as to the Moon.

Since 1984, another dedicated LLR station is continuously ranging to the Moon, namely a French station near Grasse: *Observatoire de la Côte d'Azur - Centre d'Etudes et de Recherche en Géodynamique et Astronomie* (OCA/CERGA). Since about 1985 other observatories have also been successfully contributing to Lunar Laser Ranging for limited time periods. These are *Haleakala* on the Island of Maui (Hawaii), the station *Orroral* in Australia and the German fundamental station *Wettzell*. Another joint SLR/LLR station is being built up in *Matera*, Italy (Matera Laser Ranging Observatory, MLRO). The operational *Lunar Laser Ranging Network* (LLRN) of the ILRS hence, for the time being, only consists of two stations.

The earliest LLR ranges had accuracies of several meters and were improved to 20 cm during the 1980s. Current measurement accuracies at the two LLRN observatories are about ± 1 to 3 cm. This corresponds to a relative accuracy of better than one part in ten billion ($1:10^{10}$). Sub-centimeter normal point accuracy is aimed for. Because of the high measurement accuracy, it is necessary to formulate the analysis models in post-Newtonian approximation.

The geometric relationship in the lunar laser ranging technique is explained in Fig. 8.17. The basic observable is the range, ρ, between the Earth-based observatory, O, and the reflector, R, on the lunar surface. E is the terrestrial center of mass and M the lunar center of mass. B is the barycenter of the solar system. The ephemerides of the Earth and lunar orbits refer to B.
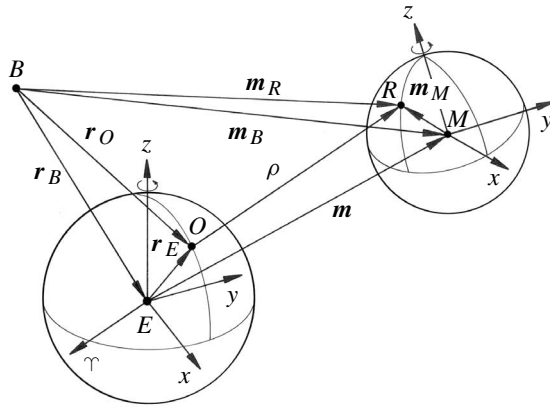
Figure 8.17. Geometrical relationship in lunar laser ranging

The equation linking the coordinates of the telescope and of the reflector is written in the mean heliocentric (barycentric) coordinate system as

$$\boldsymbol{r}_O - \boldsymbol{m}_R = \boldsymbol{\rho}. \tag{8.12}$$

Equation (8.12) is only fulfilled if several corrections are applied. The coordinates, r_E, of the telescope, written in the Earth-fixed reference system, differ from the coordinates in the barycentric system because of

- Earth rotation,
- pole coordinates,
- precession, and
- nutation.

The reflector coordinates, m_R, expressed in the barycentric system, have to be corrected for lunar motions, for example libration. The measured ranges, finally, are influenced by tides, aberration and other relativistic effects, as well as by variations of the station coordinates due to crustal motion. The modeling of the whole process hence constitutes a rather complicated problem of parameter estimation. Dickey et al. (1983) report on more than 80 Earth-Moon parameters to be introduced into the model. Basic models are given in the early literature, for example Stolz (1979); Ballani (1982). For an analysis of LLR observations in the concept of a *post-Newtonian* theory see e.g. Müller (1991); Nordtvedt (2001).

The long series of what is now more than 30 years of continuous data provides an excellent opportunity for long-term, as well as short-term, studies of the behavior of the Earth-Moon system. The LLR technique contributes and/or is expected to contribute to, among others, the following problems (Dickey et al., 1994; Soffel, Müller, 1997):

Global parameters of the Earth-Moon system
- geocentric coordinates of the tracking stations, including drift rates,
- selenocentric reflector coordinates,
- lunar orbit,
- lunar rotation (libration),
- low harmonic coefficients of the lunar gravity field,
- combined mass of Earth and Moon,
- tidal friction (momentum exchange between Earth and Moon),
- Love number of the Moon, and
- control of precession and nutation theories for a deformable Earth.

Earth rotation
- universal time (UT0), length of the day (LOD),
- polar motion, and
- long-term variation of the Earth rotation.

Gravitational physics and relativity
- test of Newton's law of gravitation (possible $\dot{G}/G$),
- test of the equivalence principle (general relativity),
- principles of special relativity (e.g. Lorentz contraction), and
- verification of the geodesic precession.

Within an iterative process, some of the parameters are held fixed in the solution or taken from other sources. For example, Earth orientation parameters may be held fixed while parameters of the Earth-Moon system are solved for. Based on the long-term series of precise ranges to the Moon it is now possible to compute a very precise ephemeris of the lunar orbit; it is precise enough to permit accurate analysis of solar eclipses as far back as 1400 B.C. From the current evolution of the orbit, it is possible to derive interesting conclusions. For instance, due to the tidal interaction, the Moon is receding from the Earth at about 3.8 cm/year.

The geocentric gravitational constant was determined from more than 20 years of lunar laser ranging as (Dickey et al., 1994)

$$GM = 398\,600.443 \pm 0.004\,[\text{km}^3/\text{s}^2]. \tag{8.13}$$

The strong influence of the Sun on the lunar orbit also permits a precise determination of the relation of masses of the Sun, Moon, and Earth (Soffel, Müller, 1997):

$$m_S/(m_E + m_M) = 328\,900.560 \pm 0.002. \tag{8.14}$$

The over thirty years of data (the longest time series available from any of the modern space techniques) are especially valuable in solving for corrections to the 18.6-year nutation terms and the precession constant (Dickey et al., 1994).
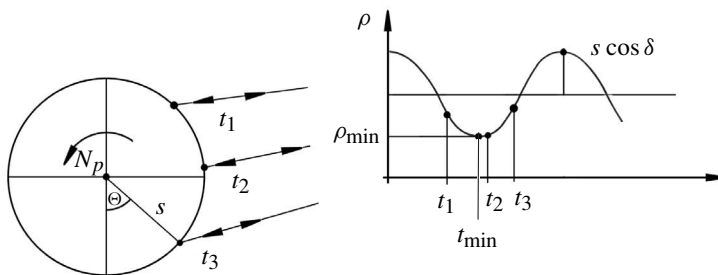
Figure 8.18. Determination of Earth rotation from lunar laser ranging

Earth rotation (UT0 and LOD) can be derived with high accuracy from LLR observations in particular. Fig. 8.18 depicts the geometrical situation. Let s be the distance of the tracking station from Earth's rotation axis. Then the range observation, ρ, must follow a cosine-function where the amplitude is a measure of s. The phase of the cosine-function is given by the moment of shortest lunar distance and hence is equivalent to the geographic longitude of the tracking station. Variations in longitude are related to variations in Earth's rotation velocity. With good modeling of the observations UT0 can be determined with a resolution of 0.05 ms.

In order to determine both pole coordinates, it is necessary to use data from at least two laser stations, sufficiently separated in geographic location. This is also true

for the determination of UT1, because Earth rotation variations cannot be separated from the pole components with only one input station (Stolz, 1979). Because of strong correlations between pole coordinates and errors in the lunar ephemerides as well as variations of UT1, lunar laser range observations are less suited for the determination of polar motion than are range measurements to artificial satellites. Within the IERS, therefore, LLR does not contribute to the determination of pole coordinates.

For results and new insights from lunar laser ranging in gravitational physics and relativity see, for example, Dickey et al. (1994); Soffel, Müller (1997). Two notable findings are that the relativistic geodesic precession of 19 mas/year is confirmed within 0.35%, and that the gravitational constant G has no detectable rate for dG/dt/G within $1.1 \cdot 10^{-12}$/year (ILRS, 2000).

8.7 Spaceborne Laser

The use of SLR equipment at a large number of terrestrial observation stations for the determination of precise coordinates is very expensive and time consuming. This is why several proposals were made early on for reversing the principle, that is to deploy the laser ranging system in an orbiting platform and to install reflectors on the ground, e.g. Mueller (1975); Kahn et al. (1980); Drewes, Reigber (1982); Cohen et al. (1990). The concept has many attractions because a dense network of ground reflector points can be installed in active tectonic areas, and be controlled on a regular basis. With the use of additional beacons in areas of tectonic stability (fiducial stations), the orbit of the spaceborne laser system can be precisely modeled.

Feasibility studies have demonstrated that spaceborne laser systems can be realized, and would provide baseline accuracies on the order of a few cm over distances from a few km to 1000 km. The concept, however, was never realized because GPS developed to be an extremely accurate and efficient tool to provide geodetic control for monitoring crustal deformation.

Instead, a spaceborne laser altimeter mission was planned and has been realized with the *Geoscience Laser Altimeter System* (GLAS) as an integral part of the NASA *Earth Observing System* (EOS) program. GLAS is the primary instrument on the *Ice, Cloud and Land Elevation Satellite* ICESAT, launched on January 12, 2003. The main scientific objective of ICESAT is to better understand the mass balance of the polar ice sheets and their contribution to sea level change. Furthermore cloud heights, topography of land surfaces, vegetation heights, and sea-ice surface characteristics will be measured (Schutz, 1998).

ICESAT flies in a near polar low Earth orbit (LEO) at an altitude of 600 km with an inclination of 94 degrees. The mission orbit sets a 183 day repeat pattern which yields 15 km track spacing at the equator and 2.5 km at 80 degrees latitude. The on-board dual-frequency GPS receiver is designed to provide 5 cm radial orbit position; SLR reflectors serve as a back-up system. On-board star cameras and gyros control the spacecraft orientation and laser pointing direction.

The GLAS instrument uses three Q-switched Nd:YAG lasers, but only one will operate at a time. The pulse length is 5 ns, the shot repetition rate 40 Hz. The laser beam, nominally in nadir direction, has a 0.110 mrad divergence and illuminates a spot on Earth's surface with a diameter of about 66 m (footprint). The surface reflected part of the signal is collected in a 1 m on-board telescope.

The laser pulse travel time provides the scalar altitude. Together with the pointing information from the orientation system and the GPS position of the spacecraft, an *altitude vector* can be determined which provides the ITRF location of the illuminated spot on the surface. The error budget is estimated as follows (Schutz, 1998):

instrument precision	10 cm,
radial orbit determination	5 cm,
pointing determination	7.5 cm,
tropospheric delay	2 cm,
atmospheric scattering	2 cm,
other	1 cm,
total	13.8 cm

The single shot error of about 14 cm enters an adjustment process as in satellite altimetry [9.4] using the crossover technique. Considering the high number of possible crossovers in high latitudes, error estimates indicate that the required accuracy of 1.5 cm/year can be met, and the surface variability of large ice sheets in Antarctica and Greenland can be determined (Schutz, 1998).

9 Satellite Altimetry

9.1 Basic Concept

Satellite altimetry is one of the more recently developed methods of satellite geodesy and has been, up to now, the only operational technique to form part of group (2) in [1.2]; namely, measurements made from the satellite to the ground (*Space to Earth*). The basic concept is very simple. The satellite is used as a moving platform for a sensor which transmits microwave pulses in the radar frequency domain to the ground, and receives the return signals after reflection at Earth's surface. The altitude, a, of the satellite above Earth's surface can be derived, as a first approximation, from the observed two-way travel time of the radar signal:

$$a = c \frac{\Delta t}{2}. \qquad (9.1)$$

Because of the favorable reflective properties of water, the method is especially suitable over the oceans. A circular area with a diameter of a few kilometers, the so-called *footprint*, is illuminated at the instantaneous sea surface, the size of which is related to the spatial resolution of the incoming microwave beam. Accordingly, the observations refer to a mean instantaneous sea surface height which differs from the geoid height by the separation, $\overline{H}$. The altitude, h, of the satellite above a global ellipsoid can be derived from an orbit computation with respect to a geocentric reference frame. If additional corrections are at first neglected, we find the basic simplified altimeter equation (cf. Fig. 9.1, Fig. 4.8, p. 144):

$$h = N + \overline{H} + a. \qquad (9.2)$$

The microwave sensor in the satellite, the *radar altimeter*, works in the frequency domain of about 13.5 GHz (Ku band), corresponding to a wavelength of 2.2 cm. The pulse length is a few nanoseconds, the related resolution of the single range measurements being several centimeters.

Fig. 9.1 demonstrates that a radar altimeter can be used to scan the sea surface directly and hence approximately the ocean geoid. Satellite altimetry is, for this reason, a very powerful geodetic tool for direct mapping of the geoid. The main significance of the method results
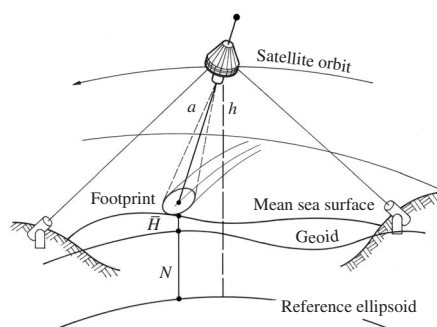
Figure 9.1. Basic concept of satellite altimetry

from the possibility of scannning large ocean areas within a fairly short time period, and determining a detailed representation of the sea surface with high resolution in space and time.

On the other hand, satellite altimetry is a very good example of the interdisciplinary nature of satellite geodesy. The quantity $\overline{H}$, namely the separation between the mean sea surface and the geoid, is a disturbance (noise) to the geodesist, who models the geoid, but constitutes the observation signal for the physical oceanographer in the study of ocean dynamics [9.5.3]. The geophysicist can, from the large scale analysis of $\overline{H}$, derive valuable insight into the structure of the ocean floor and its tectonic features [9.5.2].

A wealth of data has been obtained from existing altimeter missions, and this has led to important scientific results in geodesy, geophysics, and oceanography. Detailed discussions can be found in several dedicated issues of the *Journal of Geophysical Research* (e.g. Vol. 84, B8, 1979; Vol. 87, C5, 1982; Vol. 88, C3, 1983; Vol. 95, C3, 1990; Vol. 99, C12, 1994; Vol. 100, C12, 1995) and in the excellent handbook, edited by Fu et al. (2001).

In recent years, several new satellite altimeter missions have been launched, and further missions, carrying radar altimeters [9.2], are planned for the near future. Satellite altimetry will hence remain one of the powerful methods in satellite geodesy.

9.2   Satellites and Missions

The technique of satellite altimetry was tested for the first time during the *SKYLAB* missions SL-2, SL-3, and SL-4 (1973–1974) over orbital arcs of about 750 km length. The accuracy of the altimeter was about 1 to 2 m. This opened the way to a direct comparison of the altimeter heights with a computed $5' \times 5'$ gravimetric geoid (Marsh, Vincent, 1975). Fig. 9.2 shows this comparison for one orbital arc near the Puerto Rico Trench, where the geoidal heights change by up to 20 m over quite short distances. The correspondence is within a few meters; all the main characteristic geoidal structures are visible in the data of the SKYLAB altimeter (solid line: model geoid).

After the successful test during the SKYLAB missions, new and improved altimeter versions were flown on the early satellites GEOS-3 (1975) and SEASAT-1 (1978). *GEOS-3* (*Ge*odynamics *E*xperimental *O*cean *S*atellite, Fig. 4.11, p. 150), launched on April 9, 1975, was a multi-purpose satellite [4.3.2]. Besides the altimeter, laser-reflectors, Doppler transmitters for precise orbit determination, and a "satellite-to-satellite" tracking package [10.2]
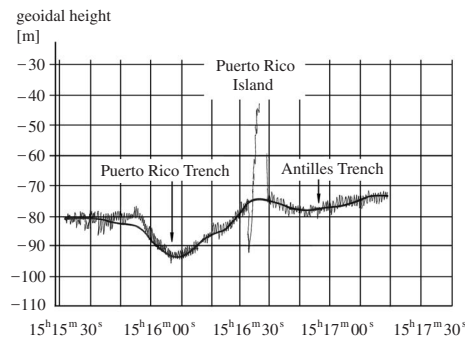
Figure 9.2. SKYLAB altimeter test

were installed on the spacecraft. Primary mission goals were improvement of our knowledge of Earth's gravity field, the geoid, ocean tides, currents, structure of Earth's crust, dynamics of the solid Earth, and remote sensing technology (Stanley, 1979). The design lifetime of the altimeter was only 1 year; however, the total amount of altimeter data spans 3.5 years.

SEASAT-1 (also designated as SEASAT–A) was launched on June 26, 1978. Its main objective was the mapping of ocean surface data through remote sensing techniques. The sensor module includes five sensors with the following tasks:

altimeter	satellite altitude, wave height, wind velocity,
scatterometer	wind velocity, wind direction,
microwave radiometer	sea surface temperature, wind velocity, atmospheric water vapor,
synthetic aperture radar	wavelength, wave direction, and
visible and infrared radiometer	feature identification.

On October 10, 1978, after about three months of operation, a short-circuit occured onboard and prevented the operation of most sensors, including the altimeter. Nevertheless, the majority of the mission objectives were reached. In particular the amount and quality of the altimeter data fully met initial requirements [9.5.1].

The U.S. Navy's *GEOSAT* (Geodetic Satellite) spacecraft was launched on March 12, 1985 into an 800 km near circular orbit with an inclination of 108° (Cheney et al., 1986; McConathy, Kilgus, 1987). The main instrument is an improved version of the radar altimeter flown on SEASAT. The precision for height measurements is about 3.5 cm (Mac Arthur et al., 1987). Additional subsystems are a dual frequency (150 and 400 MHz) Doppler beacon for spacecraft tracking based on the TRANET network [6.3], and a C-band transponder [4.4.2]. The attitude control subsystem (a gravity gradient system in the form of a 6 m scissors boom with a 45 kg end mass) was designed to point the radar altimeter to within 1 degree of nadir.

The satellite performed, in succession, two separate missions: the primary *Geodetic Mission* (GM) with data collected from March 31, 1985 through September 30, 1986, and the subsequent *Exact Repeat Mission* (ERM) from November 8, 1986 through January 5, 1990 (Marks et al., 1991).

The main objective of the primary geodetic mission was to map the marine geoid, up to latitudes of 72 degrees, at high resolution. The ground track had a near-repeat period of about 23 days. The drifting orbit produced a dense network of sea level profiles with an average track spacing of about 4 km, thus providing gravity fields of unprecedent accuracy and resolution. The GM data were initially classified but released in their entirety for public use in 1995. The full GM data set, in particular with recomputed orbits, gave rise to valuable geophysical interpretation [9.5.2].

In October 1986, upon termination of the GM phase, the spacecraft was maneuvered into an orbit whose ground track repeated every 17 days and largely corresponded to the SEASAT ground tracks. The data of this *Exact Repeat Mission* (ERM) were unclassified from the beginning, and freely accessible for geodetic and oceanographic work. The spacing between ERM ground tracks is ∼75 km at 60° latitude and hence

much wider than that of GM tracks. Fig. 9.3 gives an example for the latitude 60°S ±3°. "Exact repeat" means that the ground tracks repeat to within ±1 km for each 17-days repeat cycle. The ERM mission covered 62 complete 17-days cycles before a
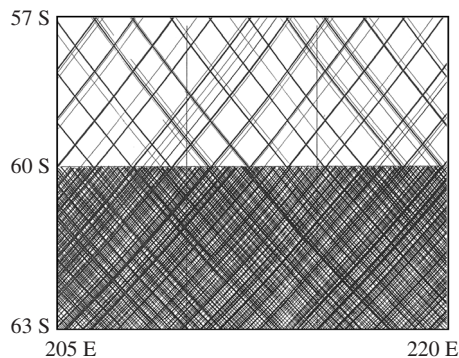
Figure 9.3. Ground track spacing of GEOSAT GM (medium lines) and ERM (thick lines) missions near 60° southern latitude; SEASAT ground tracks (thin lines) for comparison (from *Marks1991*)

tape recorder failure in October 1989 terminated the global data set. A limited amount of data was available by direct broadcasting in the North Atlantic and Gulf of Mexico until January 1990.

The more than three years of ERM data are highly valuable for oceanographic studies such as sea level variations [9.5.3]; however, the gravity fields derived from these data are not able to resolve fine-scale features fully, due to the wide spacing of the ground tracks [9.5.1].

Reprocessed so-called *geophysical data records* (GDRs), were made available through NOAA in 1997. The data set includes the entirety of the GM + ERM data with consistent JGM-3 orbits and enhanced geophysical corrections (NOAA, 1997).

The first *European Remote Sensing Satellite* (ERS-1) was launched on July 17, 1991 into a sun-synchronous, low-Earth, nearly circular orbit of 780 km and with 98.5° inclination. The satellite allowed all-weather high-resolution imaging over land, coastal zones and polar ice caps, measured ocean wave heights and wavelengths, wind speeds and directions, various ice parameters, sea-surface temperatures, cloud cover, atmospheric water vapor content, and precise altimetry over oceans and ice. The satellite carries a set of active microwave sensors supported by additional instruments (Fig. 9.4):

- Synthetic Aperture Radar (SAR),
- Radar Altimeter (RA),
- Along Track Scanning Radiometer (ATSR),
- Precise Range and Range-Rate Equipment (PRARE), and
- Laser Retro Reflector Array (RRA).

The total mass is 2 400 kg, and the over-all height is 11.8 m. The RA antenna has a diameter of 1.2 m. ERS-1 carries an attitude and orbit control system (AOCS) to main-tain the platform orientation in flight. This system contains, inter alia, infrared Earth-sensors, sun-sensors, an inertial core of six gyros, and three orthogonal reaction wheels (O'Brien, Prata, 1990).

The primary on-board system for precise orbit determination was PRARE [4.3.3.3]. Unfortunately PRARE failed during the ini-tial tests, hence precise orbits depend on satellite laser ranging.

The radar altimeter operates in the Ku-band in two modes, the ocean mode and the ice mode. Because of the near polar orbit, ERS-1 provides valuable data over ice and permits the study of the polar regions. The

Figure 9.4. ERS-1

design accuracy of the ERS-1 altimeter was 10 cm. The results, however, revealed a performance similar to the GEOSAT altimeter (∼3-5 cm).

Because of the multidisciplinary nature of satellite altimetry, different phases of operation were planned and realized. Two 3-day-repetition cycles were initiated for 3 months each during the Arctic winter to monitor ice coverage (*ice phase*). During 1992 and 1993, ERS-1 operated in the *multidisciplinary phase* with a ground track repetition rate of 35 days. In 1994, for primarily geodetic applications, ERS-1 was maneuvered into a 176-day cycle, the *altimeter phase* or *geodetic phase*. The equatorial ground track spacing was about 900 km for the ice phase, 80 km for the multidisciplinary phase, and 18 km for the altimeter phase.

The expected lifetime was about three years. ERS-1, however worked three times as long. This is why a completely new space technique, *Interferometric SAR* (InSAR) [11.2] could be applied in a *tandem phase*, with ERS-2, soon after its launch in April 1995. ERS-1 operation was terminated on March 10, 2000, because of a failure in the attitude control system. Table 9.1 gives an overview of the different mission phases (Schöne, 1997).

TOPEX/POSEIDON (Fig. 9.5), often abbreviated as T/P, is a satellite mission that carries a radar altimeter system as the primary instruments, and is jointly con-ducted by the NASA and the French Space Agency (CNES). TOPEX stands for Ocean *TOP*ography *EX*periment.

The spacecraft was successfully launched on August 10, 1992, into a circular, non-sun-synchronous orbit at an altitude of 1340 km and with an inclination of 66°. The mission includes two altimeters. The primary instrument is a dual-frequency altimeter, operating simultaneously at 13.6 GHz (Ku band) and 5.3 GHz (C band).
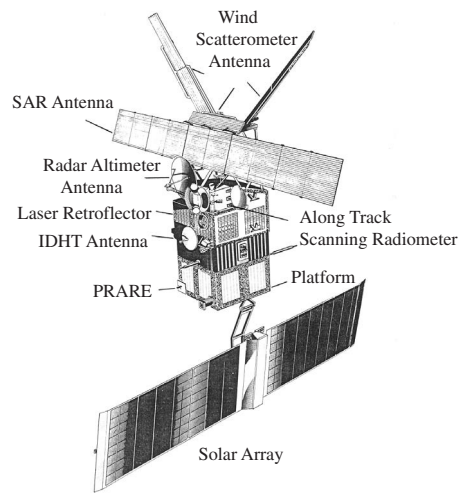
Table 9.1. ERS-1 mission phases

Operation phases	Orbit	Duration
start and test		17.07.1991 – 30.07.1991
commissioning phase	3-days	01.08.1991 – 12.12.1991
first ice phase	3-days	28.12.1991 – 10.03.1992
multi-disciplinary phase	35-days	14.04.1992 – 15.12.1993
second ice phase	3-days	24.12.1993 – 01.04.1994
geodetic phase	168-days	10.04.1994 – 19.03.1995
tandem mission phase	35-days	21.03.1995 – 10.03.2000

The measurements can hence be corrected for errors caused by free electrons in the ionosphere. As a by-product, the total electron content can be obtained. This altimeter is the first to use two-channel measurements for ionospheric range corrections.

The second instrument (POSEI-DON) is a solid-state radar altimeter experimental sensor (CNES), operating at a single frequency of 13.65 GHz. The main objective of the experimental sensor is validation of low-power, low-weight altimeter technology for future Earth-observing missions.

The orbit determination is supported by a laser retro-reflector array and the French DORIS (*Doppler Orbitography and Radiopositioning Integrated by Satellite*) tracking system [6.7]. In addition, an experimental GPS receiver is flown onboard. A radiometer provides estimates of the total water vapor content along the signal path, and is used to correct the altimeter measurements. The satellite is operated through NASA's *Tracking and Data Relay Satellite System* (TDRSS).

Figure 9.5. TOPEX/POSEIDON; courtesy NASA/JPL/Caltech

The TOPEX/POSEIDON mission is mainly designed to explore ocean circulation and its interaction with the atmosphere [9.5.3]. The ground tracks repeat exactly after about 10 days, with a spacing of about 315 km at the equator. The footprint is 3 to 5 km in diameter for typical wave heights. Measurements are taken approximately once per second, giving a spacing of about 6 km. The design lifetime of the mission was 3 to 5 years. At the time of this writing, however, this highly successful satellite is still delivering data after more than 10 years, having also flown several months in a tandem formation with its successor satellite JASON-1.

ERS-2, the follow-on mission to ERS-1, was launched by the European Space Agency, ESA, on April 21, 1995 into a nearly identical orbit to ERS-1, with the

objective to continue its work. Additional instruments measure the ozone content of the atmosphere (*Global Ozone Monitoring Experiment*, GOME). Both SLR and PRARE are used for precision orbit determination. Because of the longevity of ERS-1 both satellites were placed in a tandem configuration, separated by 400 km, for nearly five years. The orbital period is 35 days and the ground track spacing at the equator is 80 km.

GEOSAT follow-on (GFO) is a US Navy satellite mission to study physical oceanography. GFO was launched on February 10, 1998 into a near-circular orbit at 800 km altitude with 108 degrees inclination. The satellite carries a single frequency radar altimeter, water vapor radiometer, Doppler beacon, four GPS receivers, and a laser retroreflector array. The mission lifetime is scheduled to be about 10 years, following the GEOSAT *Exact Repeat Mission* (ERM) orbit with a repetition rate of 17 days after 244 revolutions. Due to problems with the onboard GPS receivers, precise orbit determination is mainly supported by SLR.

JASON-1 is a follow-on mission to T/P, again as a joint project of NASA and CNES. JASON-1 was launched on December 7, 2001, into a circular orbit very similar to that of T/P, namely with an altitude of 1336 km and an inclination of 66 degrees. The onboard instrumentation is:
  − tracking and data relay satellite transmitter,
  − microwave radiometer,
  − dual frequency altimeter (POSEIDON-2),
  − DORIS dual frequency receiver,
  − GPS receiver (TRSR), and
  − laser retro-reflector array.

As with TOPEX/POSEIDON, three independent techniques are used for precise orbit determination. The satellite weighs just 500 kg (one fifth the weight of its predecessor). The design lifetime is 5 years. The main objective is to continue the mission, started by T/P, to monitor world ocean circulation and to study interactions of the oceans and atmosphere. T/P time series are not long enough to resolve all scientific issues in oceanography. Oceanic oscillations with periods over 10 years can only be recovered by continued time-series [9.5.3].

The *Environmental Satellite* (ENVISAT-1) is the successor to the European Space Agency (ESA) Remote Sensing Satellites ERS-1 and ERS-2. The satellite was launched on March 1, 2002 into a near-circular, sun-synchronous, near-polar orbit of 800 km altitude and 98.5 degrees inclination. The orbital period is 35 days, as for ERS-2 and some phases of ERS-1. ENVISAT-1 is a multi-purpose satellite for environmental studies, and observes Earth's atmosphere, ocean, land and ice over a 5-year period, using ten complementary instruments. The most important sensors for research related to geodesy are:
  − Advanced Synthetic Aperture Radar,
  − Radar Altimeter (RA-2),
  − Microwave Radiometer,
  − DORIS Receiver, and
  − Retro-reflector Array.

Table 9.2 gives an overview of the system data of the cited altimeter missions. It is evident that a wealth of data for research in geodesy, oceanography and geophysics has been provided by so many missions. Since TOPEX/POSEIDON, altimeter noise levels are as low as 2 to 3 cm, corresponding to the noise level of the other space techniques. T/P and ERS-2 are already far beyond their expected lifetime. But three current missions, and planned successors, promise that satellite altimetry will remain a powerful technique in satellite geodesy.

Table 9.2. Characteristic data of altimeter missions

Mission	SKYLAB	GEOS-3	SEASAT-1	GEOSAT	ERS-1
Mission begin	1973	1975	1978	1985	1991
Mission end	1973	1978	1978	1989	1996
Duration (months)	days	44	4	54	57
Mean altitude (km)	435	840	800	785	785
Inclination		115	108	108	98.5
Max. latitude		65	72	72	81.5
Cycle repeat (days)				~23/17	3/35/168
Track division (km)				4/75	933/80/16
Frequency (GHz)				13.5	13.5
Altimeter noise (cm)	>100	60	10	4	4
Radiometer/Frequ.			yes	no	yes/2
Orbit determination		SLR, Doppler	Doppler	Doppler	SLR (PRARE)

Mission	TOPEX/ POSEIDON	ERS-2	GFO	JASON-1	ENVISAT-1
Mission begin	1992	1995	1998	2001	2002
Mission end					
Duration (months)					
Mean altitude (km)	1340	780	800	1340	800
Inclination	66	98.5	108	66	98.5
Max. latitude	66	81.5	72	66	81.5
Cycle repeat (days)	10	35	17	10	35
Track division (km)	316	80	165	316	80
Frequency (GHz)	5.3/13.6	13.5	13.5	5.3+13.6	3.2+13.6
Altimeter noise (cm)	2	3	3.5	1.5	2
Radiometer/Frequ.	yes/2	yes/3	yes/2	yes/3	yes/2
Orbit determination	SLR, GPS DORIS	SLR, PRARE	SLR (GPS) Doppler	SLR, GPS DORIS	SLR DORIS

For the near future two more missions are planned (or are already realized). *ICE-SAT*, carrying a laser altimeter (GLAS) [8.7] was launched on January 12, 2003. *CRYOSAT*, an altimetry satellite built by ESA, will be dedicated to polar observations.

It is scheduled for launch in 2004 or 2005 with the objective to determine variations in the thickness of continental and marine ice sheets. The satellite will carry two receiving antennas forming an interferometer in the across-track direction. Further altimeter missions are under discussion.

9.3 Measurements, Corrections, Accuracy

9.3.1 Geometry of Altimeter Observations

The simple observation equation (9.2), consistent with Fig. 9.1, has to be refined for a detailed discussion of the situation. Fig. 9.6 explains the relationship:

$$h = N + H + \Delta H + a + d. \tag{9.3}$$

Here we have:

h ellipsoidal height of the altimeter satellite, based on orbit computation,
N geoidal height,
H sea surface topography
ΔH instantaneous tidal effects,
a altimeter measurements, and
d discrepancy between the computed orbit and the actual orbit.

$H + \Delta H$ equals $\overline{H}$ in (9.2). The altimeter observable, a, has to be corrected for atmospheric influences, and it has to be referred to the spacecraft's center of mass. It is also possible to correct for ocean wave height and other influences [9.3.2], [9.3.3]. Accordingly, further refinements of equation (9.3) are possible, see e.g. Tapley, Kim (2001).

The difference between the geoid and the mean sea surface is called *sea surface topography* (SST); it can reach 1 to 2 m. Usually the *mean sea level* (MSL) is understood to be the sea surface freed from all time-dependent variations. The deviation of this level from the geoid is caused by different salinities of ocean waters, large-scale differences in atmospheric pressure, and strong currents (e.g. the *Gulf stream*). At a resolution of better than 2 m, it is thus not valid to assume that the geoid is approximated by MSL. This leads to difficulties, if height systems which are tied to different tide gauges are connected [9.5.3]. Measurements obtained at the sea surface, for instance gravity
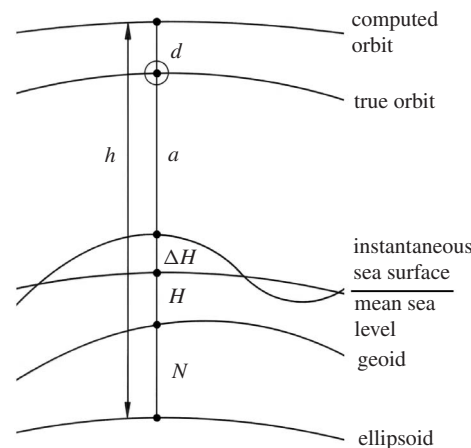
Figure 9.6. Geometrical relationship in satellite altimetry

anomalies from sea gravity observations, do not refer to the geoid but to the mean sea surface.

The sea surface topography can further be separated into one part, caused by atmospheric pressure, also called *inverted barometric effect*, and into another part, the mean *ocean dynamic topography*, mainly caused by ocean circulation (Tapley, Kim, 2001). Note that the denomination is not uniform in literature. The term "sea surface height" or "sea surface topography" may either refer to the ellipsoid or the geoid. In this book SST is defined with respect to the geoid.

9.3.2 Data Generation

The satellite altimeter antenna transmits a short rectangular impulse which is reflected back from the sea surface at the moment of contact. The simultaneously reflecting, i.e. the completely "illuminated", circular area is called the *footprint*. The footprint size depends on the sensor altitude, a, above the sea, the signal propagation velocity, c, and the pulsewidth, τ. The maximum radius of the circular area (cf. Fig. 9.7 (b)), (Chelton et al., 2001) is

$$R = \sqrt{2c\tau h}. \tag{9.4}$$
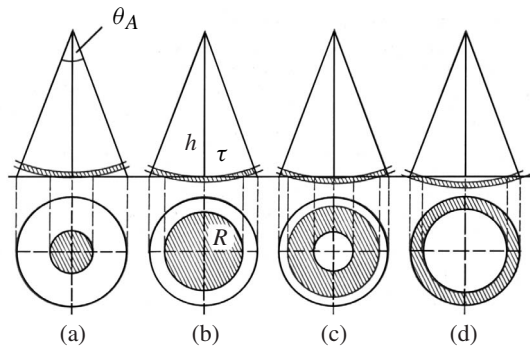
Figure 9.7. Variation of the illuminated reflecting area for a radar pulse penetrating the sea surface

Equation (9.4) is valid for a quiet sea surface. With increasing wave height the footprint radius expands because the effective pulselength is amplified as (Rummel, Sanso, 1993)

$$\tau' = \tau + \frac{2\,SWH}{c}.$$

SWH is the *significant waveheight* (see [9.3.3]). For ERS-1 and ERS-2, with $h = 800$ km and $\tau = 3$ ns, the footprint radius, R, varies between 1.2 km for quiet sea and about 6 km for rough sea (*SWH* = 10 m).

The reflected energy depends directly on the size of the reflecting area. Consequently, the returning energy continuously increases until the radar pulse is completely submerged in the reflecting surface (Fig. 9.7 (a), (b)). As soon as the outer edge of the pulse has arrived at the surface, the reflecting area converts into an annular form, with a nearly constant area, until the edge of the beam is reached (Fig. 9.7 (c), (d)).

Altimeter systems, where the size of the footprint depends on the length of the transmitted pulse in the above described way, are called *pulse-length-limited* systems. Another possibility is to define the footprint size by the beamwidth, θ_A, of the radar beam. Such *beamwidth-limited* systems require a very precise nadir adjustment of the antenna. A detailed discussion of footprint sizes, and their dependence on wave heights, is given by Chelton et al. (2001).

Fig. 9.8 shows the idealized form of the return impulse in pulse-length-limited systems. The leading edge has the length of the transmitted impulse (12.5 ns for GEOS-3), and is followed by a constant level until the final edge.
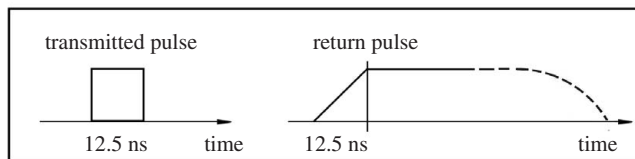
Figure 9.8. Transmitted pulse and return pulse

The range observable is derived from the time delay between the instant of pulse transmission and the moment when the return pulse reaches half of the maximum amplitude. The size of the reflecting area and the real form of the return pulse depend on the quality of the reflecting surface (water, ice, land), and on the roughness of the surface (wave heights). Information on the sea state can be derived from a pulse analysis. The wave crests and troughs within the footprint are averaged out. Remaining systematic effects, depending on the sea state, have to be determined by calibration measurements in special test areas (*ground truth*).

The altimeter pulse frequency is usually about 1 KHz. The single pulses are condensed to 10 Hz and then reduced to 1 second mean values. This process leads to a strong limitation of the observation noise. The precision of a 1-second-mean altimeter observation is about 2 cm for current missions (see Table 9.2). To convert the altimeter measurements into sea surface heights requires several corrections, such as for instrumental influences, atmospheric delays, sea state and orbit improvement (see [9.3.3]). The final accuracy for current missions is about 6 to 8 cm [9.4].

Altimeter measurements or sea surface heights are given along the satellite ground track with a spacing of about 7 km. The spacing between tracks can be much larger, depending on the mission objectives. Geodetic missions, with the aim to determine a high resolution gravity field, aim for a dense spacing and low repetition rate. Examples

are 4 km for the GEOSAT GM mission, and about 16 km for the ERS-1 geodetic phase (see Table 9.2). Oceanographic missions, aiming at measurements of sea level variations, need frequent repetition rates and thus accept a larger spacing. A good example is TOPEX/POSEIDON with a track spacing of 316 km. The European satellites select a multidisciplinary compromise with 80 km track spacing. Fig. 9.14, p. 462, shows the track pattern of SEASAT for a period of 18 days.

Altimeter data and data products are available to users through particular data centers under the control of the responsible space agencies. The data and products are provided in a mission-dependent format and also include corrections, orbit information and waveform analysis. The ENVISAT products, for example, are categorized into three distinct levels (Benveniste et al., 2001):

 − *Level* 0 (raw): unprocessed data as it comes from the instrument,
 − *Level* 1 (engineering): data converted to engineering units with instrumental calibration applied, and
 − *Level* 2 (geophysical): data converted to geophysical units, with datation, geolocation, including "re-tracked" data, such as range, wind speed, wave height.

Re-tracking means improvement of the altimeter data based on analysis of the return signal (Heidland, 1994). Different techniques are applied referring to the specific properties of the reflecting surfaces, such as water, continental ice, shelf ice or sea ice.

Data products of modern altimeter missions are available either in near-real-time (3 hours), in quasi near-real-time (2–3 days), or with the highest precision offline after several weeks. The generation of near-real-time products is supported by onboard orbit determination, for example with the DORIS-DIODE capability [6.7]. Significant advantages are to be gained by merging the data from two or more altimetry missions with different orbital patterns (see Fig. 9.13), such as ERS and T/P, or ENVISAT and JASON.

In some cases, historic altimeter measurements are reprocessed with improved orbital data based on refined gravity models. This is in particular true for the GEOSAT GM data set (NOAA, 1997). Data formats and standards are not yet homogeneous. The establishment of an "International Altimeter Service", similar to the existing services for GPS, SLR, and VLBI is under discussion. The main objective is to harmonize the use of heterogeneous data for *Multi-Mission Satellite Altimetry* (MMSA).

9.3.3 Corrections and Error Budget

Three groups of corrections have to be considered; each of them contributes to the level of achievable accuracy:

 (1) deviation of the real orbit from the calculated orbit (orbit error),
 (2) influences along the signal propagation path (altimeter error), and
 (3) deviation of the instantaneous sea surface from the geoid.

(1) *Orbit errors*
These are mainly caused by [3.2]

 − limited resolution and accuracy of the terrestrial gravity field used in the orbit
  computation,
 − errors in the coordinates of the tracking stations,
 − errors or limitations in the tracking system (Doppler, Laser), and
 − mismodeling in the orbit computation.

The predominant influence, in particular for the early altimeter missions, comes from
the *terrestrial gravity field*. Each satellite is particularly sensitive to a certain sub-set
of harmonic coefficients; it is therefore advisable to develop *tailored gravity models*,
including observations of the particular satellite, or satellites with a similar orbit. This
has been done, for example, with the gravity model GEM 10 for GEOS-3 [12.2.2],
that improved the orbit accuracy from 10 m to 1 to 2 m (Lerch et al., 1979).

The tailored gravity models PGS-S1 to PGS-S4 were developed for SEASAT, and
have improved the orbit accuracy from 5 m initially to 1 m (Fig. 9.9, Wakker et al.
(1987)). These models could also be used for ERS-1 because of its similar orbital
elements. Precise GEOSAT orbits, at
the level of 50 cm, have been com-
puted based on the GEM-T2 geopo-
tential model (Marsh et al., 1990).
GEM-T3 was a refined version (Lerch,
1992), with the inclusion of altime-
ter data from GEOS-3, SEASAT and
GEOSAT. Based on GEM-T3, the se-
ries of *Joint Gravity Models* (JGM) was
developed to compute precise orbits for
TOPEX/POSEIDON. Since none of the
previously used satellites had a sim-
ilar orbit, a refinement of the pre-
launch model, JGM-1, was started with
T/P, laser and DORIS tracking data
and direct altimetry data from GEOS-
3, SEASAT and GEOSAT, resulting in
JGM-2. The inclusion of additional SLR
data, DORIS and GPS tracking of T/P,
resulted in JGM-3 (Tapley et al., 1996).

Figure 9.9. Improvement of the radial or-
bit accuracy for SEASAT ((1) before launch;
(2), ..., (5) additional data from SEASAT
laser, GEOS-3 altimetry, SEASAT altimetry,
TRANET Doppler; (6) final accuracy)

This model provided an orbit accuracy of 10 cm for reprocessing of the GEOSAT data,
and it is also appropriate for JASON.

Based on JGM-3, a tuned model has been developed to provide improved orbits for
the ERS satellites (Scharroo, Visser, 1998), and hence also for ENVISAT. The most
recent model is EGM96 (Lemoine et al., 1998) including additional data, with a much
higher spatial resolution than models such as JGM-3.

With the various new models at hand, the orbits of altimeter satellites can be
modeled with an accuracy of better than 10 cm in the radial component (Chelton
et al., 2001). The gravity field is hence no longer the dominant factor for precise orbit
determination in satellite altimetry.

Evidently, in satellite altimetry, the radial orbit error is of particular interest. The geocentric distance of the satellite can be approximated for short orbital arcs, with (3.49), as

$$r = a(1 - e \cos E).$$

Linearization of this directly gives the radial orbit error:

$$\Delta r = \Delta a + e \, \Delta M \, a \sin M - \Delta e \, a \cos M, \tag{9.5}$$

where the eccentric anomaly, E, has been approximated by the mean anomaly, M. Obviously, the radial orbit error depends on the quantities Δa, ΔM, and Δe which can be regarded as constant over short portions of the orbit.

An analysis of the Lagrange equations (3.119) demonstrates that the errors in the gravity field coefficients produce periodic variations and resonances in the three elements [3.2.2.2]. The effect must hence be carefully modeled (Engelis, 1987).

A second important aspect in orbit modeling is the available *tracking system*. For operational purposes the all-weather radio-tracking systems GPS or DORIS are preferable. As a back-up system, and for higher accuracy requirements, for instance over test and calibration fields, or for particular projects, laser ranging campaigns are organized for altimeter satellites equipped with corner cube reflectors (ILRS, 2000). The disadvantage of SLR tracking is its weather dependency and the non-uniform global distribution of tracking stations. Good results with Doppler tracking based on TRANET [6.2.2] have been achieved for GEOSAT (e.g. Shum et al. (1990)).

The TOPEX-POSEIDON mission is mainly based on tracking data with the DORIS Doppler technique [6.7]. After failure of its PRARE subsystem [4.3.3.3], the orbit determination of ERS-1 was mainly supported by laser tracking. PRARE and SLR are used for ERS-2, and DORIS and SLR for ENVISAT, (see Table 9.2).

The geocentric coordinates of the tracking stations are usually related to the ITRF and are based on precise global observation techniques (SLR, VLBI, GPS, DORIS). The accuracy is continuously improved by new observations, and is now at the one centimeter level [12.1.2]. For the early altimeter missions, like GEOS-3, SEASAT-1, and GEOSAT, however, the coordinates of the tracking stations could be subject to errors of a few meters (Khan, 1983). A new adjustment of historic altimeter data, with improved orbit computation, may be of interest. A very successful example is the reprocessing of the GEOSAT GM altimeter data set based on the JGM-3 gravity model and complete Doppler tracking data (NOAA, 1997).

Even under favorable conditions and with modern tracking systems, the remaining orbit errors form a considerable part of the total error budget. Hence, methods for orbit improvement have to be applied. Non-dynamical techniques can also be used, because only the radial component is of interest (Sandwell et al., 1986). Most of these are based on the *crossover method* [9.4].

(2) *Influences along the signal path*
These can be subdivided into:
    − instrumental errors, and

– propagation errors.

The most important *instrumental* influences are the
  – distance between the phase center of the radar antenna and the spacecraft center
    of mass (center of gravity correction),
  – propagation delay in the altimeter electronics, and
  – timing error in the measuring system.

These effects can be minimized and estimated as the altimeter instrument is built. The overall effect of the instrumental errors, the *altimeter bias*, is determined and controlled in the *altimeter calibration* over precisely surveyed test areas (*ground truth*). Satellite passes through the zenith of laser ground stations are in particular suitable. For GEOS-3, a mean calibration value of 5.30 m was determined (Berbert, Carney, 1979). For SEASAT-1 the calibration value did not differ significantly from zero (Kolenkiewicz, Martin, 1982).

For TOPEX/POSEIDON calibration sites have been maintained at a platform off the coast of Southern California and near Lampedusa Island in the Mediterranean Sea (Chelton et al., 2001). Besides SLR, GPS and DORIS are also used (Fig. 9.10). Sea surface height is determined independently by the altimeter and direct in situ measurements with tide gauges. Magnitudes of the biases for recent altimeter missions range from 40 cm for ERS-1 to near zero for the two altimeters onboard T/P. In addition, validation of JASON and ENVISAT is assisted by a worldwide network of tide gauges collocated by DORIS and GPS measurements.

For the study of changes in the global mean sea level [9.5.3] it is important to detect and calibrate potential drifts in the measurement system. Considering a eustatic sea level rise of less than 2 mm/year, it is necessary to detect any drifts at the 1 mm/year level. This requires continuous calibrations at the calibration sites over several years.

Figure 9.10. Geometry of in situ altimeter calibration, after Chelton et al. (2001)

Another instrumental error may be caused by deviation of the beam direction from the vertical (*nadir error*). The influence depends on the pulse length and the beam width, and can be minimized by technical means. The effect of the nadir error can be neglected for the pulse-length limited systems (cf. Fig. 9.7), because the footprint is then defined by the pulse-length.

Signal propagation errors are caused by the influence of the ionosphere and the troposphere on the propagation velocity [2.3.3]. The influence of the *ionospheric refraction* in the frequency domain of 14 GHz is about 5 cm to 20 cm, depending on the level of ionization. As with other microwave systems used in satellite geodesy [6.4.2], [7.4.4], the ionospheric effect can be modeled with measurements at two frequencies.

This is why dual-frequency altimeters are flown on modern altimeter missions, for example 13.6 and 5.3 GHz on T/P and JASON (see Table 9.2).

The influence of the *tropospheric refraction* is about 2.3 m, because only vertical measurements are applied. The total tropospheric influence can be corrected with an accuracy of a few centimeters based on appropriate refraction models [2.3.3.2], in particular if information on the water vapor content along the signal path is available from simultaneous radiometer measurements from the same satellite. This is the case for all altimeter missions since SEASAT, with the exception of GEOSAT (Table 9.2). For GEOSAT, corrections with an accuracy of approximately 2 cm were derived from global fields of water vapor based on observations from meteorological satellites (Douglas, Cheney, 1990). Precise tropospheric parameters are also available as a product of the IGS global tracking network and from GPS/LEO occultation observations, see [7.6.2.9], [7.8.1].

The effect of the *sea state* on the reflected signal can also be included in the list of propagation errors. The size of the reflecting area depends on the roughness of the sea surface: a higher asperity creates a larger area. The slope of the leading edge of the return signal decays with increasing wave height, because the energy is reflected earlier from the tops of the waves (Fig. 9.11). The actual wave heights can be derived from signal analysis. Systematic corrections are required because reflections at the wave crests are less intensive than in the troughs. These corrections, also named *sea state bias*, have a linear dependency from the largest third of wave heights (*SWH*) and reach up to 2 to 4% of *SWH*, e.g. Douglas, Agreen (1983). A good discussion on the sea state bias can be found in Chelton et al. (2001).
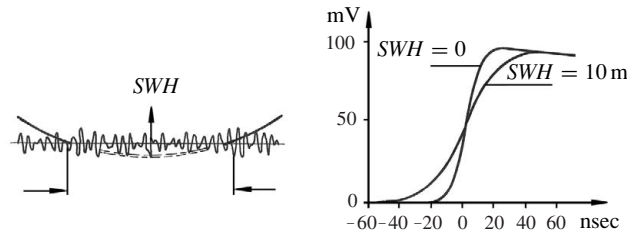
Figure 9.11. Relation between the leading edge slope of the return signal and the wave height; *SWH* = mean value of the largest third of wave heights

(3) *Deviation of the instantaneous sea surface from the geoid*

This effect can be split into one part, H, constant in time, and another part, ΔH, variable in time (Fig. 9.6). The altimeter observations have to be corrected for the variable time-dependent component before they can be used in the determination of the quasi-stationary mean sea surface. The variable deformations of the sea surface are mainly caused by tides and atmospheric loading effects; the wave-induced roughness of the sea surface can be neglected because it is already averaged out within the altimeter observation process.

The response of the sea surface to atmospheric loading is about -1 cm mbar^{-1}. This so-called *inverted barometer* effect causes global deformations of the sea surface between 10 and 20 cm. Regional effects, in particular in the tropics, are smaller. Sea surface pressure values are taken from numerical weather prediction models, with an uncertainty of 2 to 4 mbars, corresponding to 2 to 4 cm. The inverted barometer effect is hence still a major error source in the precise analysis of sea-surface height. For detailed information on the subject see e.g Wunsch, Stammer (1997).

The principal part of the variable deformation is induced by ocean tides. In the open ocean tidal amplitudes reach about 10 to 60 cm with larger values near the coasts and in shallow marginal seas. Before the launch of T/P knowledge of tides was mainly based on hydrodynamical models with empirical constraints from globally distributed tide gauges. The global ocean tide model of Schwiderski (1984), with an accuracy of about 0.1 m was widely used. With the inclusion of T/P data, the tides can now be estimated to an accuracy of 2–3 cm (LeProvost, 2001). One example is the CSR3.0 model (Eanes, Bettadpur, 1996). In shallow seas, such as the North Sea, deviations from the global model can reach quite large amounts; in such cases, local or regional tidal models must be applied.

Solid Earth tides cause deformations of the Earth body of several decimeters in height. They include the effect of direct astronomical forcing (the body tide) and the effect of crustal deformation by ocean tides (tidal upload). The modeling accuracy is about 1 cm or better (Zahran, 2000). Both, solid Earth and ocean tides have to be removed from altimeter data before they can be used for the study of ocean circulation (Chelton et al., 2001).

The constant part, H (Fig. 9.6), the *sea surface topography*, also named *dynamic sea surface height* shows amplitudes up to about 2 meters [9.3.1], [9.5.3]. Depending on the subject, the quantity H can be regarded as a correction or as the signal of interest. If altimeter observations are used in the determination of a marine geoid, the sea surface topography has to be taken as a correction to the measurements from oceanographic models. Traditionally, dynamic heights are computed from hydrographic data with respect to a reference level of equal pressure at great depths (several thousand meters). The accuracy of these models is debatable, and is at best about 20 cm (Fu et al., 2001).

In oceanography the dynamic sea surface height is of interest, and the geoid undulation is taken as a correction. Precise geoid information over the oceans, without the inclusion of altimeter data, is available for long wavelenghts only. For shorter wavelengths, with the inclusion of altimeter data the current accuracy estimate is at the few decimeter level (Chelton et al., 2001).

The difficulty in separating the geoid undulation, N, and the sea surface topography, H, hence constitutes a basic problem in the use of satellite altimetry in geodesy as well as in oceanography. Several solution concepts exist to improve the situation by the inclusion of additional information on ocean flow or satellite orbits (Rummel, Sanso, 1993). The best solution will be an independent geoid improvement for shorter wavelengths which can be expected from the new gravity field missions [10].

9.4   Determination of the Mean Sea Surface

The altimeter data are available as a sequence of altitude values of the mean sea surface above the selected reference ellipsoid along the sub-satellite track (cf. Fig. 9.12). The distance between adjacent sub-satellite points varies according to the satellite and the observation mode; for current missions it is about 7 km. The spacing between the tracks varies from 80 to 316 km for current missions, and it was 4 and 16 km for the geodetic missions of GEOSAT and ERS-1.

Figure 9.12. Example of altimeter data in the North Sea

After applying all the corrections discussed in [9.3], a systematic vertical error remains, which is partly due to the radial orbit error [9.3.3], and which contains all other residual errors.

The subsequent data processing makes use of the fact that the same ocean area is covered several times so that the altitude of the mean sea surface $N + H$ must be identical in the intersections of the ascending and the descending satellite ground tracks. This so-called *crossover technique* is widely used in the evaluation of altimeter measurements, e.g. (Rummel, Sanso, 1993). Fig. 9.12 demonstrates the principle for a small section of the North Sea and GEOS-3 data (Monka, 1984).

After correction of the altimeter measurements, the differences in the cross-points are more or less identical to the differences in the radial orbit errors of the contributing satellite orbits and residual altimeter biases.

As a first step, the crossover points have to be located. This can be achieved in an iterative process, based on the geographical coordinates of the satellite ground-tracks. Other methods start from the satellite ephemerides (Kim, 1997). Then, the heights at the crossover points are interpolated from the sequence of measured altimeter heights along the ground tracks. Simple linear or quadratic interpolation formulas can be used, or advanced interpolation techniques like the Kalman or Wiener filter. The differences between the interpolated values at the crossover points are

$$d_{ij} = \tilde{h}_i - \tilde{h}_j = d_i - d_j + \varepsilon_i - \varepsilon_j \tag{9.6}$$

with

$\tilde{h}_i, \tilde{h}_j$ interpolated altimeter observations along the ground tracks i and j for the crossover point,

d_i, d_j orbit errors of the satellite orbits i and j for the crossover point, and

$\varepsilon_i, \varepsilon_j$ observation errors.

The orbit error can be modeled along the satellite orbit by use of a polynomial in distance or time. In the simplest case, a shift or inclination of the orbital arc is

sufficient. In the general case, the observation equation

$$v_{i,j} = \sum_{e=0}^{k} \alpha_{ie}(t_i - t_{i0})^e - \sum_{e=0}^{l} \alpha_{je}(t_j - t_{j0})^e - d_{ij} + \varepsilon_i - \varepsilon_j, \qquad (9.7)$$

with

k, l	degrees of the polynomials for the modeling of orbital errors in the sub-satellite tracks i and j,
α	polynomial coefficients,
$(t - t_0)$	epoch of the observation in the crossover point relative to the start epoch, t_0, and
$v_{i,j}$	residuals,

can be formulated for each crossover point. This non-dynamical form of orbit modeling can be applied, because the orbital arcs are very short, and only the radial component is considered.

Starting from the crossover technique, other algorithms have been developed for determination of the mean sea surface. Examples are the use of two-dimensional polynomials, least squares interpolation, and the combination of dynamical orbit improvement methods (e.g. short arc), with different interpolation techniques. For an overview see Rummel, Sanso (1993) and Tapley, Kim (2001).

Fig. 9.13 shows the combined distribution of groundtracks for T/P and EN-VISAT. The track spacings are complementary to each other. The large number of cross points can be used to model the mean sea surface.
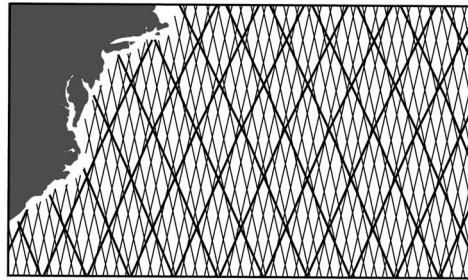
Figure 9.13. Ground tracks of T/P (solid lines) and ENVISAT (after: CNES, Aviso website)

9.5 Applications of Satellite Altimetry in Geodesy, Geophysics, and Oceanography

As has been stated before, the mean sea surface can be measured by altimetry with about 1 cm precision. It is composed of

- the marine geoid, with variations of about ± 100 m,
- the mean ocean dynamic topography, with variations of about ± 1 m, and
- the response of the ocean surface, averaged over time, to atmospheric pressure changes, with a variation of about ± 10 cm.

Although the problem of separating the geoid and the mean ocean topography is not yet completely resolved and remains one of the challenging scientific tasks in marine

geodesy, the wealth of data from satellite altimetry provides the best overall approach for determination of the marine gravity field [9.5.1].

The fine structure of the mean sea surface, derived from high resolution satellite altimeters such as GEOSAT and T/P reflects ocean bottom topography and tectonic structures of the oceanic lithosphere, and contributes significantly to marine geophysics [9.5.2]. The high frequency of track repetitions together with the centimeter resolution of modern missions like T/P, JASON, and ERS/ENVISAT provides a powerful means for continuous monitoring of ocean surface variability and related processes in oceanography [9.5.3]. Dedicated missions with high orbit inclination, like ERS-1/2, ICESAT and the forthcoming CRYOSAT, offer unprecedented opportunities to map the polar icesheets [9.5.3].

Whereas the first altimeter missions were mainly directed to the mapping of a detailed marine gravity field, also with the objective to improve the precise orbit determination of the altimeter satellites, the modern missions, with their high resolution in the time domain, are more committed to monitor time variable effects.

9.5.1   Geoid and Gravity Field Determination

To a first approximation, the mean sea surface and the marine geoid can be considered as equal. The mean sea surface heights along a satellite subtrack, resulting from an altimeter mission [9.4], can hence be treated as geoidal heights, N. Fig. 9.14 shows, as an example, the global coverage with SEASAT satellite tracks for a period of 18 days. The complete data set is accordingly much denser.
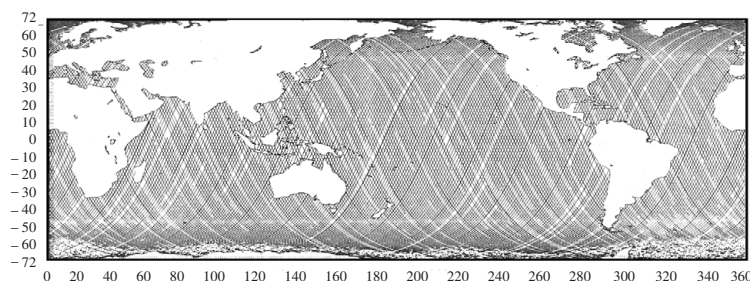
Figure 9.14. 18-day global coverage with SEASAT-1 data (Marsh, Martin, 1982)

In total, about 5 million height values are available for the GEOS-3 mission as well as for the SEASAT-1 mission; about one half are useful after data screening. Because of the inclination of the satellite orbits, the data are restricted to a belt between 72° north and south. In spite of the very short lifetime of SEASAT-1, both data sets are about equal in size. SEASAT-1 delivered data for 1648 hours, from 3 months of operation, and GEOS-3 provided data for 1745 hours, from 3.5 years of operation.

These sea surface heights can be used to determine an evenly-spaced grid of globally distributed heights, based on suitable interpolation techniques. Based on the grid

points, a representation with contour lines can be derived. Fig. 9.15 gives, as an example, the representation of the mean sea surface based on the data of Fig. 9.14. This surface roughly corresponds to the geoid and demonstrates that the early altimeter missions provided excellent global coverage and new insights into the structure of the marine geoid
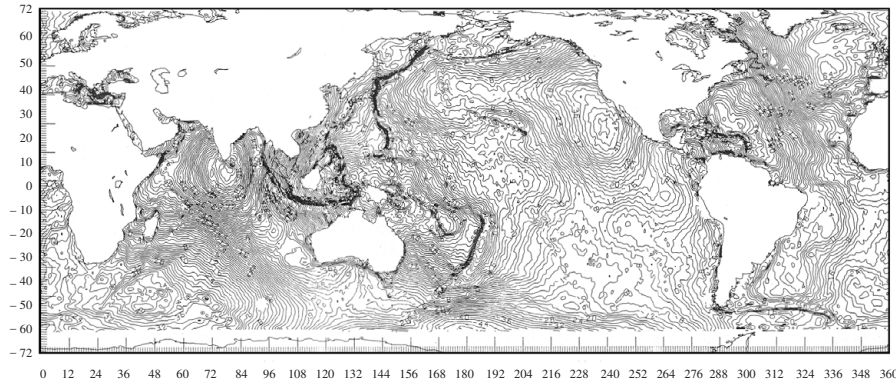
Figure 9.15. Sea surface topography from SEASAT data (Marsh, Martin, 1982)

The global altimeter data base was improved considerably after the release of the GEOSAT data from the geodetic mission (GM). In total, 30 million GEOSAT data and 20 million data from the ERS-1 geodetic phase are available (Knudsen, Andersen, 1997). The track spacing is between 4 and 8 km. A further breakthrough came with the T/P data and its high radial orbit accuracy of 3 to 4 cm. By about 1995, combined solutions of several altimeter missions led to sea surface models of about 1 dm precision. By averaging over several years, and reprocessing of older data with improved gravity field models, the current uncertainty of the mean sea surface representation is a the 1 cm level (Tapley, Kim, 2001).

Altimeter data are used as a data base to derive gravity anomalies with Stoke's inverse integral formulas (Torge, 2001). For example, the GEOSAT GM data first gave $5' \times 5'$ mean altimeter heights and were then converted into $5' \times 5'$ mean gravity anomalies (Trimmer, Manning, 1996). Usually, the sea surface topography is first removed with a global model.

Altimeter data are included in all recent Earth models [12.2.2]. While the classical analysis of satellite orbits only allows the determination of the long wavelength part of the potential field, the inclusion of altimeter data, together with surface gravity data, gives a much higher resolution, up to degree and order (360, 360), and even higher. The use of heterogeneous data from different sources has to be supported by proper weighting. Appropriate strategies have to be applied to account for the dynamic sea surface topography. EGM96, for example, solves for a separate spherical harmonic expansion of the dynamic SST (TOPEX, ERS-1 and GEOSAT) to degree

and order 20 (Lemoine et al., 1998). Comparisons between T/P mean sea surface over two years with an independent hydrographic SST model and EGM96 show an agreement of about 10 cm (Lemoine et al., 1998; Tapley, Kim, 2001). It seems that this number is the current accuracy limit for an ocean geoid model, derived from altimetry. For a better separation between geoid and dynamic sea surface topography it will be necessary to include data from the current and forthcoming gravity field missions CHAMP, GRACE and GOCE [10]. For more information on the subject see Tapley, Kim (2001) or the series of IAG symposia on topics related to geoid determination, e.g. Rapp et al. (1996); Segawa et al. (1997); Forsberg et al. (1998).

9.5.2   Geophysical Interpretation

Except for the 1 to 2 m dynamic sea surface topography associated with oceano-graphic features [9.5.3], the geoid and the sea surface coincide. The SST can be represented by rather long wavelengths. Altimeter data with a precision of a few cm can thus resolve geoid signals associated with seafloor topography. This is particularly true for the densely-gridded data from the ERS-1 geodetic mission (1994/1995) and the originally-classified GEOSAT GM data (released in 1995). These data provide a detailed view of the marine gravity field with a spatial resolution of better than 5 km.

Previously undetected *seamounts* (submarine elevations), uncharted *fracture zones* and *deep sea trenches* can be identified and located, and they can be made visible in suitable representations. Continental margins, rifts, and the orientation of fracture zones can be traced over thousands of kilometers, and give a wealth of information for geophysical interpretation. Mid-ocean spreading ridges, the longest mountain chains on Earth, are now nearly entirely mapped. Lineations have been detected in every ocean basin that are the result of tectonic plate motions. Satellite altimetry hence provides a confirmation of plate tectonics. For detailed information on the subject see Cazenave, Royer (2001), with many impressive pictures.

Geoid anomalies related to submarine tectonic features have wavelengths below about 3000 km. Hence, as a first step in analysis a long wavelength reference geoid, such as JGM-3, is subtracted from the data. Another procedure is to use a high pass filter. For detailed investigation it can be useful to remove wavelength components larger than 200 km by filtering. The remaining data allow detailed mapping of the seafloor tectonic fabric.

Fig. 9.16 demonstrates the relationship in the case of a seamount. Such seamounts generate an excess of mass relative to the adjacent oceanic plain. This excess produces deformations (little bumps) on the mean sea surface (in the geoid height) that can be measured by the satellite altimeters. The height variation corresponds to variations in shipborne gravity. The geoid anomaly reaches about 1 to 2 m for a 1 to 3 km high seamount with a typical base diameter of 10 to 50 km. The related gravity anomaly varies between 20 and 200 mGal. More than 10 000 seamounts have been identified alone on the Pacific plate (Cazenave, Royer, 2001).
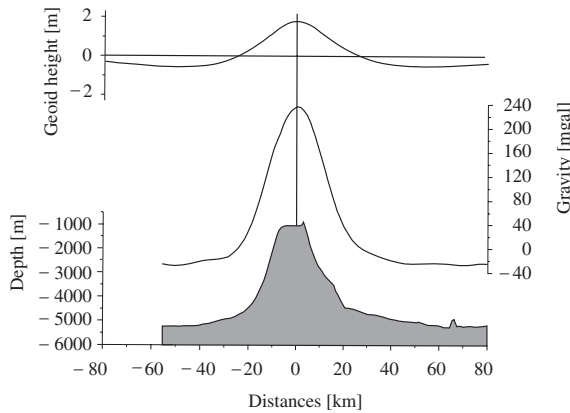
Figure 9.16. Gravity signal and geoid variation over a seamount, after Cazenave, Royer (2001)

Satellite altimeter data are also a valuable support to plan ship cruises for in situ investigation of the ocean floor.

9.5.3 Oceanography and Glaciology

In physical oceanography, the altimeter data are used to study the variations in space and time of the sea surface, with respect to an equipotential surface. Whilst sea surface topography is a disturbance (noise) in precise geoid determination, it contains a wealth of information on flow and tides for the oceanographer.

For determination of the dynamic sea surface topography it is necessary to derive a geoidal model with an accuracy of better than 0.1 m from other data, and to subtract it from the altimeter heights. This is a very exacting requirement which has, up to now, only been fulfilled for long-wavelength structures over thousands of kilometers. A significant improvement can be expected from current and forthcoming satellite missions for the determination of a high resolution gravity field, for example with *satellite-to-satellite* tracking or *satellite gradiometry* [10.1].

Fig. 9.17 shows a contoured global representation of the sea surface topography in which large-scale circulation features can be identified (Lemoine et al., 1998). The surface was computed from TOPEX, ERS-1 and GEOSAT data with a spherical harmonic expansion to degree and order 20 along with the gravity model EGM96.

The dynamic SST is directly related to large scale ocean circulation. The large scale flows are in so-called *geostrophic* balance, described by the equations

$$fv = \frac{1}{\rho}\frac{\partial p}{\partial x}, \quad \text{and} \quad fu = -\frac{1}{\rho}\frac{\partial p}{\partial y}. \tag{9.8}$$

Here, u and v are the velocities in the x (east) and y (north) directions. p is the pressure, ρ the sea-water density, and f is the Coriolis parameter, defined as $f = 2\Omega \sin \Phi$,
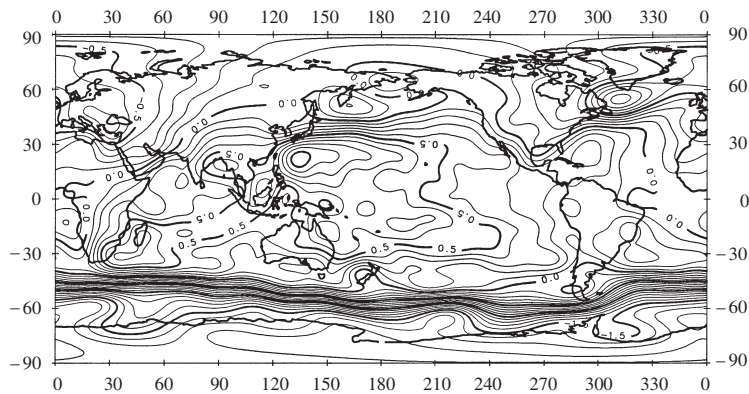
Figure 9.17. Global representation of sea surface topography; contour interval 0.1 m, after Lemoine et al. (1998)

where Ω is the Earth rotation rate and Φ the latitude (Robinson, 1995). For a precise modeling of ocean circulation two requirements must be fulfilled:

- precise altimeter measurements (1 to 2 cm, available with current altimeter missions),
- precise marine geoid (1 to 2 cm, expected for the near future with gravity field missions GRACE, GOCE).

The variable part of sea surface topography can already be precisely monitored with the repeat missions of the current satellites, namely 10 days for TOPEX/POSEIDON and JASON, and 35 days for ERS and ENVISAT. The 10-days T/P cycles have been available for more than 10 years. Analysis of the cycle solutions gives information on

- currents,
- tides, and
- sea level changes.

The major *ocean currents*, such as the Gulf stream, generate a sea surface slope of about one meter. These currents are meandering and this can be detected by altimetry. Fig. 9.18 shows an early example. Two passes of GEOS-3 data were taken along the same track in the North Atlantic, but separated by five months. The geoid undulation of 1.5 m at the position of a seamount remains unchanged. The shaded area is a 120 cm change in height due to a meander of the Gulf stream (Cheney et al., 1984).

The sea surface slope associated with currents can increase or decrease by about 0.1 m with seasonal changes in volume transport. With the 10 days cycles of T/P or JASON, such variability can be regularly identified and located.

The largest variability of sea level is due to *ocean tides* with a mean global amplitude of about 30 cm. Repeated altimeter measurements can be regarded as a huge number of tide gauge measurements distributed over the global ocean. The altimeter data are assimilated into mathematical tide prediction models. The current best ocean tide
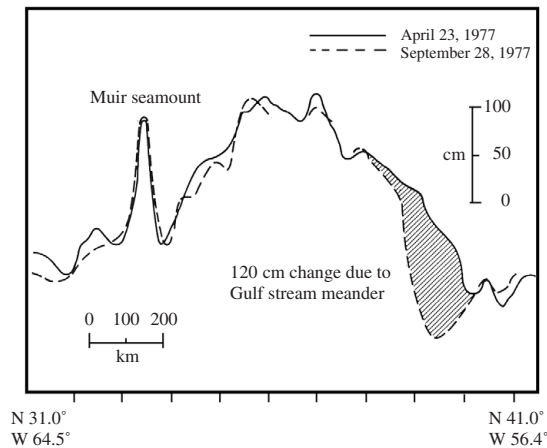
Figure 9.18. Variation of the sea surface topography due to a Gulf stream meander

models derived from T/P data have an accuracy of about 2 to 3 cm (Desai, Wahr, 1995; Fu, Chelton, 2001).

The study of *sea level changes* shows periodic as well as secular effects. The results have been verified by comparisons to globally distributed tide gauges (Cazenave et al., 1999). Each cycle (10 days for T/P and JASON) delivers a model of the mean sea state, which can be gridded. For the grid points the sequence of altimeter heights can be analyzed in a time series, e.g. a Fourier series. Annual periods show amplitudes up to 20 cm; the secular trend indicates a sea level rise of about 2 mm/year. The annual variations can be referred to as *ocean seasons*, caused by variable temperature.

The differences between spring and autumn reach about 20 cm. Well known is the *El Niño* effect, resulting in a sea level anomaly of about 10 cm caused by increased water temperature. Altimeter data can identify the eastward wander of this anomaly over several month-long periods across the Pacific Ocean. Satellite altimetry hence contributes to global climate studies. Oscillations with periods over 10 years will be recovered from the long time series of TOPEX/POSEIDON, now continued with JASON-1.

One particular advantage of satellite altimetry, as compared to tide gauge readings, is that an absolute sea level can be determined, referred to Earth's center of mass, and independent of crustal motion. This contributes, for example, in connecting separated tide gauges and distinct height reference systems with the objective of creating a unified global height datum (Fig. 9.19). GPS observations at the tide gauges provide the precise geometrical connection between the stations [7.6.2.3].

Characteristic features of the *sea state* (waves and wind) can be analyzed from the shape and intensity of the return pulse wave form. A calm sea is a good reflector and returns a strong pulse; rough seas scatter the signal and hence return a weak pulse. The
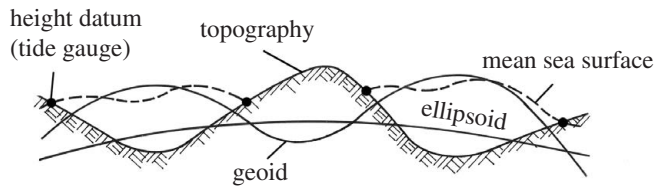
Figure 9.19. Relationship between separated tide gauges

amount of scattered power can be used to identify the sea state, observe sea ice extent and the rate of accumulation of snow on glaciers (Heidland, 1994; Zwally, Brenner, 2001).

Due to the near-polar orbits of altimeter missions like ERS-1 [9.2], significant contributions were expected to the mapping of ice topography and ice sheet variability. For ERS-1 two particular ice-phases, with a 3-day repetition cycle, were realized. The orbital parameters allowed mapping of the whole Greenland ice sheet and about 80% of Antarctica. ERS-2 and ENVISAT continue the observations.

Polar ice sheets and sea ice play an important role in the global climate system, because of their high albedo and their role as a large store of fresh water. With ERS altimeters it has been possible to monitor ice sheet mass balance and changes in sea ice thickness since 1992. This monitoring has already revealed a significant thinning of Antarctic glaciers (Benveniste et al., 2001), that can continue to be observed with ENVISAT.

A current difficulty in mass balance estimation is controlling the wave penetration within the snowpack (Rémy et al., 2001). Two-frequency altimeters will improve the estimation of ice accumulation. Further improvement of our knowledge on ice mass balance is expected from the GLAS laser altimeter on ICESAT [8.7] and the CRYOSAT altimeter, devoted to the survey of ice sheets and sea ice.

10   Gravity Field Missions

10.1   Basic Considerations

According to Newton's law of gravitation the attraction between two particles is proportional to the product of their masses and inversely proportional to the square of the distance between them [3.1.2]; accordingly, the gravitational effect of mass anomalies on the orbit of a near-Earth satellite must decrease with the square of increasing orbital height. The relationship is explained simply in Fig. 10.1 (Rummel, 1986).

A mass inhomogeneity at 1 km depth causes a certain gravity signal at the point P on Earth's surface. In order to generate an identical signal at P, a mass inhomogeneity at 10 km depth must be one hundred times as large. To produce the same signal at a satellite orbiting at 200 km altitude, the inhomogeneity must be stronger by a factor of 40 000. This consideration demonstrates that only a highly sensitive satellite sensor is capable of measuring small inhomogeneities of Earth's gravity field.
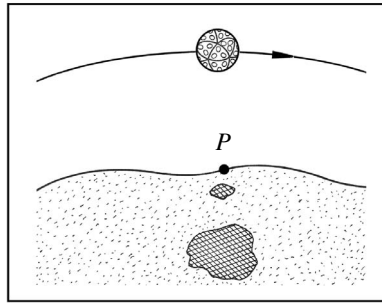
Figure 10.1. Mass inhomogeneity and satellite orbit

The terrestrial gravity field is usually expressed in terms of a series of spherical harmonics up to a maximum degree, N [3.2.2.1], [12.2], which can be associated with a shortest resolvable wavelength, λ, at Earth's surface according to

$$\lambda = \frac{360}{N} [^\circ]. \tag{10.1}$$

An equivalent representation refers to a certain block size, S, on the sphere in relation to representative mean values, such as mean free air anomalies (Torge, 2001). The resolution of the associated gravity field expansion is as given in Table 10.1. In many cases the half wavelength, $\lambda/2$, is considered.

The factor $(a_e/r)^n$ in equation (3.116) describes attenuation of the field with the altitude of the satellite orbit. The series in (3.116) is usually truncated at the maximum resolvable degree, $n = N$, which is then transformed into the corresponding spatial resolution with the wavelength, λ, or half wavelength, D, given in km with $D = 20\,000/N$. The degree of development is, for example,

$N = 25$ for LAGEOS $(h = 6000$ km$)$,
$N = 50$ for STARLETTE $(h = \ \ 950$ km$)$,
$N = 60$ for ERS $(h = \ \ 780$ km$)$.

Table 10.1. Typical subdivision of the gravity field expansion

Subdivision	long	mean	short	very short
wavelength, λ [km]	> 8000	> 1000	> 200	< 200
degree and order, N, of the representation	< 5	< 36	< 200	> 200
block size, S [degrees] (mean anomalies)	> 10	> 5	> 1	< 1

For orbit determination of high satellites, hence, only the long-wavelength components of the gravity field are required.

For gravity field modeling, the orbits of a large number of satellites with altitudes between 800 km and 20 000 km have been analyzed, using such geodetic observation techniques as cameras (directions) [5.1], Doppler [6], SLR [8], PRARE [4.3.3.2], and GPS [7].

Current gravity field models, based on such analyses, have a spatial resolution of about 500 km (half wavelength), corresponding to a degree of $N = 36$. The related accuracy is about 1 m for the geoid, and 5 mGal for the gravity anomaly. One example is the model GRIM5-S1 (Schwintzer, et al., 2000), see Fig. 12.6, p. 521

With the inclusion of satellite altimetry [9] and surface gravity data, models can be improved. However, due to the inhomogeneity of surface data and the discrepancy between geoid and mean sea level, improvement in global geoid models below the level of about 1 m cannot be expected with conventional methods.

Today's requirements in geodesy, geophysics and oceanography for a global, fine structured gravity field, and a related geoid model, however, are about two orders of magnitude higher, namely

resolution $N > 200 \mathrel{\widehat{=}} \lambda < 200$ km,
accuracy   <1 mGal for gravity anomalies,
       $< \pm 1 \dots 2$ cm for geoidal heights.

Such a gravity field model would have the same level of accuracy as is routinely attained for global position determination. Some of its contributions to different fields will be (Schuyer, 1997):

Geodesy   global vertical datum, height determination with GPS,
       ice and land vertical movements,
Geophysics   processes in the Earth's core and mantle,
       continental lithosphere (post-glacial rebound),
       ocean lithosphere (subduction processes),
Oceanography absolute circulation, sea level change, climatic change.

The only foreseeable possibility for meeting these requirements are dedicated high-resolution gravity-field missions. The essence of these is the use of satellites as gravity

probes in Earth's gravity field. To overcome the limitations of ground tracked satellites, as they are present in traditional tracking techniques, three fundamental criteria have to be fulfilled:

- orbit altitude as low as possible (200 to 500 km),
- uninterrupted tracking over large orbital arcs in 3 spatial dimensions, and
- discrimination between gravitational and non-gravitational forces acting on the satellite.

Two concepts are under consideration, and have already been tested (see also Rummel et al. (2002)). These are:

- satellite-to-satellite tracking (range and range-rate measurements between satellites), and
- satellite gravity gradiometry (measurement of gravity differences within the satellite).

In the first concept we distinguish between the high-low and the low-low configuration. Fig. 10.2, (a) to (c), demonstrates all three techniques.
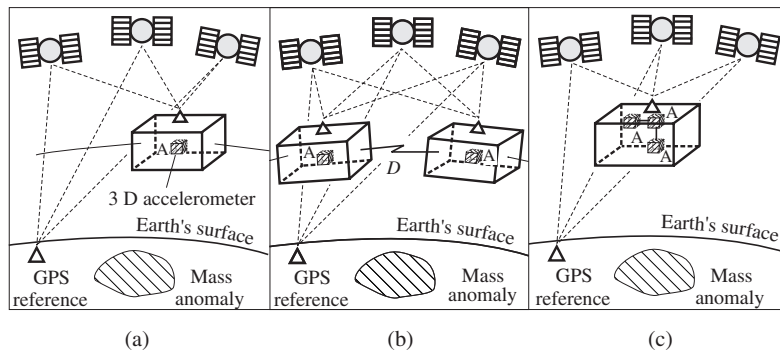
Figure 10.2. Different concepts of dedicated gravity field missions; SST-HL (a), SST-LL (b), SGG (c), after Rummel et al. (2002)

Satellite-to-satellite tracking in the *high-low mode* (SST-HL, Fig. 10.2 (a)) means that a LEO spacecraft is tracked by high orbiting satellites like GPS, GLONASS or GALILEO, relative to a network of ground stations. The non-gravitational forces acting on the low orbiter are measured by accelerometers. The LEO is a probe in the Earth's gravity field which can be precisely tracked without interruption. The observed 3-D accelerations correspond to gravity accelerations.

Satellite-to-satellite tracking in the *low-low mode* (SST-LL, Fig. 10.2 (b)) means that two LEO satellites are placed in the same low orbit, separated by several hundred kilometers, and that the range D between both spacecrafts is measured by an inter-satellite link with the highest possible accuracy. Again, the effect of non-gravitational forces acting on the two LEOs can either be measured or compensated (see [4.3.3.1]). In essence, the acceleration difference between the two LEOs is measured. The LL-configuration can be combined with the HL-concept. One advantage over the pure HL-technique is that differencing of observables provides a much higher sensitivity.

Satellite gravity gradiometry (SGG, Fig. 10.2 (c)) means that acceleration differences are measured directly in the satellite. Since the spacecraft is in free fall, the accelerations have to be measured away from the satellite's center of mass, ideally in all three dimensions. One important advantage, compared with the SST technique, is that non-gravitational accelerations are the same for all measurements inside the spacecraft and hence vanish by differencing.

In the first case (SST-HL), the first derivatives of the gravitational potential are measured, and in the second case (SST-LL) the difference of the first derivatives over a long baseline. In the third case (SGG) the second derivatives are determined. In short, the methods can be characterized as (Rummel et al., 2002):

SST-HL   measurement of accelerations of one LEO,
SST-LL   measurement of acceleration differences between two LEOs,
SGG   in situ measurement of acceleration gradients within one LEO.

A large number of proposals has been elaborated for all three concepts during the last 30 years or so, among them the *Geopotential Research Mission* (GRM), (Keating et al., 1986), *ARISTOTELES* (Visser et al., 1994) or *STEP* (Satellite Test of the Equivalence Principle). For an overview see Sneeuw, Ilk (1997). Although these projects were not realized, the principles developed for them have nevertheless entered most of the existing or planned dedicated gravity field missions.

It is remarkable that all three above mentioned techniques could be or will be realized during the first decade of the new century with the missions CHAMP, GRACE, and GOCE. This period is therefore dubbed the *Decade of Geopotential Research*. The missions have different characteristics and hence satisfy different aspects of high precision gravity field determination. Fig. 10.3 gives an impression. All three missions will considerably improve the best existing gravity field model, EGM96, by several
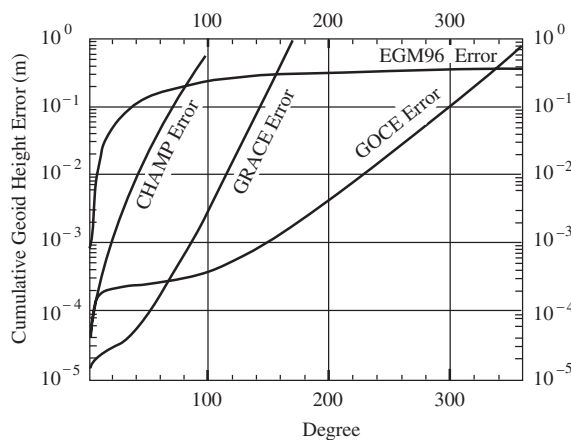
Figure 10.3. Cumulative geoid errors for EGM96 gravity field and dedicated gravity field missions, after Gruber et al. (2000)

orders of magnitude: CHAMP up to coefficients of degree and order 70, GRACE up to about 140, and GOCE up to about 350. Whereas GRACE shows the highest accuracy for the low harmonics up to 70, and hence can detect gravity field variations with time at this scale, GOCE shows best performance between degrees 70 and 350, and can hence also provide a 1 cm geoid for short half wavelengths down to about 80 km. More details about concepts and missions are given in the next two sections and in the cited literature.

10.2 Satellite-to-Satellite Tracking (SST)

10.2.1 Concepts

In this method, range changes between two satellites are measured with a very high resolution. The method belongs to group (3) in [1.2] (*Space to Space*). Two concepts can be applied, as has been indicated before:

(1) *high-low concept*, between a high-orbiting satellite (geostationary, GPS, GLONASS or GALILEO) and a low-orbiting spacecraft, possibly launched from the space shuttle (Fig. 10.2 (a)), and

(2) *low-low concept*, based on two satellites following each other along the same orbit, a few hundred kilometers apart (Fig. 10.2 (b)).

For both concepts, the spacecraft in the low orbit are the sensors in Earth's gravity field. One-way and two-way microwave intersatellite tracking systems can be used to measure the relative velocities. The irregular variations of this velocity contain gravitational information. The lower the satellite's orbit the more pronounced and detailed this information becomes. The basic observable is the radial velocity range rate between the two spacecraft (Rummel et al., 1978):

$$\dot\rho = \dot{X}_{12} e_{12}. \tag{10.2}$$

The inter-satellite range is ρ; $\dot{X}_{12} = \dot{X}_2 - \dot{X}_1$ is the difference in the velocities of the two satellites, S_1 and S_2 (Fig. 10.4), and

$$e_{12} = \frac{X_2 - X_1}{|X_2 - X_1|} = \frac{X_{12}}{\rho} \tag{10.3}$$

is the unit vector from S_1 to S_2. The range rate change is then (Reigber et al., 1987):

$$\begin{aligned}\ddot\rho &= \ddot{X}_{12} e_{12} + \dot{X}_{12} \dot{e}_{12} \\ &= \ddot{X}_{12} e_{12} + \dot{X}_{12}(\dot{X}_{12} - \dot\rho e_{12})\rho^{-1} \\ &= \ddot{X}_{12} e_{12} + ((\dot{X}_{12})^2 - (\dot\rho)^2)\rho^{-1}\end{aligned} \tag{10.4}$$

because

$$\dot{e}_{12} = \frac{d}{dt}(X_{12}\rho^{-1}) = (\dot{X}_{12} - \dot\rho e_{12})\rho^{-1} = C\rho^{-1}. \tag{10.5}$$
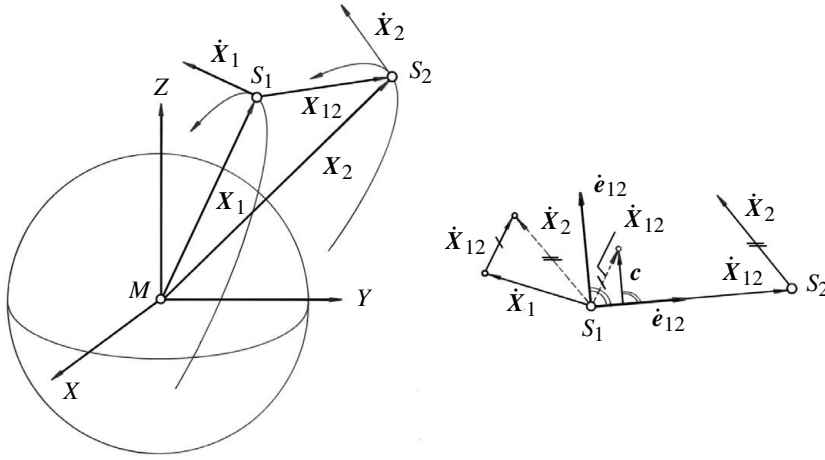
Figure 10.4. Relationship for satellite-to-satellite tracking

The observable, $\ddot{\rho}$, is then of particular interest if $\dot{e}_{12}$ is very small, i.e. if the change in orientation is low. This is the case for the "high-low" concept.

The rationale behind the method of satellite-to-satellite tracking is to establish a relationship between the parameters of the terrestrial gravity field and the observables $\dot{\rho}$ and $\ddot{\rho}$. Usually the spherical harmonics, C_{nm} and S_{nm}, of the expansion (3.109) are taken for such parameters, β_n. Approximate values, X^c, $\dot{X}^c$ and $\dot{\rho}^c$, $\ddot{\rho}^c$, are derived from initial conditions for both satellites, and the gravity field is approximated by β_n^c, hence

$$\beta_n = \beta_n^c + \Delta\beta_n, \qquad n = 1, \ldots, N. \tag{10.6}$$

Linearizing (10.2) gives

$$\Delta\dot{\rho} = \dot{\rho} - \dot{\rho}^c = \frac{\partial}{\partial\beta_n}(\dot{X}_{12}e_{12})\Delta\beta_n = \left(e_{12}\frac{\partial\dot{X}_{12}}{\partial\beta_n} + \rho^{-1}C\frac{\partial X_{12}}{\partial\beta_n}\right)\Delta\beta_n. \tag{10.7}$$

Accordingly, equation (10.4) develops to

$$\Delta\ddot{\rho} = \ddot{\rho} - \ddot{\rho}^c = \frac{\partial}{\partial\beta_n}((\ddot{X}_{12}e_{12} + ((\dot{X}_{12})^2 - (\dot{\rho})^2))\rho^{-1})\Delta\beta_n \tag{10.8}$$

$$= \left(e_{12}\frac{\partial\ddot{X}_{12}}{\partial\beta_n} + 2\rho^{-1}C\frac{\partial\dot{X}_{12}}{\partial\beta_n} + \rho^{-1}(a - 2\dot{\rho}\rho^{-1}C - \rho^{-1}(C)^2e_{12})\frac{\partial X_{12}}{\partial\beta_n}\right)\Delta\beta_n,$$

where

$$a = \ddot{X}_{12} - (\dot{X}_{12}e_{12})e_{12}$$

is the acceleration vector orthogonal to the interconnecting line between both satellites. The observables $\dot{\rho}$ and $\ddot{\rho}$ are not measured directly, but derived from signal processing. The subsequent numerical development is rather complicated and tedious. Several research groups have developed solution concepts (cf. Rummel (1979); Krynski (1983); Colombo (1984); Ilk (1990); Jekeli (2000)). A number of publications with new approaches certainly will appear in connection with the current gravity field missions.

Fig. 10.5 gives an example (Douglas et al., 1980), of how a $1° \times 1°$ anomaly of 10 mGal produces a velocity change in the along-track component of a satellite at 200 km altitude. The low frequency disturbance over nearly the complete revolution can hardly be separated, because all block-anomalies contribute in a similar manner. The distinct high-frequency disturbance can be discriminated if the second satellite is sufficiently separated ($> 1°$) from the first one. Fig. 10.5 (right) shows how a given disturbance generates a change in the range between a satellite pair flying in-echelon. With increasing separation ($> 1° \triangleq 100\,\text{km}$) the disturbance signal becomes more significant.
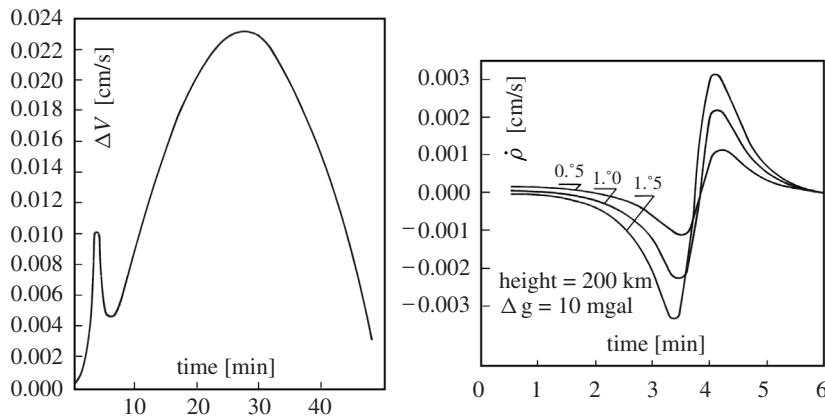
Figure 10.5. Effect of a $1° \times 1°$ 10 mgal anomaly on the velocity of a satellite at 200 km altitude (left), and on the range rate between a pair of satellites with 0.5°, 1.0°, and 1.5° separation (right)

The basic problem of the method becomes clear from the diagrams. The orbital expression of a $1° \times 1°$ anomaly of 10 mgal is very small. The influences of neighboring blocks (mass inhomogeneities) are superimposed upon each other. In order to resolve detailed structures ($< 10\,\text{mgal}$) of the gravity field with associated wavelengths $< 200\,\text{km}$ it is necessary to observe and to analyze range rates with a resolution of better than $10\mu\text{m/s}$.

10.2.2 High-Low Mode, CHAMP

The concept of satellite-to-satellite tracking was proposed and tested quite early in the 1960s. SST in the high-low mode was applied during the NASA lunar APOLLO program for Earth-based control of the lunar orbiter (Vonbun, 1977b). Subsequent analysis of the data led to the discovery of strong anomalies in the lunar gravity field (Sjögren et al., 1972).

With respect to Earth's gravity field, SST in the *high-low mode* was tested in 1975 with measurements between the geostationary satellite ATS-6 and the low orbiting space vehicles GEOS-3, NIMBUS-6, and APOLLO-SOYUZ.

From a comparison between measured range rates, $\dot{\rho}^m$, and the computed range rates, $\dot{\rho}^c$, based on a global gravity model (GEM 7), the anomalous gravity structures of the Java Trench and the Himalayan mountains, for example, were clearly visible (Fig. 10.6). The efficiency of the method was demonstrated in this test. However, a
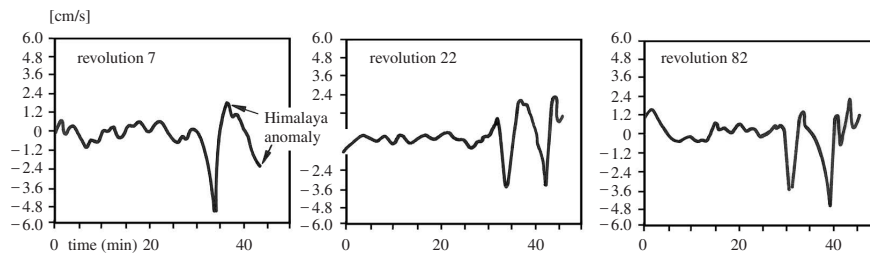
Figure 10.6. Recovery of anomalous gravity structures from range-rate observations in the high-low mode (Vonbun, 1977b)

dedicated satellite mission with a much higher resolution in the range and range-rate observations has only been realized after about 25 years, with the CHAMP mission.

The *Challenging Mini-Satellite Payload for Geophysical Research and Application* (CHAMP) was launched under the scientific responsibility of the GeoForschungsZentrum (GFZ), Potsdam, Germany, on July 15, 2000, into an almost circular near polar orbit of about 450 km altitude and an inclination of about 87.3 degrees. The design lifetime of the satellite is 5 years. Due to atmospheric drag, the altitude will decrease over the mission time to about 300 km or less. This change in altitude is intentional and makes the satellite sensitive to a broad variety of coefficients. The spacecraft (Fig. 10.7) only weighs 500 kg.
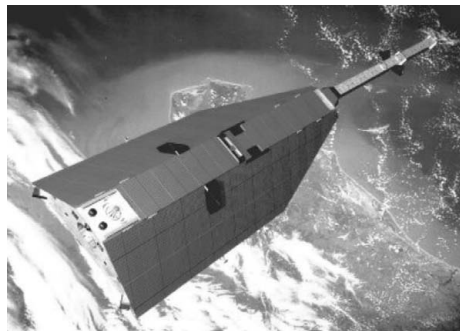
Figure 10.7.   Challenging Mini-Satellite, CHAMP; courtesy GFZ

The main scientific goals of the mission are:
- mapping of the global gravity field,
- mapping of the global magnetic field, and
- profiling of the ionosphere and troposphere.

To achieve these goals the satellite carries the following scientific instruments:
- a space-borne 16 channel dual-frequency GPS receiver connected to a multiple-antenna system,
- a three-axis accelerometer (at the spacecraft's center of mass) to measure the non-gravitational accelerations acting on the spacecraft,
- a laser-retro reflector (LRR) for backup tracking from the ground,
- a magnetometer, and
- a digital ion drift meter.

The low orbiting CHAMP satellite is a sensor in free fall in Earth's gravity field. The gravitational orbit perturbations are continuously monitored with respect to the high orbiting GPS satellites, using precise GPS orbits [7.4.3.2] based on a worldwide tracking network [7.8.1]. The concept of differential GPS [7.5] can be applied and provides position and velocity information for the CHAMP spacecraft with an accuracy of a few centimeters.

CHAMP is not a drag-free satellite [4.3.3.1]. For gravity field modeling the gravitational perturbations alone are required, hence the non-gravitational perturbations from drag, solar radiation pressure, albedo, thrust and so on [3.2.3] have to be measured independently. This is done by the three-axis STAR accelerometer from ONERA (Touboul et al., 1998), with a resolution of about $3 \cdot 10^{-9} \text{m/s}^2$. To avoid misalignment, the accelerometer has to be placed as close as possible to the spacecraft's center of mass, and the satellite's orientation has to be controlled by star sensors [5.3.1].

With CHAMP data it is expected to improve the accuracy of existing gravity field models at long and medium wavelengths by a factor of about 5 to 10 (Sneeuw, Ilk (1997); Gruber et al. (2000), see also Fig. 10.3).

Earth's magnetic field is measured by scalar and vector magnetometers, fixed to the end of a 4 m boom, together with the star sensors (Reigber et al., 1999). Two particular GPS antennas at the rear of CHAMP receive signals from setting GPS satellites at the spacecraft's horizon; these signals are used for the technique of *limb sounding* [7.6.2.9].

For more details on the CHAMP mission see e.g. Balmino et al. (1999); Reigber et al. (1999) and the forthcoming literature on CHAMP results.

10.2.3 Low-Low Mode, GRACE

The first experiment in the *low-low mode* was carried out during the APOLLO-SOYUZ rendezvous maneuver in 1975 (Vonbun, 1977b). The results, however, were not significant because of the low resolution in the observables. NASA was for several years developing a promising low-low mission under the name *Geopotential Research*

Mission (GRM) (Keating et al., 1986). The mission is no longer being pursued. Nevertheless, the GRM concept is still a viable technical option for precisely modeling Earth's gravity field. In this concept two co-orbiting space vehicles would be positioned from the Space Shuttle in a 160 km altitude circular polar orbit at an adjustable separation between 150 and 550 km.

The selected orbital height is always a compromise between the technical effort involved in maintaining a low orbit, and the desired resolution of the gravity field. Fig. 10.8 demonstrates this relationship for two orbital heights, 160 km and 200 km. It is evident that the relative velocity between both satellites must be measured with an accuracy of about 1 μm/s.

Because of the low orbital height the space vehicles are exposed to strong surface forces, in particular air drag [3.2.3.3]. For mapping of the gravity field, however, it is required that only the velocity changes that are due to gravitational effects are measured. The spacecraft must hence either carry a *Disturbance Compensation System* (DISCOS) [4.3.3.1] which measures and immediately corrects the dislocation of the space vehicle caused by external forces, or the non-gravitational forces have to be measured independently with a three-axis accelerometer.

Figure 10.8. SST signal, $1° \times 1°$ 1 mgal gravity anomaly, GRM mission

A rather large amount of propellant is required for orbit corrections, and to maintain the orbital height. An alternative is to start with a larger altitude and to accept the decrease of the semi-major axis during the mission's total lifetime.

Instead of GRM, the GRACE mission, with very similar parameters and objectives has been realized. GRACE is a joint project between NASA and the German Space Agency, DLR. The name stands for *Gravity Recovery and Climate Experiment*. Besides the high resolution precise mapping of Earth's gravity field, the secondary science objective of GRACE is *limb sounding* for the determination of tropospheric and ionospheric parameters [7.6.2.9].

Two identical satellites were launched on March 17, 2002, into a near-polar orbit of about 500 km altitude with an inclination of 89°. In

Figure 10.9. GRACE mission

the nominal configuration the satellites fly in-echelon, 220 km apart, within ±50 km. Orbit maneuvers are necessary every one or two months in order to maintain the

separation between the two spacecraft. The design lifetime of the mission is 5 years. Each satellite is about 3 m long, weighs about 480 kg, and carries the following scientific instruments (mostly redundant):
  − ultra-stable oscillator (USO),
  − GPS receiver,
  − accelerometer (SuperSTAR),
  − K-band ranging system (KBR),
  − star cameras, and
  − laser retro reflectors (LRR).

The GPS receivers can track up to 10 satellites and provide navigation data as well as range and range-rate in the high-low mode. As with CHAMP, the STAR accelerometer is required to separate gravitational and non-gravitational disturbances. The sensor unit consists of a metallic proof mass inside an electrode cage. In order to make precise measurements of the non-gravitational accelerations, the proof mass must be located within 0.1 mm of the center of gravity of the spacecraft. The LRR array is used for precise absolute orbit determination, and the star cameras are required for precise pointing of the satellites towards one another.

The key instrument is the *K-Band Ranging System* (KBR). Each satellite transmits carrier phase signals to the other satellite at two frequencies (24 and 32 GHz), allowing for ionospheric corrections. Two one-way ranges between both satellites are obtained, each by comparing the on-board generated phase with the received phase. Both phases are generated by the same ultra-stable oscillator. The ranges are obtained at a sampling rate of 10 Hz, and are then filtered to produce range-rates and range-rate-rates (accelerations in the line of sight) at a sampling rate of 0.1 Hz. The estimated accuracy of the filtered range-rate is 10^{-6} m/s (Jekeli, 2000).

In essence, the twin GRACE satellites can be considered as one instrument in which
  − variations in the gravity field cause variations in the range between the two satellites; areas of stronger gravity will affect the lead satellite first and accelerate it away from the following satellite,
  − range variations are measured by a high-accuracy microwave link; the relationship to the global reference frame is given by GPS, and
  − the observed range variations are corrected for non-gravitational effects by a precise accelerometer.

The observations will produce monthly global gravity maps with a spatial resolution of about 300 km on the ground, and a precision superior by a factor of up to 100 over existing models (see Fig. 10.3). Besides mapping a static global gravity field down to mean wavelengths with unprecedent accuracy, GRACE will in particular be able to monitor fluctuations in the gravity field. Changes in the geoid can be monitored to a sub-millimeter level per year. These variations contain information on changes in the distribution of masses between the atmosphere, oceans and solid Earth, and contribute to the monitoring of
  − surface and deep currents in the ocean,
  − ground water storage on land masses,

- mass variations within the Earth, and
- exchange between ice sheets, glaciers and the oceans.

The GRACE concept can also be regarded as a one-dimensional gradiometer with a very long baseline. The original GRACE observations can be used to derive gravity gradients with an accuracy comparable to the planned gradiometer missions [10.3], (Keller, Heß, 1999).

Another of GRACE's mission goals is to provide a better knowledge on the atmosphere by limb sounding [7.6.2.9].

10.3   Satellite Gravity Gradiometry

10.3.1   Concepts

A gradiometer is a sensor that can measure the change of the gravity acceleration in space, i.e. the gravity gradient. The first derivatives of Earth's gravitational potential $V = V(X, Y, Z)$ are given with the vector, $\boldsymbol{g}$, of the gravity acceleration. A gradiometer is hence capable of measuring the second derivatives. In total, the second derivatives, given by

$$V_{ij} = \frac{\partial^2 V}{\partial i \, \partial j}$$

form a tensor, the *gravity gradient tensor* or *Eötvös-tensor*,

$$V'' = \begin{pmatrix} V_{XX} & V_{XY} & V_{XZ} \\ V_{YX} & V_{YY} & V_{YZ} \\ V_{ZX} & V_{ZY} & V_{ZZ} \end{pmatrix}. \tag{10.9}$$

X, Y, Z is an orthogonal triple. Only five of the 9 elements in the Eötvös-tensor are mutually independent. It holds that

$$V_{XY} = V_{YX}, \quad V_{XZ} = V_{ZX}, \quad V_{YZ} = V_{ZY}, \tag{10.10}$$

as does the Laplace condition (i.e. a vanishing trace of the tensor):

$$V_{XX} + V_{YY} + V_{ZZ} = 0.$$

A gravity gradiometer which measures all of the elements contained in the tensor (10.9) is called a *full-tensor gradiometer*. The components of the tensor describe the local structure of the gravity field by the curvature of this field. This implies a conceptual superiority of the gradiometer if compared with other sensors for the mapping of the gravity field.

The development of gravity gradiometers can be traced back to the Hungarian baron *Roland von Eötvös*, who built a stationary torsion balance in about 1900, based on the early work of *Cavendish* (1731–1810) and others (Torge, 1989). Eötvös was able to measure one part of the components of the second derivatives of the gravitational field at the surface.

A *torsion balance* consists of two equal masses arranged at different heights and connected by a rigid system of negligible mass, which is suspended by a torsional thread (Torge, 1991). Due to the unequal gravitational acceleration of the two masses, the suspended beam experiences a torsional moment that can be measured. The torsion balance was used worldwide in geophysical fieldwork during the first third of the last century, before it was substituted by the easier to handle relative gravimeters. The main difficulty in the use of the torsion balance is the strong influence of topographic masses near the instrument.

A fresh impetus to the development of gradiometry came from the requirements of space flight, in particular for the determination of directions under free-fall conditions (Forward, 1974). New gravity gradiometer concepts were developed, now also applied for continuous dynamic measurements in satellites and aircraft.

Under free-fall conditions inside a satellite, one can only measure the difference in the acceleration of gravity between the points where an accelerometer is located, and the center of mass of the spacecraft. The situation can be described as a *tidal field* in the satellite-fixed reference frame (Ilk, 1990). The observations are, however, only meaningful if the locations of the accelerometer and of the center of mass are precisely known. Since this is not possible, two accelerometers are used, precisely located with respect to each other, and the difference in the gravity acceleration is observed. Several problems arise:

- the satellite, with the instruments in it, rotates,

- the orbit has to be precisely known,

- the orientation of the instrument and spacecraft with respect to an external frame is needed,

- the data are corrupted by external forces (e.g. drag, radiation pressure) and instrumental errors (drift, scale errors, etc.), and

- the accuracy requirements are extremely high.

For a detailed discussion see e.g. Colombo (1989); Aguirre-Martinez, Cesare (1999); Müller (2001).

Changes in the gravity acceleration are measured in the *Eötvös* unit (E.U.), (1 E.U. $\cong 10^{-9}\,\mathrm{s}^{-2}$). With respect to current gravity field missions, the term 1 mE is also used for 10^{-3} E.U. Converted to the units that are usually applied in gravity, 1 E.U. corresponds to a change of $10^{-9}\,\mathrm{ms}^{-2}$, or $0.1\,\mu$Gal, over 1 meter. It can be derived from simulation studies that a gravity anomaly of 1 mGal and 100 km × 100 km size at the surface produces a signal of 0.001 to 0.0001 E.U. at 200 km altitude (Balmino, Bernard, 1986). Fig. 10.10 explains the relationship. From this consideration, the following requirements can be derived for a satellite-based gradiometer mission (Rummel, 1985):

orbital height	160–240 km
orbital inclination	$\sim 90°$
eccentricity	< 0.001
mission duration	> 6 months
resolution of the gradiometer	10^{-2} to 10^{-4} E.U.
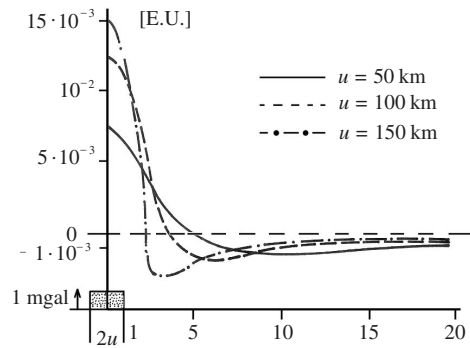
Figure 10.10. Signal of the vertical gravity gradient at 200 km altitude, caused by a 1 mGal isolated anomaly at Earth's surface (after Balmino, Bernard (1986))

10.3.2   GOCE mission

Several proposals and studies over the last about 20 years, for example Balmino et al. (1984) or Rummel, Schrama (1991), finally led to the GOCE mission which is scheduled for launch early in 2006. GOCE stands for *Gravity Field and Steady-State Ocean Circulation Explorer* and forms part of the ESA *Earth Explorer* program.

The mission is based on a sensor fusion technique (see Fig. 10.2 (c)), namely a combination of very precise orbit determination using GPS high-low SST, and satellite gravity gradiometry (SGG). GOCE will be flown in a nearly circular near polar sun-synchronous orbit of about 97° inclination and an altitude of 240–250 km. The satellite will have a launch mass of about 1000 kg and a small cross section of about 0.9 m^2. It will be totally symmetrical to minimize the influences of non-gravitational surface forces. The mission duration will be about 2 years. Continuous data will be generated in two eclipse-free cycles of 6 months each. The main objective is to measure the geoid with an accuracy of about 1 cm, gravity anomalies of 1mGal and a spatial half-wavelength resolution of about 70 km. The two core instruments of GOCE are

  − a GNSS (GPS/GLONASS) receiver, and
  − a gravity gradiometer.

The *GNSS receiver* will play twin roles. It will be used for precise orbit determination (location of the gravity gradiometer) at the 1 cm level, and for an analysis of the long and medium wavelength features of the gravity field by the SST-HL technique [10.2.2].

The *gravity gradiometer* consists of three pairs of highly sensitive accelerometers, located in the close vicinity of the satellite's center of mass. Observables are differences of accelerations over a short baseline of about 50 cm. The six three-axis accelerometers are mounted in a so-called *diamond configuration* (Sneeuw et al., 2001). Two accelerometers are placed on each axis of the instrument triad.

In principle, two types of gradiometer design exist, the inductive and the capacitive. The inductive gradiometer operates inside a superconduction environment, the capacitive at ambient temperature. A capacitive gradiometer will probably be used for GOCE, because of its higher maturity (Müller, 2001). The sensitivity is about 3 to 5 mE/$\sqrt{\text{Hz}}$.

The spacecraft will be drag-free, that is, the non-gravitational disturbances are compensated by a control system using thrusters. Further, the spacecraft will be kept Earth oriented with its

- x-axis pointing along track,
- y-axis pointing cross-track, and
- z-axis pointing radially outwards.

The observations are hence made on a rotating platform and inertial forces (centrifugal and Euler) are sensed. The observed gradient signal is

$$\Gamma = V'' + \Omega\Omega + \dot{\Omega}. \tag{10.11}$$

The first two terms, $\Omega\Omega$, are symmetric and $\dot{\Omega}$ is antisymmetric. For details see Müller (2001). The angular velocity, Ω, and the angular acceleration, $\dot{\Omega}$, in the Earth-pointing mode are of the order of 10^{-3} rad/sec and 10^{-5} rad/sec^2, respectively. In the inertial configuration they would be about zero. The orientation of the spacecraft has to be known with an accuracy of $5 \cdot 10^{-3}$ rad/$\sqrt{\text{Hz}}$, which will be controlled by star trackers.

The design of GOCE is not completely finalized. Until the launch, certain modifications still may occur. In particular, more sophisticated gradiometer error models will be developed as well as calibration techniques and refined data analysis procedures.

The recovery of a gravity field with high spatial resolution, e.g. up to degrees of 200 or 300, implies 50 000 to 100 000 unknown coefficients. There are basically two alternative solution approaches under discussion, e.g. Rummel, Schrama (1991); Klees et al. (2000). In the *space-wise* approach the gradiometry observations are considered as gravity related functionals given on a spherical surface at the satellite altitude. The approach is analogous to the *geodetic boundary value problem*, and applies the methods and theoretical considerations devoted to this problem in recent years, cf. (Torge, 2001). One disadvantage of the space-wise method is the lack of connection to the orbit.

In the *time-wise* approach the gradiometer measurements are considered as a discrete time series, spanning the entire mission period. The spherical harmonic series of Earth's gravitational field are connected to the orbit via inclination functions [3.2.2.1]. In this way, a linear relationship between the observable functionals and the spherical harmonic coefficients is established (Colombo, 1989).

The mission objectives, as stated above, are quite challenging. A geoid resolution of 1 cm with a half wavelength of about 70 km will considerably contribute to geodesy and oceanography. The remaining omission error for small geoid structures will be at the 1 decimeter level (Torge, 2001, p. 274) and can be more reliably recovered by surface or airborne gravimetry. For GPS height-determination [7.6.2.3] a significant input can

be expected. In oceanography, a 1 cm geoid completely fulfills the requirements for large scale and meso-scale ocean circulation models [9.5.3].

The development of satellite gradiometry is still at its beginning. The forthcoming satellite gravity gradiometry missions will certainly give rise to further intensive theoretical developments and broad discussion in the literature.

11 Related Space Techniques

11.1 Very Long Baseline Interferometry

Very Long Baseline Interferometry (VLBI), is not, strictly speaking, a method of satellite geodesy. Nevertheless, the fundamentals of this technique and its possible applications in geodesy and geodynamics have been included in this book because

- VLBI is a geodetic space technique that is used solely, or in combination with other satellite techniques in the recovery of geodetic, astrometric, and geodynamic parameters,
- the observation and adjustment techniques for the geodetic use of the GPS [7] were significantly influenced by the VLBI technology, and
- satellite-borne VLBI missions [11.1.4] are in their initial stages of realization.

11.1.1 Basic Concept, Observation Equations, and Error Budget

The technique of long baseline interferometry was developed in radio astronomy with the objective of studying the detailed structure of compact radio sources with a high angular resolution (Hey, 1984; Wohlleben et al., 1991; Rohlfs, Wilson, 1996; Sovers et al., 1998). The frequencies usually applied are between 0.5 and 22 GHz (75 cm to 1.3 cm), in the so-called *radio window* of the terrestrial atmosphere [2.3.3]. Receivers for 43 GHz, 86 GHz and above are being added (Walker, 2000). In order to improve the rather low angular sensitivity and resolution of a radio telescope, the effective diameter of an antenna dish is amplified by interconnecting several individual telescopes. The approximate relation is

$$\varepsilon \approx \frac{\lambda}{d} \tag{11.1}$$

with

ε resolution,
λ wavelength of the particular radiation, and
d telescope diameter.

For the emission wavelength of cosmic hydrogen ($\lambda = 21$ cm), the telescope diameter, or the distance between the connected individual telescopes, must be at least 42 km if the required angular resolution, ε, is $1''$. Cable connections over these distances are technically difficult and very expensive.

VLBI overcomes this problem by linking the independently operating telescopes via precise atomic clocks. With this technique, the distance between the participating telescopes is no longer a problem, and even telescopes on different continents can be integrated. The maximum telescope size is nearly the diameter of the Earth, and the corresponding angular resolution is better than $0.001''$, or 1 mas (milliarcsecond).

A generic interferometer (cf. [4.2.6]) consists of two antennas, arranged at a fixed distance, b, the interferometer base, and an appropriate processing unit. The

radiowaves received at both antennas, if combined, may cause interferences (Hey, 1984).

The basic principle of VLBI is explained in Fig. 11.1. The radio signals arriving from a particular radio source, S, are recorded at both participating stations on magnetic tapes together with the timing signals from an ultra-stable oscillator. The tapes are then delivered to the VLBI data processing facility and played back through the correlator as if there had been direct connections from the individual telescopes. Correlation means that both data-streams are shifted stepwise against each other, and
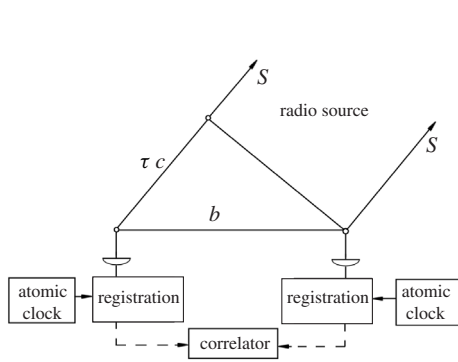
Figure 11.1. Basic principle of VLBI
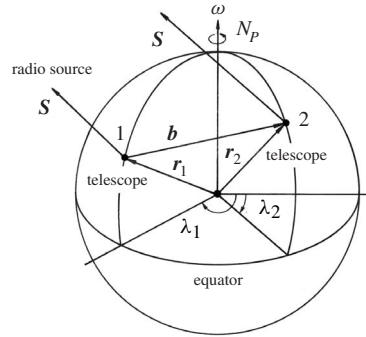
Figure 11.2. Geometrical relationship for VLBI

cross-multiplied, until both signal streams are perfectly aligned. The correlator output shows an interference pattern with maximum amplitude for signal alignment. The associated time delay, τ, is the difference in signal arrival time at the two antennas. The *signal delay*, τ, is a function of the time variable, t, defined by the station clocks, because the Earth, with the antennas, is moving in inertial space. The quantity $\tau(t)$ is the most important VLBI observable for geodetic applications.

Following the geometrical relationship in Fig. 11.1 and Fig. 11.2, and adding the main correction terms, we obtain the fundamental observation equation (Campbell, Witte, 1978; Campbell, 2000b):

$$\tau(t) = -\frac{1}{c}\boldsymbol{b} \cdot \boldsymbol{s}(t) + \Delta\tau_{\text{Ab}}(t) + \Delta\tau_{\text{Instr}} + \Delta\tau_{\text{Atm}}, \tag{11.2}$$

including a term for *diurnal aberration*:

$$\Delta\tau_{\text{Ab}}(t) = -\frac{1}{c}(\dot{\boldsymbol{r}}_2\boldsymbol{s})\tau(t), \tag{11.3}$$

and a correction for the bias and drift of the station clocks:

$$\Delta\tau_{\text{Instr}} = a_1 + a_2 t. \tag{11.4}$$

$\Delta\tau_{Atm}$ is a correction for tropospheric refraction, based on suitable models (cf. (Campbell, 1979, 2000a; McCarthy, 2000).

Following Fig. 11.2, the scalar product $b\cdot s(t)$ from equation (11.2) can be re-written with respect to an Earth-fixed Cartesian reference system (instantaneous equatorial system [2.1.2]), so that

$$bs(t) = b_x \cos\delta_s \cos h_s + b_y \cos\delta_s \sin h_s + b_z \sin\delta_z, \qquad (11.5)$$

with

b_x, b_y, b_z components of the baseline vector, b,
α_s, δ_s spherical equatorial coordinates of the radio source, S, and
$h_s = GST - \alpha_s$ Greenwich hour angle of the radio source.

In addition to the observable $\tau(t)$, which is primarily used in geodesy and geodynamics, the *fringe frequency*, $f(t)$, is obtained. The minima and maxima of the interferometric amplitudes show a certain frequency, caused by Earth rotation:

$$f(t) = \frac{1}{2\pi}\frac{d\Phi(t)}{dt}. \qquad (11.6)$$

The phase difference, $\Phi(t)$, is related to the propagation delay, $\tau(t)$, and to the received signal frequency, ν_0, by

$$\Phi(t) = 2\pi\nu_0\tau(t). \qquad (11.7)$$

Hence it follows that

$$f(t) = \nu_0\frac{d\tau(t)}{dt}. \qquad (11.8)$$

By analogy with (11.2) we find the observation equation of the fringe frequency:

$$f(t) = -\frac{\nu_0}{c}b\cdot\dot{s}(t) + \Delta f_{Ab} + \Delta f_{Instr} + \Delta f_{Atm}. \qquad (11.9)$$

According to (11.5) the equation for the scalar product becomes

$$b\cdot\dot{s}(t) = -\omega(b_x \cos\delta_s \sin h_s - b_y \cos\delta_s \cos h_s), \qquad (11.10)$$

where Earth's rotational velocity, ω, is the derivative in time of h_s.

An inspection of (11.10) shows that the baseline component, b_z, is not contained in the observation equation. The observable "fringe frequency" hence delivers only one part of the information which is of interest in geodesy.

The observable $\tau(t)$ is often referred to as *group delay*, whereas $\Phi(t)$ is called *phase delay* (cf. Shapiro (1978); Campbell (2000b) [2.3.1.2]). $\Phi(t)$ can be measured more precisely, but involves the problem of ambiguity resolution (cf. [7.3.2.3]).

The group delay is determined with high precision by using a large bandwidth. The group delay resolution is proportional to the inverse of the SNR (height of the main peak above the noise) and the total spanned bandwidth. If we increase the bandwidth at a given SNR by a particular factor, the group delay uncertainty will be reduced by the same factor. A major breakthrough for high precision geodetic VLBI came

with the *bandwidth synthesis technique*. Here, the total recorded bandwidth is split into several smaller units and distributed over the much wider receiver unit window. With modern analysis techniques, for the 8.4 GHz X-band frequency used for geodetic VLBI, a bandwidth of 720 MHz can be spanned. Fringe analysis allows determination of the group delay to about 1% of the peak width, corresponding to 15 picoseconds, or 5 mm (Campbell, 2000b).

Other than with GPS, the phase delay cannot be used successfully for geodetic VLBI because timing errors cannot be eliminated by single- or double-differencing [7.3.2]. The antenna can only be pointed to one source at a time. Simultaneous observation of different sources is not possible, because the signals emitted by quasars [2.1.2.1] are at least six orders of magnitude weaker than signals emitted by GPS satellites (Campbell, 2000a).

According to the observation equation (11.2) and (11.5), a total of $(3 + 2n)$ fundamental parameters are introduced into the basic parameter estimation process for a single baseline. These are

3 components of the baseline vector, b_x, b_y, b_z, and
2 coordinates, α, δ, for each radio source.

In the linearized observation model these are the corrections to the approximate values. In practice, more parameters are included in the adjustment process.

In a typical experiment in astrometry and geodesy several stations make simultaneous observations. Station location and clock parameters of one reference station are fixed. For each remaining station

− 3 site coordinates,
− 1 zenith tropospheric parameter, and
− 2 linear clock parameters

are introduced. Global parameters, common to the entire network, are, among others:

− Earth orientation (5 parameters: polar motion, UT1, nutation), and
− position of the radio sources (2 parameters).

The right ascension, α, for one radio source has to be kept fixed to obtain the origin in the celestial frame. In addition, a certain number of physical parameters and hardware effects have to be considered to fit the observations to the model. An extended list of parameters, including the already formulated principal terms in the observation equations, is (Shapiro, 1978; Campbell, 1979, 2000b):

(1) general and special relativity, aberration,
(2) precession, nutation,
(3) proper motion and structural changes in the radio sources,
(4) hardware effects:
      − oscillator instabilities,
      − signal delays in the receiver electronics,
      − deformation of the telescope,
(5) signal propagation effects:
      − ionospheric refraction,
      − tropospheric refraction,

(6) geodynamical effects:
 - polar motion,
 - variable Earth rotation,
 - solid Earth tides,
 - crustal deformation.

The mathematical and physical details of the individual parameters are exhaustively discussed in the literature, e.g. Shapiro (1978); Schuh (1987); Preuß, Campbell (1992). A comprehensive modern review is given by Sovers et al. (1998). An excellent short overview of the development of models from its basic form up to the extensive relativistic formulation is given by Campbell (2000b).

In accordance with current accuracy requirements, modeling of observables and determination of instrumental and environmental corrections have to be performed at the level of 0.1 mas, and several millimeters respectively.

The modeling of the basic geocentric observation equation (see 11.2) includes terms for diurnal and annual aberration, relativistic light deflection, and general and special relativity. The model is formulated in the barycentric reference system (Campbell, 2000b), (parameter group 1).

The orientation of Earth, with respect to the celestial system (precession, nutation), shows periodic variations with amplitudes of about 5 to 10 mas, when compared with the IAU 1980 nutation series [2.1.2.3]. The new *Precession-Nutation Model IAU 2000* will account for these effects, (parameter group 2).

The international celestial reference frame (ICRF) is based on radio sources (quasars) well outside our galaxy, ensuring minimal proper motions. The physical nature of quasars is still under debate, but numerous astrophysical studies during the last two decades have demonstrated that these compact extragalactic radio sources be inhomogeneous showing internal structure at the level of several mas. Changes in the structure of ICRF sources limit the accuracy of a reference frame based on them. Permanent monitoring of sources' structure, however, is possible with the same VLBI data alongside with the other analyses. Structure correction will become routine for extended radio sources used in geodetic/astrometric work (Sovers et al., 1998; Campbell, 2000b), (parameter group 3).

The necessary stability of the time base is achieved through an assembly of atomic clocks, including hydrogen masers [2.2.5]. Usually, second order polynomials are used as clock models. To account for sudden breaks, an additional delay is included that models the station clock behavior as a piecewise quadratic function of time (Sovers et al., 1998). The instrumental delay changes are monitored by the calibration system, which is part of the adjustment procedure. In general these effects can be modeled as clock errors. Large telescopes exhibit elevation dependent changes in the focal distance, which can be modeled to the level of millimeters (Campbell, 2000b), (parameter group 4).

Atmospheric effects on VLBI observations are considered to be the most critical factors limiting the achievable accuracy. VLBI stations are widely separated, hence the elevations of the telescopes during an observation session are quite different, as are the meteorological conditions along the signal paths. The ionosphere is a dispersive

medium for radio frequencies [2.3.3.1] and can hence be modeled by using two different observing frequencies. In geodetic VLBI, a frequency pair of 2.3 GHz (S-band) and 8.4 GHz (X-band) is selected throughout. In most cases, in order to model the influence of water vapor on tropospheric signal propagation [2.3.3.2], water vapor radiometers are used at VLBI stations. For the dry part of the correction, an appropriate mapping function has to be selected. The situation is similar to that in GPS. It can be expected that the forthcoming availability of near real-time global and regional tropospheric models, coming from ground-based and space-based GPS observations, will contribute to improved data correction (Schüler, 2001), see [7.4.4.2], (parameter group 5).

Of particular interest to applications in geodesy and geodynamics are the parameters of group (6). The motion of Earth's axis with respect to the crust (polar motion) has to be determined with the same accuracy as all the other parameters, i.e. 0.1 mas, corresponding to 3 mm. The same is true for the phase angle of Earth's rotation, which corresponds to the requirement to determine the UT1-variations (variations in the length of the day (LOD)) to better than 0.01 ms of time. Our understanding of the geophysical processes behind these variations has not yet reached this level of accuracy; hence the variations form part of the unknown parameters in the data adjustment.

The periodic crustal deformations caused by tidal effects could be seen rather early in VLBI data (Campbell, 2000b). Solid Earth tides cause diurnal and semidiurnal variations with vertical amplitudes of about 40 cm and horizontal displacements of about 10% of the vertical effect. The tidal loading effects of the oceans amount to about a decimeter for coastal and island sites. Models are being improved (Zahran, 2000).

The VLBI stations are also subject to horizontal and vertical crustal motion associated with plate tectonics. In order to define a terrestrial reference frame, a priori constraints are required, for example the "no-net rotation" assumption [12.4.1]. Problems may arise when different sets of defining stations are selected in the global network (Campbell, 2000b).

The parameters of group 6, as mentioned above, together with the parameters of group 2 and the radio source positions (group 3) are of particular interest in geodesy, astrometry and geodynamics. They are dealt with in more detail in section [11.1.2].

Currently, two modes of observations are carried out: 24-hour multi-station sessions and 90-minute single-baseline sessions. While the first mode is appropriate to provide all components of Earth orientation parameters, and contributes to celestial and terrestrial reference frames, the second is used for rapid determination of UT1-UTC (see [11.1.2]). Within one session, various sources (quasars) are observed following a pre-programmed schedule.

The correlation and adjustment process is done with a sophisticated hard- and software installation at particular processing centers. The Mark III system has been widely used since about 1980, with several refinements until 1999. The transition to the next generation, Mark IV, configuration started in 2000. With the Mark III data recording and processing system it was possible to record a data stream of 112 Megabits per second on 28 parallel tracks of tape. The Mark IV system is designed to

handle 1 Gigabit per second (Campbell, 2000b). Up to 4 independent experiments or sub-nets, or up to 16 stations can be processed simultaneously.

The transition to the Mark 5 system has already started by 2002. The disc-based data system will directly replace the Mark IV tape drives. The storage capacity of the discs will increase to 1000 GB within the next years. The Mark 5 system will also support electronic transmission of VLBI data (e-VLBI); it is hence an important prerequisite for future near real-time processing of VLBI data.

11.1.2 Applications

As it has been stated above, the main contributions of VLBI to space geodesy are
- to establish and maintain the *International Celestial Reference Frame* (ICRF),
- to establish and maintain the *International Terrestrial Reference Frame* (ITRF), and
- to establish and maintain the time dependent *Earth Orientation Parameters* (EOP) that relate the ITRF to the ICRF.

VLBI is unique in that it is the only technique for establishing and maintaining the ICRF, and the relationship between the ITRF and the ICRF, by directly monitoring the nutation parameters and UT1. As well as this, it is the only geodetic space technique that contributes to all three of the above mentioned items. Other advantages, when compared with satellite techniques, come from the fact that VLBI is independent of the gravity field. As a consequence (Drewes, 2000):
- VLBI is not affected by satellite orbit errors caused by gravity field mismodeling,
- VLBI is not influenced by variations of the geocenter, and
- VLBI is independent of the uncertainty of the GM value and hence of the related scale problems.

Compared with satellite laser ranging [8], VLBI has the advantage of being weather independent. Disadvantages can be stated as follows:
- VLBI is a rather expensive technology, hence only a limited number of telescopes is available,
- instrumental errors, like telescope deformation, are difficult to handle, and
- results are not yet available in real-time.

VLBI also does not provide absolute coordinates with respect to the geocenter, but baselines between stations or relative coordinates with respect to some arbitrarily selected origin.

Due to the high efficiency of modern satellite techniques like GPS, the VLBI technology is not used for operational positioning in geodesy and geodynamics. VLBI, due to its unique capacities, will however remain the primary geodetic technique for maintaining the fundamental reference frames and their inter-relationship.

Inertial Reference Frame and Source Positions
Under the assumption of a non-rotating universe (Walter, Sovers, 2000) only extra-galactic objects (quasars), which do not participate in the rotation of the galaxy, can provide stable fiducial points for the establishment of an inertial reference frame [2.1.2.1].

VLBI is at present the only astrometric technique for the determination of directions to quasars with good accuracy.

The ICRF (see Fig. 2.3, p. 15) is realized by the coordinates of 608 extragalactic radio sources, positioned with an accuracy between 0.4 mas and 1 mas. They are divided into three categories following quality criteria (Capitaine, 2002):

212 defining sources (best observed set of sources),

294 candidate sources, and

102 other sources.

The stability of the axes, derived from the defining sources, is estimated as 20 microarcseconds.

The coordinates are based on the analysis of about 1.6 million observations between 1979 and 1995 using 24-hour VLBI sessions. The observations continue with the objective to improve the source position accuracy to 0.25 mas for as many sources as possible, and to improve the overall sky distribution of sources (Schuh et al., 2002). In particular, the absence of proper motions or structural changes of the sources at the sub-milliarcsecond level is essential (Walter, Sovers, 2000). For details about the definition and realization of the ICRS see Ma et al. (1997), as well as the already cited literature.

Baselines, Plate Motion, and Terrestrial Reference Frame

The possibility of determining precise baselines with radio telescopes over very large distances was recognized at an early stage, and has been applied within the scope of the NASA *Crustal Dynamics Program*, and its predecessors, since about 1972. Routine observations over several baselines, in particular in North America, started in about 1980 (Anderle, Malyevac, 1983). A major early problem encountered by the technique was that the great majority of radio telescopes was primarily used in astronomical research, so that only very little and sporadic observation time could be assigned to geodetic applications. For this reason a number of dedicated VLBI telescopes for geodetic-geodynamic programs has been installed in subsequent years. The 20-m radio telescope at Wettzell (Fig. 11.3) belongs to this group. It has been incorporated since 1983 into numerous international projects.

Figure 11.3. 20-m VLBI telescope, Wettzell; courtesy BKG, Frankfurt

During the NASA Crustal Dynamics Program (CDP, 1979–1991) vector baselines between selected sites were measured repeatedly. In 1990 the global VLBI-network consisted of about 20

fixed stations participating regularly in geodetic VLBI projects, and about 40 platforms for mobile stations with at least one occupation per year. A VLBI experiment usually lasted 24 hours, and included observations of 12 to 18 different radio sources (Campbell et al., 1992). The CDP was phased out by 1991, and substituted by the DOSE (Dynamics of the Solid Earth) program, including the VLBI activities. The CDP provided the first contemporary measurements of relative motions between Earth's tectonic plates, and it demonstrated the internal rigidity of the continental and oceanic plates.

Figure 11.4. Transportable 6-m VLBI telescope (TIGO), cf. [12.5.2]; courtesy BKG, Frankfurt

Since station coordinates form part of the solution in most observing programs, station velocities for as many as 60 sites are today available from historical data. Currently, about 30 stations are observing within the framework of the *International VLBI Service*, IVS [11.1.3], cf. Fig. 11.6.

In most cases the baseline length changes detected by VLBI confirm the tectonic models to a surprisingly high degree (Schuh, 2000). Fig. 11.5 shows the evolution of the baseline length between Wettzell (Germany) and Westford (USA) over 17 years. The 6000 km baseline shows an increase of 1.70 cm ±0.01 cm/year. The predictions of
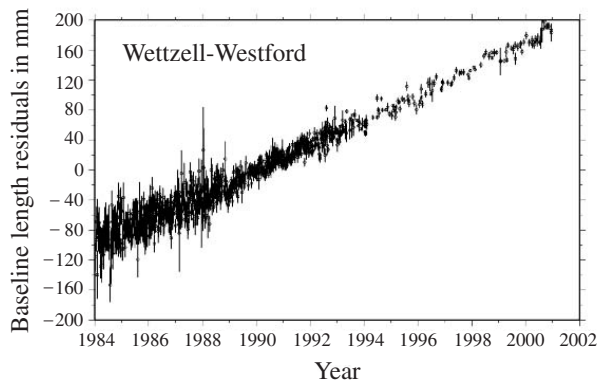
Figure 11.5. Evolution of the 6000 km VLBI baseline Wettzell-Westford; source: Geodetic Institute, University Bonn

the NUVEL geotectonic model, for the same baseline, is 1.9 cm/year. The increasing precision with time is clearly visible. In future, it should be possible to discern non-linear effects in the baseline evolution, such as periodic or episodic changes, if they occur.

One problem in the determination of global plate motions with VLBI is the fact that some of the VLBI stations are placed in deformation zones between the stable plates (Campbell, Nothnagel, 1998). The stability of reference stations has to be carefully established with local surveys, for example using dense GPS control networks.

Current accuracies for coordinates are 5–20 mm from 24 hours of observation and 1–4 mm for annual solutions. The accuracies of coordinate velocities from the annual solutions are estimated to be 0.1 to 1 mm per year (Schuh et al., 2002).

The Terrestrial Reference Frame (TRF) may be determined from single session solutions or from a global solution. Station positions and velocities are usually computed from the same 24 hour sessions that are used for the estimation of Earth Orientation Parameters. In most cases, a global VLBI solution including all available sessions is introduced into the formation of the ITRF by the responsible IERS product center (see [12.4.2]). For the creation of the ITRF2000, for example, 3 global VLBI solutions spanning 20 years of data have been included (IERS, 2001). For a detailed description of the analysis procedure, see the related ITRF documents.

A particular strength of VLBI is its contribution to the scale of the ITRF (Schuh et al., 2002). Current disadvantages are the unequal global distribution of stations, especially the lack of stations in the southern hemisphere. For the future it would be desirable, and sufficient, to have two or three stations in stable areas on each of the major tectonic plates (Campbell, Nothnagel, 1998).

Earth Orientation Parameters

Traditionally, the Earth orientation parameters (EOP) are understood to include polar motion and Earth rotation (see [2.1.2.3]). In modern discussion, as in the documents of the IERS, the EOP group is understood to include also the conventional precession-nutation model and the differences between this model and observations, named *celestial pole offsets*. The EOP, hence, include all parameters describing the transformation between the CRF and the TRF. These are:

x_p, y_p pole coordinates of the Celestial Ephemeris Pole (CEP)
            relative to the IERS Reference Pole (IRP),

UT1   related to the Greenwich mean sidereal time (GMST)
            and expressed as the difference UT1−UTC, and

$d\psi, d\varepsilon$ celestial pole offsets, in longitude and in obliquity, with respect to its
            position defined by the conventional IAU precession/nutation theory.

VLBI is unique in its ability to make rapid, accurate measurements of all five parameters. UT1-UTC and the nutation offsets within the final IERS EOP solution come exclusively from VLBI. Typically, the EOP are observed in several 24 hour sessions per week within regional networks of five or more stations (see [11.1.3]). In addition, so-called intensive sessions of 60 minutes take place on a single long baseline several times per week. The intensive session can only be used to determine UT1−UTC. The

currently achievable accuracy for the pole coordinates is 200 μas for x_p and 100 μas for y_p. The difference is due to the unfavourable geometry of the station networks. UT1 can be determined with an accuracy of 5 μs over 24 hours, and 20 μs over the one hour intensive sessions (Schuh et al., 2002).

An impression of polar motion during recent years is given in Fig. 12.14, p. 533, taken from the IERS Annual Report 2000. The Earth rotation data have revealed various new phenomena, for example the atmospheric excitation of Earth's rotation due to winds in the upper atmosphere or effects like El Niño. A rich spectrum of oscillations can be detected in UT1, including tidally induced phenomena (Schuh, 2000). An even deeper insight into the complete spectrum of oscillations can be expected as soon as continuous VLBI observations (7 days a week) become available.

Based on VLBI observations, new theoretical models could be developed in nutation theory. The new IAU 2000 precession-nutation model is mainly based on VLBI observations. A comparison of the model with VLBI observations shows an agreement at the order of 200 μas (Capitaine, 2002). With future continuous observations an improvement of a factor of 5 to 10 can be expected (Schuh, 2000).

The effects of precession and nutation are caused by the gravitational forces of the Sun and the Moon on Earth. Earth's response to these forces depends on its structure. This is why precise VLBI data and the analysis of precession/nutation serve as highly sensitive probes of Earth's interior structure.

In addition to the above mentioned geodetic products, VLBI contributes information on solid Earth tides, ocean loading, and atmospheric loading as well as tropospheric zenith delay gradients, ionospheric parameters, and the *relativistic light deflection* parameter, γ (Schuh et al., 2002).

Table 11.1 gives a summarizing overview of the currently achievable accuracy of VLBI products.

Table 11.1. Present status of geodetic and astrometric VLBI, cf. Schuh et al. (2002)

Type	Product	Accuracy	Frequency	Resolution	Delay
CRF	α, δ	0.25–3 mas	1 year		3–6 months
TRF	session				
	coordinates	5–20 mm	∼3d/week	1d	3–4 months
	annual				
	coordinates:	1–4 mm	1 y	–	3–6 months
	velocities:	0.1–1 mm/y			
EOP	UT1 from				
	24h session	5 microsec	∼3d/week	1d	1–4 months
	60min. sess.	20 microsec	∼4d/week	1d	1 week
	x_p	200 μas	∼3d/week	1d	1–4 months
	y_p	100 μas	∼3d/week	1d	1–4 months
	dε, dψ	0.1–0.4 mas	∼3d/week	1d	1–4 months

11.1.3   International Cooperation, International VLBI Service (IVS)

VLBI experiments require international cooperation. In particular, the regular determination of Earth orientation parameters is only possible when several stations cooperate following a strict schedule, and when the data flow and the correlation process is well organized. The VLBI part of the NASA Crustal Dynamics Program was organized based on the *experiment by experiment* philosophy. By about 1980 the *monitoring concept* was formulated, with the objective to increase the frequency of observing sessions to a level that allowed continuous monitoring of EOP (Campbell, 2000b).

After the successful participation of several VLBI observatories during the MERIT campaigns in 1980 and 1983/1984 (Robertson, Carter, 1985) [12.4.2], a number of observatories continued with geodetic/astrometric VLBI on a regular basis.

The control network POLARIS (*Pol*ar Motion *A*nalysis by *R*adio *I*nterferometric *S*urveying), originally established in North America, was expanded to include telescopes on other continents in the IRIS (*I*nternational *R*adio *I*nterferometric *S*urveying) network. From 1984 to 1990 IRIS observed routinely at a 5-day interval. Afterwards it was continued under the responsibility of NASA and the US Naval Observatory (USNO) under the acronym NEOS (*N*ational *E*arth *O*rientation *S*ervice). NEOS operation consists of one 24 hour observing session per week for EOP and a daily one hour or 90-minute, intensive, observation on one baseline for UT1. A network of 4 to 6 stations surrounding the Atlantic ocean has been in operation since 1986 as IRIS South (IRIS-S) under the responsibility of the University of Bonn. IRIS-S observes monthly 24 hour sessions for stabilization of EOP and TRF. It is the longest running VLBI series.

In 1997 a program named CORE (*C*ontinuous *O*bservation of the *R*otation of the *E*arth) was proposed by NASA, with the objective to coordinate VLBI observations in sub-networks of about 5 stations, each observing on different days of the week. Two sub-networks (CORE-1 and CORE-2) started in 2000 with monthly sessions (IERS, 2001).

Other examples of cooperation are the *Very Long Baseline Array* (VLBA), a 10 station network operated in North America, and the European geodetic VLBI network. VLBA is mainly devoted to astronomical research, but six observing days per year are allocated for geodesy/astrometry. The European geodetic VLBI network includes up to 10 stations and is mainly devoted to the determination of vertical crustal motion in Europe (Campbell et al., 2002).

Based on the experience of many years' international cooperation, and stimulated by the great success of the "International GPS Service" (IGS), the geodetic VLBI community decided to establish the *International VLBI Service for Geodesy and Astrometry* (IVS) and to continue and coordinate all international activities under this umbrella. The IVS started as a service of the "International Association of Geodesy" (IAG) on July 1, 1999, and has also been recognized as a service of the "International Union of Astronomy" (IAU) since August 2000. It cooperates closely with the "International Earth Rotation Service" (IERS). The organizational structure is similar to that of the IGS. Besides a *Directing Board* which determines policies, adopts standards

and sets the scientific goals for the observing programs, the IVS (status 2002) consists of (Vandenberg et al., 2002):

  31 Network Stations, acquiring high performance VLBI data,

   3 Operation Centers, coordinating the activities of the Network Stations,

   6 Correlators, processing the acquired data and providing processed data for analysis,

   6 Data Centers, distributing products to users, storing and archiving data,

  21 Analysis Centers, analyzing the data and producing the results and products,

   7 Technology Development Centers, developing new VLBI technology, and

   1 Coordinating Center, coordinating daily and long term activities.

Details can be taken from the IVS Annual Reports, e.g. Vandenberg et al. (2002). These publications also contain information on all participating telescopes. A global map with the current distribution of participating VLBI stations is shown in Fig. 11.6. The inhomogeneous distribution is clearly visible. A future goal is to operate 3 to 4 telescopes on each of the major tectonic plates, if possible outside of deformation zones. This objective can be contributed to by transportable VLBI telescopes like the VLBI component of the *Transportable Integrated Geodetic Observatory* (TIGO) [12.5.2], cf. Fig. 11.4.
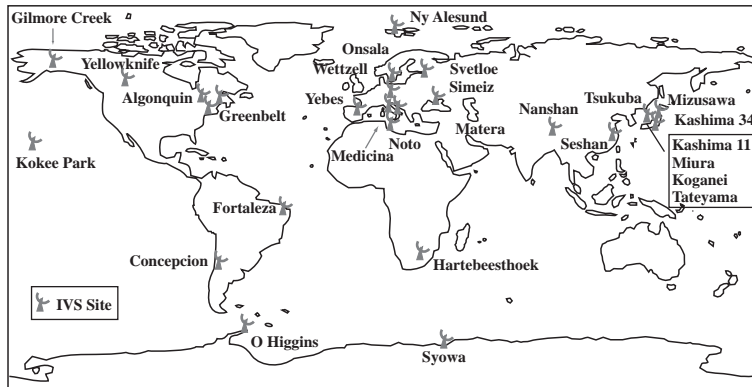
Figure 11.6. Network Stations of the International VLBI Service, Status 2002

The IVS started with a continuation of existing programs such as IRIS-S, NEOS, and CORE, and has defined its own observing programs since 2002. The long term goals are among others, e.g. (Schuh et al., 2002):

  − improve EOP and TRF by a factor of two to four,

  − improve the sky distribution of the CRF,

  − decrease the average time delay for the delivery of products, and

  − increase the frequency of observing sessions up to 7 sessions per week.

Within the IVS working groups, decisions were made to facilitate the exchange of data and results. A platform independent VLBI exchange format (PIVEX) has been developed, and an extension of the SINEX format [7.3.4] for VLBI parameters. SINEX 2.0 permits an easy combination of different VLBI solutions as well as their combination with results from other space techniques (Nothnagel, 2002).

The newest information on the development of geodetic VLBI can be obtained from the reports and meeting proceedings of the IVS. The reports are available from the IVS homepage, currently operated by the NASA GFSC. An excellent overview of current status and prospects is given by Schuh et al. (2002).

11.1.4   VLBI with Satellites

High orbiting satellites, emitting radio frequency signals, can be compared to the radio sources in the VLBI technique. Unlike the latter, however, they are not located at infinity but in the near-Earth space. The wavefronts arriving at two antennas can hence not be considered as straight lines, as in Fig. 11.1. Instead, a different geometrical model is required (Fig. 11.7) (Maniatis et al., 1987). The accuracy of the technique is substantially dependent on the accuracy of the satellite orbits.

As in the VLBI method, the difference in the time of arrival of the satellite signal wave front at both antennas, 1 and 2, is used as the observable. The difference of the arrival times can either be formulated as a time measurement,

$$S_{1,2}(t) = \tau(t)c, \qquad (11.11)$$

or in the form of phase differences, $\Phi_{1,2}(t)$:

Figure 11.7. VLBI with satellites

$$S_{1,2}(t) = \frac{1}{2\pi}\Phi_{1,2}(t)\lambda + N\lambda \qquad (11.12)$$

with an unknown number, N, of full wavelengths, λ.

Some early proposals included the idea of using the GPS satellite signals for geodetic purposes in a form analogous to the VLBI technique. Counselman, Shapiro (1979) proposed installing additional transmitters on the GPS satellites and using the signal for interferometric techniques with fairly simple ground equipment. The project was called *Miniature Interferometer Terminal for Earth Surveying* (MITES), but it was never built in the proposed way. Instead, an alternative instrument was developed, the *Macrometer*, based on the fundamental observational principle of MITES, but without the necessity for additional equipment on the GPS satellite [7.2.4.1]. Basically, it refers to the solution of equation (11.12). The same equation is applied for the use of carrier phase measurements in the GPS technique [7.3.1]. Many of today's algorithms for GPS carrier phase measurements can be traced back to this development.

The basic principle of equation (11.11), namely the direct observation of the propagation time delay through correlation of GPS signals received at two antennas, was proposed by Mac Doran (1979). In this technique, which was called *Satellite Emission Radio Interferometric Earth Surveying* (SERIES), the GPS signals are considered to be pure noise. The realization of this proposal requires rather bulky instruments; this is the reason why the proposal was not successful in GPS technology. A modified version uses the binary structure of the GPS time signals (Mac Doran, 1983), but at a rather low accuracy level. Some of these early ideas for the codeless use of GPS were taken up again in modern GPS receiver technology, to gain access to the full carrier signal of L2 under activated AS [7.2.3].

One major advantage of satellite signals, when compared with radio sources is the much stronger signal, and the existence of predetermined structures. Hence, much simpler concepts can be used for the receiver design and the data processing techniques. The proposals, originally developed for GPS, may be used again for future satellite systems. If signals from GPS satellites and from radio stars are observed with the same radio telescopes, the GPS observations could be linked directly to the CIS reference frame [2.1.2.1], [11.1.2].

Proposals were made early on to launch radio telescopes into Earth orbit and to integrate them into the ground based VLBI networks (Fejes, Mihály, 1991; Fejes, 1994). This *Space-VLBI* offers a wide range of applications in the field of geodynamics and satellite dynamics. One potential application is related to the connection and unification of reference frames. The TRF can be tied directly to the CRF, because the space antenna is related by interferometric baselines to the TRF, realized by the network of ground-based antennas. Another aspect is the inclusion of directional information into the space data for gravity field determination (Ádám, 1999). Because of the lengthened baseline, a detailed study of the structure of sources used in the establishment of the CRF, will be possible. Particular difficulties are that precise orbit information is required and that the correction for ionospheric delay may be difficult if no dual frequency observations from the space-telescope are available

The feasibility of Space-VLBI was at first demonstrated experimentally using the 5 m diameter antenna of a geosynchronous TDRS satellite [4.3.2]. A first dedicated Space-VLBI satellite, named *HALCA*, was launched successfully on February 12, 1997, from Japan into an elliptic orbit. HALCA, with its 8 m diameter antenna, forms the orbiting element of the international *VLBI Space Observatory Program* (VSOP). It operates at 1.6 GHz and 5 GHz. Together with ground-based telescopes, HALCA creates an effective telescope diameter up to 30 000 km.

The Russian RADIOASTRON satellite has been approved, but the launch is delayed until the second half of this decade due to funding problems. The satellite carries a 10 m antenna, and will be launched into an elliptical orbit. An apogee radius in the range of up to 350 000 km is under discussion. This enormous baseline will allow the study of quasar structure with unprecedented angular resolution.

Space-VLBI projects are primarily dedicated to astrophysical research, but will also contribute to geodesy and geodynamics. First experiments with data from HALCA gave encouraging results (Meyer et al., 2000).

11.2   Interferometric Synthetic Aperture Radar (InSAR)

Satellite borne radar techniques do not belong to satellite geodesy but are usually treated within the field of *remote sensing*. Still, some fundamentals and possible applications are addressed here because of their close relationship to satellite geodesy and other geodetic techniques. Most spacecraft carrying SAR equipment (like ERS-1/2, ENVISAT) are also used in satellite altimetry [9.2], and some of the results from differential InSAR can be considered as complementary to GPS in geodynamic deformation studies. Further connections are the determination of satellite orbits, the signal propagation, and the use of GPS for spacecraft and ground control. InSAR is a rather complicated and very demanding discipline, requiring a comprehensive treatment of its own. Within this text only a very rough and basic idea is presented. For further studies the reader is referred to the special literature, e.g. Leberl (1990); Bamler (1998); Gens (1998); Lillesand, Kiefer (2000).

11.2.1   Basic Concepts, Synthetic Aperture Radar (SAR)

RADAR stands for *Radio Detecting and Ranging*. A radar instrument illuminates an area with microwaves and measures the travel-time and strength of the returned signal. From these the range between the reflecting object and the radar antenna can be determined. Typical wavelengths are:

  3 cm   X-band,

  6 cm   C-band, and

 24 cm   L-band.

One problem encountered by radar techniques is the low resolution of microwaves. In general, this is determined by the frequency, the range of the object and the size of the aperture (see also equation (11.1)). For satellites at a height of about 800 km a several hundred meter long antenna would be required to achieve the aperture necessary for a resolution of 100 m on the ground. This is technically not feasible. However, when the radar measurements are taken from a moving platform (satellite or airplane) then the reflected signals along the flight path can be collected and combined. The aperture is hence created synthetically during the signal processing. This technique is called *Synthetic Aperture Radar* (SAR). As a consequence, the radar achieves a high resolution in the along-track direction (also called the azimuth direction). In the range direction, perpendicular to the flight path, the resolution is determined by the duration of the transmitted pulse. In practice, frequency modulated pulses are transmitted, and the phase of the return signal is measured.

Fig. 11.8 demonstrates the basic principle (Bamler, 1997). A SAR transmitting antenna illuminates the Earth's surface in a side-looking mode. The return signals are recorded with respect to intensity (magnitude) and phase. Phase means a relative shift of the received sine signal with respect to the transmitted signal. The resolution in range and azimuth defines the smallest picture element (pixel). For ERS the pixel size is

about 12.5 m × 12.5 m. Fig. 11.8 also explains that the SAR process transforms a 3-D object, e.g. the topography, into a two-dimensional radar image with the coordinates *range* and *azimuth*. This type of geom-
etry creates distortions and makes inter-
pretation of SAR images difficult, in par-
ticular over mountainous terrain.

Figure 11.8. Basic principle of SAR

Compared with traditional optical methods, SAR images have particu-
lar properties and advantages (Bamler, 2000):

— microwave radiation is not af-
  fected by clouds or heavy rain and
  can produce images in all weather
  conditions; it also can penetrate
  (partially) snow and soil,
— SAR is an active source technique
  and hence is independent of sun-
  light; continuous day and night operation is possible,
— SAR can be considered as a technique complementary to optical remote sensing;
  different properties of the same objects can be recorded,
— SAR is a coherent imaging method; this makes the interferometric approach
  possible, and
— SAR is capable of observing dynamic processes such as ocean currents or sea
  ice motion.

On the other hand SAR has some disadvantages. The resolution is rather low, and the received signals are heavily affected by a noise-like phenomenon termed *speckle*, caused by a large number of scatterers in the image formation process.

A SAR image contains geometric and radiometric information. Each pixel's bright-
ness is determined by the back-scattered radiation from a surface element on the ground. A strong signal results in a bright pixel value. The signal strength depends on many influences, such as topography, size of scatterers, radar wavelength, surface humidity, and incident angle. The interpretation of a SAR image hence requires a profound understanding of the image formation process, see e.g. Bamler (1998).

The use of spaceborne imaging radar in remote sensing started with the launch of SEASAT-1 [9.2] in 1978. In spite of the short lifetime of this satellite the capability of SAR for mapping the Earth's surface was demonstrated. The first complete maps of some entire countries, in particular those in tropical areas covered by clouds, were generated in 1992 using ERS-1. A particular use of SAR in topographic mapping was NASA's Magellan mission to planet Venus, that is permanently covered by dense clouds. SAR maps of the Venusian surface were generated in 1990.

Another important field of SAR uses, besides topographic mapping, is environ-
mental monitoring. Examples are
— land use,
— erosion, deforestation, desertification,
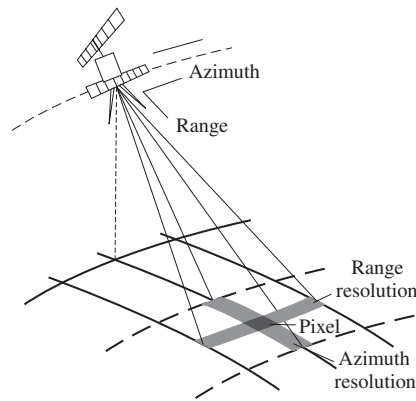
  − sea state, flood monitoring, ice flow,

  − sedimentation and oil-slick detection.

Other particular applications are in reconnaissance for military purposes, and the identification of geological and archeological structures.

  A rather large number of spacecraft have been launched carrying SAR equipment. Table 11.2 gives an overview (see Hartl et al., 1995). SIR stands for *Shuttle Imaging Radar*. SIR-A and SIR-B were operated for one day each on separate missions. SIR-C was flown on several space shuttle missions and was also used for 11 days along with the SRTM mission in 2000, see [11.2.2] .

Table 11.2. Satellite SAR systems

SAR-System	year	country	band	wavelength [cm]	frequency [GHz]
SEASAT-1	1978	USA	L	23,5	1.3
SIR-A	1981	USA	L	23,5	1.3
SIR-B	1984	USA	L	23.5	1.3
ERS-1	1991	ESA	C	5.7	5.3
J-ERS-1	1992	Japan	L	23.5	1.3
SIR-C	since	USA	L, C,	24.0; 5.7	1.3; 5.3
	1993	Germany	X	3.1	9.6
ERS-2	1995	ESA	C	5.7	5.3
RADARSAT	1995	Canada	C	5.7	5.3
ENVISAT-1	2002	ESA	C	5.7	5.3

11.2.2   Interferometric SAR

The usual SAR image is an "intensity-image" whose pixels characterize the brightness of radar echos (Hartl et al., 1995). The SAR image can also be treated as a "phase-image", based on the phase difference between the transmitted and received microwave signal. Pixels can be characterized as brightest for phase values of 2π and black for 0. Due to the small wavelength of the microwave signal compared with the size of the backscattering element (pixel size about 10 m × 10 m or larger) the phase-image will show an almost random character and is of no practical use.

  The situation changes, however, if a second image of the same ground scene is recorded at a slightly different antenna position (Fig. 11.9). This is the interferometry principle. The phase difference,
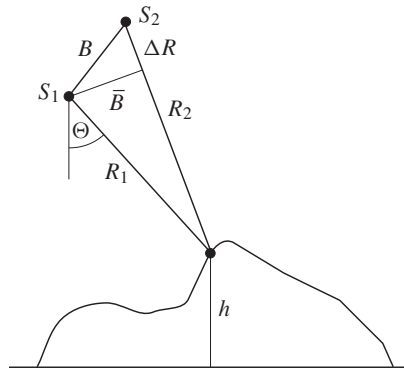
Figure 11.9. InSAR geometry

$\Delta\phi$, corresponds to the difference in range, $R_2 - R_1$. The basic idea is as follows (Bamler, 1997).

The phase of a pixel, ϕ, includes information on the phase shift caused by scattering influences, ϕ_s, and on the phase delay caused by signal propagation from the sensor antenna to the ground element and back, ϕ_p:

$$\phi = \phi_s + \phi_p. \tag{11.13}$$

The portion due to signal propagation is related to the range, R, by

$$\phi_p = \frac{4\pi}{\lambda} R, \tag{11.14}$$

with λ the radar wavelength. The phase, ϕ, can only be measured modulo 2π.

In the phase difference, or *interferometric phase*, the scattering part, ϕ_s, will vanish from (11.15) if the scattering mechanism did not change between both images, hence $\phi_{s_1} = \phi_{s_2}$. The interferometric phase is then

$$\Delta\phi = \phi_2 - \phi_1 = \phi_{p_1} - \phi_{p_2}. \tag{11.15}$$

SAR interferometry (InSAR) hence offers the possibility to use the stereo effect and to measure stereo parallaxes to the fraction of a wavelength. Fig. 11.9 demonstrates that the range parallax, $R_2 - R_1$, is related to the look angle, θ, and is hence also a measure for the terrain height, h. SAR interferometry is thus obviously a powerful method to map topography and to generate *Digital Elevation Models* (DEM).

The two receiving antennas, required for interferometric SAR, can be arranged in a different manner. Along-track antenna arrangements are used for motion measurements. The configuration in Fig. 11.9 is particularly suitable for DEM generation and is called an *across-track interferometer* (Bamler, 1997). Two options are in use. These are the

- single-pass interferometer, and the
- repeat-pass interferometer.

With the *single-pass interferometer* two SAR images are recorded simultaneously. The configuration requires one transmit/receive antenna and a second receive antenna at some distance. This is technically difficult and could only recently be realized with the *Shuttle Radar Topography Mission* (SRTM). The second antenna was located at the end of a 60 m boom, at a very precisely known distance.

In the *repeat-pass interferometry* a second image of the same area is taken during a different pass by the same or another satellite after several days or even months. This technique has been widely used to generate elevation maps since the start of ERS-1. During the tandem mission, ERS-1/ERS-2, the second satellite flew over the same area one day after the first with an orbit separation of a few hundred meters.

The InSAR technique only works properly when the scattering behavior of the images does not change too much between both observations and when the baseline between both antenna positions is not too long. Otherwise the images decorrelate

and make it impossible to derive phase differences. The maximum baseline for ERS InSAR was found to be about 1000 m. On the other hand the baseline should not be too short because of the height-to-phase sensitivity of a repeat-pass SAR interferometer (Bamler, 1997):

$$\frac{\partial \Delta \phi}{\partial h} = \frac{4\pi \overline{B}}{\lambda R \sin \theta}. \tag{11.16}$$

Equation (11.16) also gives the relationship between a full phase cycle, 2π, and the difference in height, Δh, as

$$\Delta h = \frac{\lambda}{2} \frac{R}{B} \tan \theta. \tag{11.17}$$

One full cycle, 2π, in the relative phase creates one fringe in an interferogram. For ERS-1/2, with an incident angle, $\theta = 23°$, we find for a baseline of 100 m a fringe size of about 100 m. With a resolution of 10% for the relative phase this corresponds to a height resolution of about 10 m.

The phases can only be determined modulo 2π, i.e. the interferometric phase is ambiguous. The process of resolving the ambiguities (compare with [7.3.2.3] for GPS phases) is called *phase unwrapping*. For details see e.g. Bamler (1998).

The accuracy of the final product, in general a DEM, mainly depends on the following influences (Bamler, 1997):

   − accuracy of the phase measurement (phase noise),
   − accuracy of the SAR geometry (orbit errors), and
   − modeling of the wave propagation (atmospheric effects).

Phase noise has instrumental causes (receiver noise), and is also created by temporal decorrelation in the repeat-pass interferometry, for example caused by a change in vegetation. For single-pass interferometry, only the instrumental causes are of relevance. *Orbit errors* propagate directly into the geometry of the derived DEM. The same is true for errors in the range between the two SAR antennas. Orbit determination is being continuously improved by spaceborne GPS, DORIS, and SLR [3.3.2.3].

Propagation effects are caused by the troposphere, ionosphere, and local cells of water vapor. Most effects can be averaged out by using several interferograms of the same area. Ionospheric influences can be modeled for dual frequency SAR, as with SIR-C/X. IGS global ionospheric and tropospheric models [7.8.1] may help to improve SAR image processing. For single-pass interferometry the effects will be mostly identical for both antennas and hence will not contribute to decorrelation.

Single-pass interferometry obviously has many advantages over repeat-pass interferometry. The first space-based mission of this type was the *Shuttle Radar Topography Mission* (SRTM), flown on Space Shuttle STS-99 for 11 days from February 11 to February 22, 2000. The orbital height was 230 km and the inclination angle 57°. Due to the side-looking geometry the mission was able to acquire topographic data between 60°N and 56°S. The horizontal resolution is about 30 m × 30 m; the projected vertical accuracy is 16 m. First analysis of the data indicate this accuracy was exceeded. The total data set will provide a digital elevation model based on WGS84 for a large part of the Earth with unprecedented accuracy, homogeneity and completeness. For details see Bamler (1999).

11.2.3   Differential Radar Interferometry

When three or more SAR images of the same area are generated from different passes at approximately the same antenna position it is possible to derive at least two interferograms. These can be differenced and used to produce a *differential interferogram* or "double-difference interferogram". The differential interferogram shows phase changes only where surface changes occured between the times of observation and hence have caused a change in the slant range to the antenna. One complete phase cycle corresponds to about 3 cm for ERS-1. Hence the method reveals surface changes like swelling and buckling with a resolution of centimeters or even millimeters.

The potential of differential radar interferometry was recognized early in 1989 (Gabriel et al., 1989), and was applied to the determination of surface deformation in volcanic or seismically active areas. Famous examples are the *Landers Earthquake* in Southern California (Massonet, et al., 1993), and the deflation of *Mount Etna* (Massonet, 1997). Applications are seen for

- earthquake and volcanic research,
- research into tectonic processes and crustal deformation,
- glaciology, ice sheet monitoring, and
- monitoring land sliding and subsidence.

Differential InSAR can be effectively combined with other geodetic techniques, like continuous GPS arrays [7.6.2.2], to provide a highly detailed and accurate picture of crustal deformation. An early example is the Landers Earthquake, where several GPS stations monitored continuously during the event (Bock, Williams, 1997). One main advantage of GPS is the highly accurate (few mm) continuous monitoring of absolute 3-D displacements over large areas. One main disadvantage is that observations are taken at an irregularly spaced set of stations, not all of which are optimally placed with respect to the earthquake displacement zone. The main advantage of InSAR is its much better spatial coverage. On the other hand InSAR interferograms are restricted to smaller regions and the temporal resolution is limited. Images can be easily decorrelated by changing conditions in vegetation and humidity. A short review of some strengths and weaknesses is given in Table 11.3.

Table 11.3. Comparison of GPS and D-InSAR, after Bock, Williams (1997)

	Continuous GPS	Differential InSAR
Strengths	high temporal density 3-D positioning mm accuracy	high spatial density remotely sensed no monumentation necessary
Weaknesses	limited spatial density stable monumentation siting difficulties	limited temporal density 1-D scalar measurement image decorrelation

12 Overview and Applications

The possible applications of individual satellite techniques were discussed in chapters [5] to [11] where the related observation equations can also be found. In this final chapter, the solution concepts for a given group of tasks and problems are summarized and reviewed. The historical background is indicated. Certain common views are emphasized; some aspects are supplemented and treated in more depth. Hints are given about relevant literature.

12.1 Positioning

12.1.1 Concepts, Absolute and Relative Positioning

The relationship between the satellite position, S, the observation station, B, and Earth's center of mass, M, is given through the fundamental equation (cf. (4.1), Fig. 12.1 (a)

$$r_S = r_B + \rho. \tag{12.1}$$

If only the geometric relationship between the ground station, B, and the satellite, S,
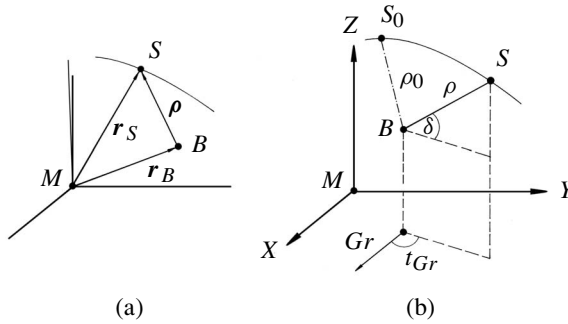
Figure 12.1. Positioning from satellite observations

is considered, we find for the components of the interconnection vector ρ, Fig. 12.1 (b) that

$$\rho = \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix} = \begin{pmatrix} X_S - X_B \\ Y_S - Y_B \\ Z_S - Z_B \end{pmatrix} = \begin{pmatrix} \rho \cos \delta \cos t_{Gr} \\ \rho \cos \delta \sin t_{Gr} \\ \rho \sin t_{Gr} \end{pmatrix}, \tag{12.2}$$

where
  δ declination in the equatorial system, and
  t_{Gr} Greenwich hour angle.

Observables may be
  − directions (cameras, CCD) [5.1], [5.2],

- ranges (laser) [8],
- pseudoranges (GPS) [7], and
- range differences (Doppler) [6].

The associated original, non-linearized, observation equations are
for *directions:*

$$\tan t = \Delta Y / \Delta X, \quad \tan \delta = \frac{\Delta Z}{\sqrt{\Delta X^2 + \Delta Y^2}}, \tag{12.3}$$

for *ranges:*

$$\rho = \sqrt{\Delta X^2 + \Delta Y^2 + \Delta Z^2}, \tag{12.4}$$

and for *range differences:*

$$\rho - \rho_0 = \sqrt{\Delta X^2 + \Delta Y^2 + \Delta Z^2} - \sqrt{\Delta X_0^2 + \Delta Y_0^2 + \Delta Z_0^2}. \tag{12.5}$$

Approximate values of the station and satellite coordinates are used for the linearization of the observation equations. The associated differentials can be found in published sources, e.g. (Campbell et al., 1973).

The positioning problem can be treated both dynamically and geometrically [4.1]. In the pure *geometric approach* the satellites are regarded as high orbiting targets. The important requirement is that the same satellite is observed simultaneously from as many ground stations as possible. The observations (12.3) to (12.5) can be used to establish regional and global spatial satellite triangulation and trilateration networks. The fundamental work on this problem started very early, e.g. Wolf (1967); Rinner et al. (1969); Moritz (1970).

In a geometrical adjustment, only *relative coordinates* are determined. There remains a so-called *datum defect* [2.1.5]. The scale, orientation, and origin of the reference frame have to be fixed independently by other means if they are not provided by the observation method itself. The scale of the network can be fixed by range observations between some of the ground stations, or between ground stations and satellites, e.g. with satellite laser ranging. The orientation of the network can be derived from the stellar CIS [2.1.2.1] along with camera observations, [5.1] or with VLBI [11.1]. The origin can be fixed, if coordinates are introduced for at least one of the network stations.

The purely geometric determination of relative coordinates played an important role in the early years of satellite geodesy, and was used in the establishment of several global and regional satellite networks. One eminent example is the *BC4 Global Network* with 45 globally distributed stations [5.1.5] and a mean coordinate accuracy of ±4...5 m (cf. Fig. 5.10).

In the *dynamical concept* the satellite orbit is computed in the acting force field from initial conditions and boundary values [3.3], and is used in the solution of the positioning problem. The satellites can be considered as carrying their own coordinates with them. They form a network of control points in space, and the terrestrial stations are tied to this network through the observation elements (12.3) to (12.5).

The satellite coordinates, and also the derived station coordinates are *by nature* geocentric, because the satellite motion is referred to the gravitational center of the central body (geocenter). No datum problem exists in the dynamical concept. *Absolute coordinates* are determined in the same reference frame as that in which the satellite orbit is computed.

Most of today's operational observation techniques in satellite geodesy deliver absolute coordinates in the related *satellite datum*, for example GPS [7] in the World Geodetic System WGS 84. One important feature of absolute methods in satellite geodesy is that coordinates can be determined from observations at one station only. The methods are therefore appropriate for navigational purposes.

Some of the characteristic elements of absolute and relative positioning are summarized in Table 12.1. For the absolute methods the achievable accuracy of the coordinates is directly dependent on the accuracy of the available orbit data [3.3]. This is of particular interest for the operational satellite navigation systems. The broadcast orbits mostly suffer from difficult to model surface forces like drag [3.2.3.3] or solar radiation pressure [3.2.3.4], and from the inhomogeneous distribution of tracking stations. The GPS orbits, for many years, were corrupted by intentional accuracy limitations [7.1.6]. As a consequence, the absolute accuracy of the position determination is usually far inferior to the relative accuracy. For GPS the difference may reach as much as three orders of magnitude (1 cm against 10 to 20 m).

Table 12.1. Characteristic aspects of dynamical and geometrical methods in satellite positioning

Dynamical methods	Geometrical methods
datum provided by satellite orbits	datum undetermined (datum defect)
absolute coordinates	relative coordinates
point positioning concept possible (single station)	simultaneous multi-station concept necessary
absolute coordinate accuracy (e.g. GPS) 5–15 m	relative coordinate accuracy (e.g. GPS) 1 cm

Depending on the task, absolute as well as relative information is extracted from the observables for the solution of practical problems. This means that the necessary datum information comes partly from the satellite orbits and partly from the terrestrial networks or from other sources. With this background in mind the question of datum transformation and the combination of satellite and terrestrial networks may be assessed. The following levels in the use of datum information can be distinguished.

(1) *The complete datum is taken from the satellite orbit*
This is, for example, the case in navigation with a single receiver, or in point positioning with GPS, GLONASS or the future GALILEO. The related accuracy is about $\pm 5 \ldots 15$ m for GPS without SA. (With activated SA it only was $\pm 30 \ldots 100$ m). The

position results refer to the datum of the satellite system, for example WGS 84 [2.1.6]. The coordinates, obtained from the observations, can be regarded as a *realization of the satellite datum* for the given epoch at the given place with a given uncertainty. The transformation into a particular local reference system can be done with generally accepted datum-transformation parameters [2.1.5]. However, the accuracy of the transformation is no better than the related accuracy of the actual datum-realization, i.e. $\pm 3 \ldots 15$ m for GPS.

If several stations are operating simultaneously, or if they are interconnected through a network in the working area, all observations can be used and provide a mean realization of the datum during the observation period. The accuracy of the realization corresponds approximately to the realization accuracy of observations at a single station. However, the simultaneously observed stations show a high relative accuracy caused by the high correlations between the simultaneous observations. The results hence contain absolute *and* relative information.

(2) *Only one part of the datum is taken from the satellite orbit*

(a) *Orientation and scale*. This may be the case if two satellite receivers are operated simultaneously (e.g. DGPS or RTK [7.5]), and the resulting baseline vectors are used to connect new points to existing control points of the terrestrial network (e.g. Fig. 7.71 in [7.6.1.3]). The position information of the datum here comes from the terrestrial network.

(b) *Scale only*. This is the case if, for example, GPS observations with two receivers are exclusively used as a ranging method for terrestrial trilateration. This mode was frequently used during the first years of GPS application in surveying.

(c) *Orientation, or one part of the orientation, only*. This depends on the kind of parameter selection in the transformation formulas between the satellite network and the terrestrial network [2.1.5].

(3) *No part of the datum is taken from the satellite orbit*

This is the case if the satellite network is first computed from only the simultaneous satellite observations (case (1)), and is then transformed with a 7-parameter transformation via identical points onto a terrestrial network. In this case the complete datum is taken from the terrestrial network; the information from the satellite datum is eliminated in the transformation process. The satellite observations are only used to determine the geometry of the network. This is also the case for networked GPS reference stations [7.5.3.2].

(4) *The satellite orbit is partly or completely recomputed*

(a) *The satellite orbit is recomputed over the observation area.*

(b) *The satellite orbit is provided by an agency like IGS.*

The procedure (a) is also referred to as the "short arc technique" or semi short arc technique" [3.3.3.3], [4.1]. The satellite orbit is recomputed over the observation area, or is given some degrees of freedom, based on the current observations. In these cases part of the datum comes from the broadcast orbit, part stems from the coordinates of the existing network, and part is taken from the orbit adjustment model with the

input gravity field. Geometric as well as dynamic elements enter the solution. The datum information in the broadcast satellite orbit is improved using the datum information inherent in existing coordinates of a selected number of control points. This last-mentioned aspect is, for example, of importance for high-precision application of GPS in geodynamic research, if reference observations are available from fundamental stations with precisely known geocentric coordinates. This is the case for ITRF and/or IGS stations. The procedure is also called the *fiducial point concept*. Fiducial stations can be introduced either as error-less, or with appropriate variances. A proper weighting of all observations entering the network adjustment is essential.

The IGS orbits for GPS satellites [7.4.3.2] (procedure b) are completely recomputed and refer to ITRF2000 which is essentially identical to WGS 84 [2.1.6]. For observations at a single station (precise point positioning [7.3.4]) the datum is fully derived from the precise post-processed ephemerides. When IGS orbits are used together with observation data from IGS stations as fiducial points, both data sets are fully compatible, and new stations are interpolated into the datum of the IGS stations which is practically identical to ITRF2000.

12.1.2 Global and Regional Networks

One of the fundamental objectives in satellite geodesy, identified from the beginning, is to determine precise geocentric coordinates for globally distributed control stations within a uniform reference frame (cf. [1.1], [1.2]). One early contribution to this fundamental task was made with the purely geometric BC4 world network [5.1.5]. Another geometric solution was provided with the equatorial SECOR network [4.4.1]. With the inclusion of TRANSIT Doppler observations both sets of coordinates were made nearly geocentric.

Most of the early global sets of coordinates for a given number of observation stations formed part of comprehensive dynamical *solutions* where gravity field parameters were adjusted along with the station coordinates. These so-called *Earth models* are discussed in chapter [12.2]. The Smithsonian Astrophysical Observatory *Standard Earths* (SAO-SE) and the *Goddard Earth Models* (GEM) belong to this category. A comprehensive review and discussion of several of the early solutions is given by Mueller (1975), and is also included in the report on the U.S. "National Geodetic Satellite Program" (Henriksen, 1977).

With the worldwide use of TRANSIT and GPS, an increasingly dense field of geocentric station coordinates has been built up. If broadcast ephemerides have been used, the absolute coordinates refer to WGS 84 (or to WGS 72 for TRANSIT before 1988 and for GPS before 1987). The accuracy of the station coordinates was quite different. Where precise ephemerides were used for the reduction of TRANSIT Doppler measurements, the absolute accuracy of the single station coordinates was about $\pm$ 1 to 2 m, depending on the geographical location [6.2.2]. The total number of Doppler stations, determined with precise ephemerides, may reach several thousand. This global set of precise Doppler coordinates was one of the best sources for geocentric station coordinates until about 1990.

Sets of global station coordinates can be supplemented and densified by *regional campaigns* and solutions. Regional Doppler campaigns, with the inclusion of a large number of simultaneously operating TRANSIT receivers were very successful. Examples are the *European Doppler Campaign*, EDOC (Boucher et al., 1979), the *German-Austrian Doppler Campaign*, DÖDOC (Rinner et al., 1982), and the *African Doppler Survey*, ADOS (Chodota, 1987). These projects provided sets of absolute coordinates for the working areas with the related system accuracy, as well as sets of relative coordinates for all participating stations with a much higher accuracy.

The focus of interest since about 1985 has been the installation and monitoring of a very precise set of station coordinates for the establishment of a *Conventional Terrestrial Reference System* (COTES), based on all available geodetic space methods.

In the course of the MERIT campaign (1983–1984) [12.4.2] several solutions, mainly based on SLR and VLBI were computed for 15 to 35 stations and compared with one another. The absolute coordinates, at that time, showed differences of up to 10 to 20 cm between the individual solutions.

Since then, new solutions have been produced, nearly every year, with increasing accuracy. Since 1988 the organization responsible for compiling a combined set of coordinates has been the *International Earth Rotation Service* IERS [12.4.2]. The current status can be found in the annual reports of the IERS (e.g. IERS, 2001). Recent solutions agree at an accuracy level of about 1 cm. The latest solution is the ITRF2000 [2.1.2.2].

Laser ranging to the LAGEOS satellites provides particularly significant contributions. Geocentric coordinates with a standard deviation below ± 1 cm can be determined from the analysis of long orbital arcs for appropriately equipped laser tracking stations [8.5.4]. SLR is hence of particular importance to the ITRF origin. VLBI provides global relative accuracies at the order of a few millimeters and also contributes significantly to the scale [11.1.2].

SLR and VLBI form the backbone of the global reference frame, which is then densified mainly by DORIS [6.7] and GPS [7.6.2.1]. Contributions from GLONASS and GALILEO can be expected. One early regional densification project is the European reference frame EUREF, since 1989. ITRF stations in Europe (Laser and VLBI) were used as fiducials and densified by GPS. Another example is SIRGAS for South America [7.6.2.1].

12.1.3 Operational Positioning

The reader is referred to the comprehensive discussion in [7.6.2]. With respect to the instrumental effort, until about 1986 it was mainly the TRANSIT Doppler method which was applied in operational positioning [6.6]. For the time being, GPS techniques are almost exclusively used. Three main fields of application can be distinguished:

(1) geodynamics (crustal deformation),
(2) control surveys, and
(3) special surveys.

(1) *Geodynamics*

Relative position information is primarily needed for the analysis of crustal deformation. The required accuracy of the coordinates is in most cases ± 1 cm or better. This corresponds to

- $1 \cdot 10^{-6}$ relative accuracy for 10 km station separation,
- $1 \cdot 10^{-7}$ relative accuracy for 100 km station separation, and
- $1 \cdot 10^{-8}$ relative accuracy for 1000 km station separation.

The only applicable operational observation technique so far is GPS [7.6.2]. With respect to the required high accuracy level, only dual-frequency receivers should be used, even over very short distances (cf. ionospheric disturbances [7.4.4.1]). Precise orbits are available from the IGS. Observations should always be connected to ITRF or IGS stations as fiducial points, if necessary via some densification stations based on several days or even weeks of GPS data collection.

In areas of high risk (e.g. earthquake, volcanic activities) permanent GPS arrays play an increasingly important role [7.6.2.2]. Geodynamic processes can also be monitored at stations equipped with DORIS receivers [6.7].

(2) *Control surveys*

Considering the high accuracy potential and the three dimensional character of operational satellite techniques, a *fundamental network in the satellite datum* (i.e. WGS 84 or ITRF) will be required for all countries, even in already well-surveyed areas. All follow-up surveys must be related to this fundamental network. Fig. 7.77 (p. 359) gives an example, the German GPS reference network DREF [7.6.2.1].

In the *first step*, a homogeneous network has to be established with as many simultaneously working satellite receivers as possible. The inter-station distances may vary between 50 and 150 km, depending on the situation. The network should be tied via simultaneous observations to an existing geocentric datum, in general the latest ITRF solution or a densification thereof, with GPS.

The DREF network is tied to about 15 EUREF stations, most of them have been determined by SLR or VLBI. The interstation distances are about 100 km. In most cases the number of available satellite receivers is not sufficient to observe all stations in the national network simultaneously. The connection between the individual observation phases (sessions) is achieved through permanently operating reference stations and a certain number of overlapping stations.

If the network is not tied to ITRF stations or regional densifications like EUREF or SIRGAS, the set of coordinates after the network adjustment provides a mean realization of the satellite datum (mostly WGS 84) for the observation period. The absolute accuracy is only a few meters (for GPS under activated SA it only was 30 to 100 meters). This classical method of datum realization through observations today is no longer adequate.

If the network is tied to one or more reference stations which are already linked to ITRF, the datum of the network is taken partly or completely from the reference points [12.1.1]. Observations in WGS 84 can easily be used for the interpolation of new stations in the ITRF reference frame.

The set of coordinates in the fundamental network, once it has been determined in one of the ways described, must be held fixed for subsequent densification work, because the high accuracies desired can only be achieved in the relative mode. The set of coordinates represents the satellite datum in the particular country at a particular epoch. With respect to the increasing international cooperation in geodesy, it is advisable to link the national network to the latest solution of the ITRF.

For the *densification* of the fundamental network see [7.6.2.1]. In order to avoid misunderstanding, it should again be stressed that network densification with satellite methods is only possible with observations made relative to at least one known station. Otherwise the uncertainty of the datum realization (several meters to tens of meters) would enter the result. This is also true if precise datum transformation parameters are available for the area.

Because of the high accuracy requirements, even in areas with existing classical geodetic control, the GPS technique will in future nearly exclusively be used for the establishment and densification of networks, possibly alongside GLONASS and GALILEO. The issue of combining satellite networks with existing terrestrial geodetic control will continue to wane in importance.

For some years of transition, the existing terrestrial networks will be maintained as national geodetic reference frame in many countries, and GPS will only be used as a method of densification. Because of the distortions and inhomogeneities in most classical networks, GPS results have to be incorporated into existing networks using local transformation parameters. Usually, the high internal homogeneity and relative accuracy of the satellite results is lost due to this procedure. On a long-term basis it is hence advisable to transform all existing surveying points into the distortion-free reference frame defined by GPS reference stations [7.6.2.1].

The hierarchic structure of classical networks from first to third or fourth order will disappear, and be substituted by a new hierarchy. The future structure of this hierarchy and its accuracy standards is already discernible (cf. [7.6.2.1]) and includes three levels:

A: Continental (or Sub-Continental) Reference Frame, based on the ITRF, with interstation distances between 200 and 500 km, and an accuracy goal of 1 cm over 500 km.

B: Nationwide Fundamental Network, with interstation distances of 50 to 100 km, and an accuracy goal of 1 cm over 100 km.

C: All other GPS measurements, which must be connected to stations of level B. Accuracy standards correspond to the particular purpose. For control surveys the requirement is to maintain 1 cm relative accuracy.

(3) *Special Surveys*

This group of tasks belongs to level C. The accuracy requirements may be quite different, and range between meters for GIS applications and millimeters for high precision engineering surveys. For details see [7.6.2.4], [7.6.2.6].

12.2   Gravity Field and Earth Models

12.2.1   Fundamentals

One of the main objectives of the *dynamical method* of satellite geodesy is to obtain information about Earth's gravity field from the dynamic behavior of satellite orbits [1.2]. The basic equations and explanations are given in chapter [3.2.2]. Earth's gravitational potential is usually represented in the form of functions on a unit sphere, the so-called *surface harmonics*. Following the fundamental equations in potential theory we find for the gravity potential W (Torge, 2001), that

$$W = V + \Phi, \qquad (12.6)$$

with V the *gravitational potential* and Φ the *centrifugal potential*. With respect to the orbital motion of satellites, only the gravitational potential, V, is of interest. V is determined by the mass distribution inside the body being orbited, including its atmosphere. According to Fig. 12.2, the following relationship exists:

$$V = G \iiint_{\text{Earth}} \frac{dm}{e}, \qquad (12.7)$$

where

  A attraction point,
  dm mass element,
  G gravitational constant,
  e distance of the mass element from the attraction point,
  a distance of the mass element from the geocenter, and
  ψ central angle, corresponding to the spherical distance on a unit sphere.
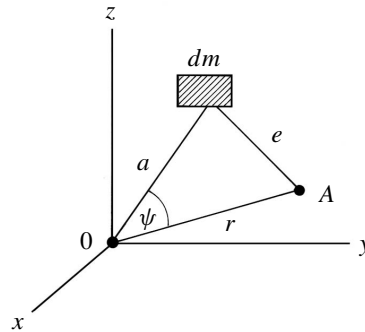
Figure 12.2. Relationship in the development of the gravitational potential

V is expanded into a series with the distance, e, written in spherical coordinates a, r, ψ. For the reciprocal distance, e, it follows that

$$\frac{1}{e} = \frac{1}{r} \sum_{n=0}^{\infty} \left(\frac{a}{r}\right)^n P_n(\cos \psi), \qquad (12.8)$$

where $P_n(\cos \psi)$ represent polynomials of nth degree in $\cos \psi$, known as *Legendre polynomials*, or *zonal harmonic functions* (e.g. Heiskanen, Moritz, 1967; Torge, 2001). Introducing (12.8) into (12.7) leads to

$$V = G \iiint_{\text{Earth}} \frac{dm}{r} \sum_{n=0}^{\infty} \left(\frac{a}{r}\right)^n P_n(\cos \psi). \qquad (12.9)$$

In practice, the series is truncated at a certain degree, i, i.e. the potential, V, is sub-divided into a first part, V_i, which is described in mathematical form, and into the remaining part, the *disturbing potential*, T, so that

$$V = V_i + T. \qquad (12.10)$$

In older literature, V_i is also known as the *level spheroid of rank i*. Introducing spherical coordinates (polar angle, ϑ, and longitude, λ) we obtain the representation of the gravitational potential as it is used in satellite geodesy, written in *Laplacian spherical harmonics*, see (3.109),

$$V = \frac{GM}{r}\left(1 + \sum_{n=1}^{\infty}\sum_{m=0}^{n}\left(\frac{a_e}{r}\right)^n (C_{nm}\cos m\lambda + S_{nm}\sin m\lambda)P_{nm}(\cos\vartheta)\right). \quad (12.11)$$

The harmonic coefficients, C_{nm}, S_{nm} are mass integrals and describe the distribution of masses inside Earth. a_e is the equatorial radius, and P_{nm} are the *associated Legendre functions*. Usually the origin of the coordinate system is shifted to the center of mass, hence the terms with $n = 1$ vanish, see (3.110). If rotational symmetry can be assumed, only *zonal coefficients*, dependent on the latitude, are present, leaving

$$V = \frac{GM}{r}\left(1 + \sum_{n=2}^{\infty}\left(\frac{a_e}{r}\right)^n C_n P_n(\cos\vartheta)\right). \qquad (12.12)$$

In many cases, in particular in older literature, the coefficients

$$J_n = -C_n, \quad J_{nm} = -C_{nm}, \quad K_{nm} = -S_{nm} \qquad (12.13)$$

are used in satellite geodesy. Also, the *fully normalized* coefficients $\overline{C}_{nm}$, $\overline{S}_{nm}$ are applied (3.188).

A particular harmonic coefficient can be explained geometrically as a certain deviation from the sphere. Fig. 12.3 illustrates the low zonal harmonics. The coefficient

$$J_2 \qquad\qquad J_3 \qquad\qquad J_4 \qquad\qquad J_5$$
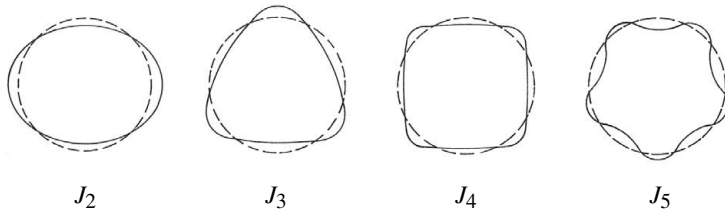
Figure 12.3. Low zonal harmonics ($J_2, \ldots, J_5$) and the figure of Earth

J_2 corresponds to Earth's flattening, J_3 to a triangular form, J_4 to a quadratic shape, and so on. The superposition of all the terms leads to a spectral representation of the

gravitational potential. The corresponding representations of the *gravity anomalies*, Δg, and the *geoid undulations*, N, are (Khan, 1983), see also (Torge, 2001):

$$\Delta g = \left(\frac{GM}{a_e^2}\right) \sum_{n=2}^{\infty} \sum_{m=0}^{n} \left(\frac{a_e}{r}\right)^{n+2} (n-1)$$
$$\times (\delta C_{nm} \cos m\lambda + \delta S_{nm} \sin m\lambda) P_{nm}(\cos \vartheta), \qquad (12.14)$$
$$N = \left(\frac{GM}{a_e \gamma}\right) \sum_{n=2}^{\infty} \sum_{m=0}^{n} \left(\frac{a_e}{r}\right)^{n+1} (\delta C_{nm} \cos m\lambda + \delta S_{nm} \sin m\lambda) P_{nm}(\cos \vartheta),$$

with

$$\delta C_{nm} = \text{observed } C_{nm} - \text{reference } C_{nm},$$
$$\delta S_{nm} = \text{observed } S_{nm} - \text{reference } S_{nm}.$$

A rotational ellipsoid, defined by C_{20} and C_{40}, is usually selected as a reference figure, hence $C_{nm} = S_{nm} = 0$ for $m \neq 0$. The coefficients δC_{nm} and δS_{nm} are then the "observed" potential coefficients, and γ is the *normal gravity* (cf. Torge, 2001).

In *dynamical* satellite geodesy the potential coefficients C_{nm} and S_{nm} are derived (observed) from the analysis of perturbations in satellite orbits, and then introduced into equations (12.11) to (12.14). The coefficients are also called *field parameters*. They form part of *Earth models* [12.2.2] and contribute with (12.14) to the computation of a global geoid.

With an increase in the degree of the series development of the potential, the standard deviations of the observed coefficients also increase. Only coefficients up to degree and order 36, and some higher order resonance terms, can be determined significantly from orbit analysis. The numerical reason for this comes from the "attenuation factor", $(a_e/r)^n$, in equation (12.11), that decreases rapidly with growing n. A much higher resolution cannot be expected, even with significant improvements in the accuracy of satellite tracking (cf. Fig. 12.4, Balmino (1983)). Only dedicated gravity field satellite missions [9], [10], or the combination with terrestrial gravity data can improve the situation.

The coefficient J_2 delivers an interesting geodetic result, namely the flattening, f, of a mean (best fitting) reference ellipsoid. It holds (Torge, 2001)
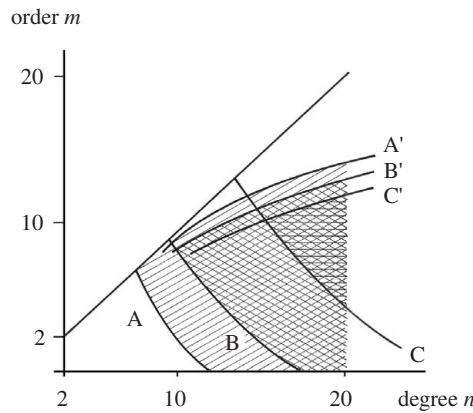
Figure 12.4. Tracking accuracy and gravity field resolution; the coefficients within the hatched area cannot be recovered with a given tracking accuracy (A = 10 m, B = 1 m, C = 0.2 m)

that

$$f = \frac{3}{2}J_2 + \frac{m}{2} + \frac{9}{8}J_2^2 + \frac{15}{28}J_2m + \frac{3}{56}m^2, \tag{12.15}$$

with

$$m = \frac{a_e^2 b\omega^2}{GM}.$$

The currently adopted best value is (Moritz, 2000)

$$f = 1 : 298.257,$$

that forms part of the *Geodetic Reference System 1980*. The value given in the IERS Conventions 2000 is

$$1/f = 298.25642 \pm 0.00001.$$

The existence of a coefficient J_3 was confirmed very early on through satellite observations. J_3 denotes an asymmetry in Earth's figure with respect to the equator (Fig. 12.5). This discovery became famous at the time as Earth's "pear-shape". This assignment is misleading, however, because the asymmetry, if compared with Earth's flattening, is smaller by a factor of 10^3. Even geoid undulations are much larger in some parts of the world.
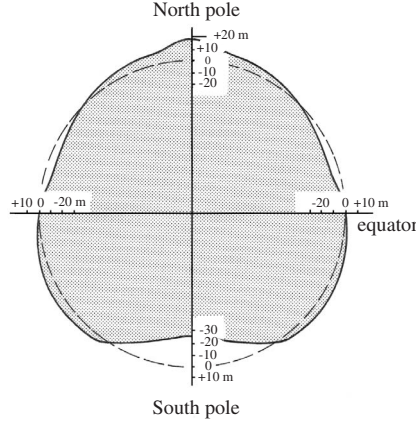
Figure 12.5. "Pear-shape" of the Earth due to the zonal coefficient J_3

In the general concept of dynamical satellite geodesy (cf. [4.1]), the parameters describing the geometrical figure and Earth's gravity field, and the elements of the satellite orbits, are solved for simultaneously. The basic concept is as follows (cf. Sigl, 1984), starting from equation (12.1),

$$r_S = r_B + \rho,$$

with
  r_S geocentric position vector of satellite S,
  r_B geocentric position vector of station B, and
  ρ topocentric observation vector of S from B.
(12.1) is rewritten in a generalized form as

$$r_S^T(t) = r_S^B(t). \tag{12.16}$$

The left hand side is computed as the "theoretical satellite vector", from orbital theory, as

$$r_S^T(t) = r_S^T(t; \alpha_1, \ldots, \alpha_6 ; p_1, \ldots, p_n), \tag{12.17}$$

with

$\alpha_1, \ldots, \alpha_6$ orbit parameters, and

$p_1, \ldots, p_n$ gravity field parameters.

The right hand side is the "observation vector" and contains

$r_{B1}, \ldots, r_{Bq}$ the coordinates of the tracking stations

as well as

$l_1, \ldots, l_m$ the corresponding satellite tracking observables (e.g. laser ranging),

and hence yields

$$r_S^B(t) = r_S^B(t; l_1, \ldots, l_m;\ r_{B1}, \ldots, r_{Bq}). \tag{12.18}$$

In the next step equations (12.17) and (12.18) are linearized:

$$r_S^T(t)\Big|_{p_0}^{\alpha_0} + D_\alpha r_S^T d\alpha + D_p r_S^T d p = r_S^B\Big|_{r_{B0}}^{l_0} + D_l r_S^B dl + D_{r_B} r_S^B dr_B \tag{12.19}$$

with

α	orbital parameters ($\alpha_1, \ldots, \alpha_6$),
p	parameters of the gravity field ($p_1, \ldots, p_n$),
l	observations ($l_1, \ldots, l_m$),
r_B	geocentric station positions ($r_{B1}, \ldots, r_{Bq}$),
$D_\alpha, D_p, D_l, D_{r_B}$	matrix of coefficients, and
$d\alpha, d p, dl, dr_B$	corrections.

The expression

$$r_S^T\Big|_{p_0}^{\alpha_0} \tag{12.20}$$

denotes a reference orbit, which is computed from approximate orbital elements, α_0, and approximately known gravity field coefficients, p_0.

$$r_S^B\Big|_{r_{B_0}}^{l_0} \tag{12.21}$$

is the observed orbit, starting from the approximate coordinates, r_{B_0}, of the tracking stations and the original, uncorrected observations, l_0. From the parameter estimation process corrections, $d\alpha$, $d p$, and dr_B are estimated for the orbit, the gravity field coefficients, and for the coordinates of the tracking stations. dl are the corrections to the observations, and are also called *bias parameters*.

In the concept of dynamical satellite geodesy, therefore, the deviations of the actually observed orbit from a precomputed approximate reference orbit are analyzed. The simultaneous estimation of improved station coordinates and gravity field parameters by the adjustment of all available satellite observations provides an *Earth model*.

At the same time, the semi-major axis of a best-fitting global ellipsoid, and the geocentric gravitational constant, GM, can be solved for. In order to stabilize the

solution, the zonal harmonic coefficients can be determined with comparatively high accuracy beforehand, and then be introduced as known parameters in the adjustment of the longitude-dependent tesseral coefficients. In a similar way the station coordinates can be determined first, and then held fixed in the solution of the gravity field models.

12.2.2   Earth Models

Satellite observations have been collected and introduced into data banks since the launch of the first artificial satellites, and they are used from time to time for refined Earth model computations. Up to 1997 more than 60 such models have been published, and their number is still growing. A current overview is given by Rapp (1998). In the following some of the best-known models are summarized.

(1) *Standard Earths of the Smithsonian Astrophysical Observatory* (SAO-SE)
The first SAO Standard Earth was published as early as 1966 (Lundquist, Veis, 1966), and was based on more than 45 000 directional observations with Baker–Nunn cameras from 12 ground stations to 13 satellites. As a result, geocentric coordinates of the 12 tracking stations, zonal coefficients up to J_{14}, and a complete development of the potential field up to degree and order 8 were published. In addition, numerical values were determined for GM and a_e:

$$GM = 398\,603.2 \text{ km}^3\text{s}^{-2} \quad \text{and} \quad a_e = 6\,378\,165 \text{ m.}$$

In 1969 an updated Standard Earth *SAO-SE II* was published, and in 1973 another update with *SAO-SE III* (Gaposchkin, 1973). In SE II the first laser measurements to satellites were included. In SE III camera and laser observations to a total of 25 satellites, and surface gravity data were used. The parameters of the gravity field were complete up to $l = 18$. Geocentric coordinates were determined for 90 tracking stations.

(2) *Goddard Earth Models* (GEM)
A long series of Earth models has been determined at the NASA Goddard Space Flight Center (GSFC). The first model GEM 1, exclusively based on satellite data, was published in 1972 and included a development of the gravity potential field up to $l = 12$. The model GEM 9 from 1979 was the last model in this series to be exclusively derived from satellite observations. In total, 840 000 observations to satellites were included, namely
  − 150 000 camera observations [5.1],
  − 477 000 electronic observations [4.4], [6], and
  − 213 000 laser ranges [8].
The potential development is complete up to $l = 20$; geocentric coordinates were determined for about 150 tracking stations.
  Since the sensitivity of orbital analysis stopped at about $l = 20$, altimeter data [9] and surface gravity data were included for further refinements. The related GEM models are (Lerch et al., 1978):

GEM 10	(1977),	$l = 22$,	satellites + surface gravity (1)
GEM 10A	(1978),	$l = 30$,	satellites + surface gravity (1) + altimetry (1)
GEM 10B	(1978),	$l = 36$,	satellites + surface gravity (1) + altimetry (2)
GEM 10C	(1978),	$l = 180$,	satellites + surface gravity (2) + altimetry (3).

About 700 altimeter passes from GEOS-3 are contained in GEM 10B; the model GEM 10C includes 2 300 altimeter arcs from GEOS-3, and in addition 384 000 mean $1° \times 1°$ gravity anomalies from surface data. GEM 10C was for many years one of the most used high resolution spherical harmonic models.

Along with the adjustment of the models GEM 9 and GEM 10, an improved numerical value for GM was determined, mainly based on LAGEOS laser ranging:

$$GM = 398\,600.64 \text{ km}^3\text{s}^{-2}.$$

The semi-major axis of a best-fitting global ellipsoid was determined from an analysis of the geocentric station coordinates as

$$a_e = 6\,378\,139 \text{ m}.$$

The dedicated, tailored gravity model GEM-L2 was published in 1983 for analysis of LAGEOS observations (Lerch et al., 1983). The model was developed up to $l = 20$, and was based on about 400 000 laser ranges to LAGEOS from 2.5 years, in addition to the GEM 9 data. The geocentric coordinates of the 20 tracking stations involved were known to an accuracy of ± 6 cm. In particular, the long-wave components up to $l = 4$ were extremely precise and allowed the modeling of the related geoid with an accuracy of better than 10 cm. The GM value of the solution is

$$GM = 398\,600.440 \text{ km}^3\text{s}^{-2}.$$

With respect to the TOPEX/POSEIDON mission [9.2] new tailored models were developed for precise orbit determination. The pre-launch model GEM-T1 was complete to degree 36 and based on satellite tracking data from 27 satellites. GEM-T3 (Lerch, 1992) was complete to degree 50 and included, in addition, satellite altimeter data from GEOS-3, SEASAT, and GEOSAT, as well as surface gravity data.

(3) GRIM Earth Models

The first Earth model GRIM 1 was published in 1976 jointly by German and French research groups (Balmino et al., 1976a). GRIM 1 was exclusively based on satellite observations, and was developed up to $l = 10$. A combination of the satellite solution with $1° \times 1°$ surface gravity anomalies (GRIM 2) was developed up to $l = 30$, and published in the same year (Balmino et al., 1976b).

The solution GRIM 3 (Reigber et al., 1983) yields a complete potential development up to $l = 36$, and geocentric coordinates of 95 tracking stations. The solution is based on camera, laser and Doppler observations to 22 satellites, some 25 000 $1° \times 1°$ terrestrial gravity anomalies, and nearly 28 000 $1° \times 1°$ gravity anomalies from GEOS-3 altimetry.

The development of the GRIM 4 series of solutions started in 1991 (Schwintzer et al., 1992). The last version was GRIM4 S-4, GRIM4 C-4 (Schwintzer, et al., 1997). S stands for "satellite-only" model and C for "combined model". GRIM4 S-4 includes in total 2.65 million single observations to 34 satellites with orbital heights between 800 km and 20 000 km. The model is complete to degree $l = 72$ and order $m = 72$. This corresponds to a spatial half wavelength of $D = 300$ km. For several years GRIM4-S4 was considered to be one of the best available satellite-only geopotential models (Rummel et al., 2002). The GM value for GRIM4 S-2 is (Schwintzer et al., 1992)

$$GM = 398\,600.4369 \pm 0.0028 \text{ km}^3\text{s}^{-2}.$$

The latest solution is GRIM 5, including observations to GFZ-1, and has been available since October 1997 (Schwintzer, et al., 2000). The satellite-only solution GRIM 5-S1 includes data from 21 satellites and is developed to degree $l = 99$ and order $m = 95$ (Fig. 12.6). The combination model GRIM 5-C1 includes in addition
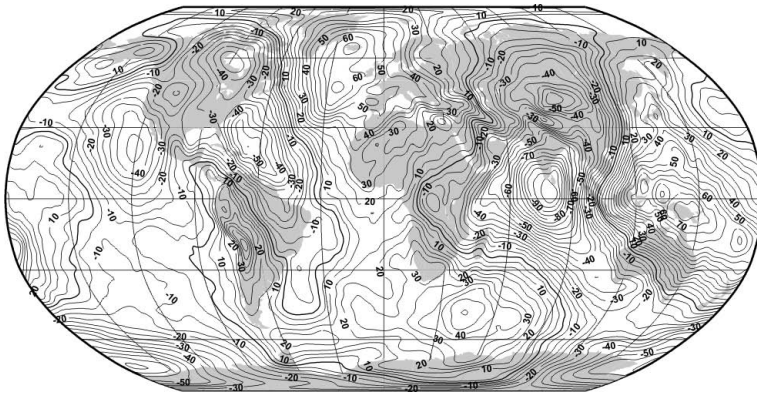
Figure 12.6. GRIM5-S1 Geoid (5 m contour lines), Schwintzer, et al. (2000)

gravity anomalies from surface data and altimeter data. The degree of development is $l = m = 120$.

(4) *Other Earth Models*

A large number of other Earth models exists; they cannot all be mentioned in this book. Reviews are given by Wenzel (1985); Rapp (1998); Torge (2001). High resolution models are always based on a combination of satellite data with terrestrial data. An ultra-high resolution model with 1800×1800 coefficients was developed by Wenzel (1999).

Several *dedicated gravity models* have been developed for particular satellites like LAGEOS, STARLETTE [8.5.3], GEOS-3, SEASAT-1, GEOSAT, and ERS-1/2, TOPEX/POSEIDON [9.3.3]. One of the currently most used models is the *Joint Geopotential Model* EGM96, complete to degree and order 360 (Lemoine et al., 1998), see also [8.5.3].

Together with the determination of Earth models, and with the increasing amount of satellite tracking data, the numerical value for the semi-major axis, a_e, of a best fitting mean Earth ellipsoid can also be continuously improved. Fig. 12.7 shows the development of our knowledge of a_e over the course of the last century. The result from 1967 contains, for the first time, satellite tracking data. The latest value from the IERS Conventions 2000 is
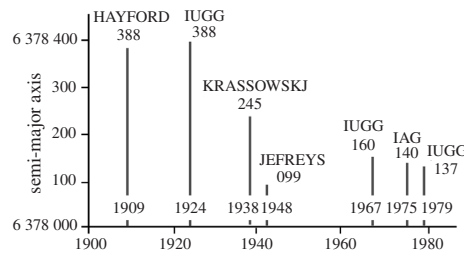
Figure 12.7. Numerical values for the semi-major axis of a mean Earth ellipsoid since 1900

$$a_e = 6\,378\,136.6\,\text{m}.$$

(5) Geocentric Gravitational Constant GM

The product, GM, of the mass of the Earth, M, and the universal gravitational constant, G, is a very important parameter in dynamical satellite geodesy because it contributes to the scale in the coordinate results. GM is one of the defining constants in the *Geodetic Reference System* 1980, and is usually treated as an errorless quantity. The parameter is always given in the form of the product, because this can be derived from satellite observations with a much higher accuracy than can the single factors.

In astronomical research, GM has been derived from analysis of the lunar orbit, following Kepler's 3rd law. A much higher accuracy can be achieved from observation of the "free fall acceleration" acting on space vehicles, in particular on interplanetary probes and distant satellites. This is explained by the fact that during the first days of a space flight, the motion of the space vehicle is primarily governed by the central term of the terrestrial gravitation.

The first reliable results were obtained from analyses of the *Mariner, Ranger, Surveyor, Lunar-Orbiter* and *Venera* probes (NGSP, 1977, Vol. 1 p. 292). The most accurate numbers today come from several years of analysis of laser ranging to the Moon and to LAGEOS. Some results are given in Table 12.2.

Low orbiting satellites are less suited for the determination of GM because of the rather strong non-gravitational perturbations. Fig. 12.8 illustrates development in the determination of GM. The dramatic improvement of the related accuracy is noticeable.
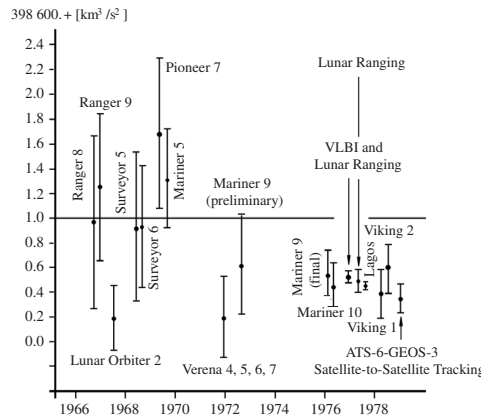
Figure 12.8. Development in the determination of the geocentric gravitational constant GM (Martin, Oh, 1979)

Table 12.2. Different determinations of GM, [km^3 s^{-2}]

Method/Source	Numerical Value	
Lunar orbit (O'Keefe, 1958)	398 620	±6
C_{20}, C_{40} terrestrial data (Rabe, 1962)	398 603	±6
Ranger (6–9) (1966)	398 601	±0.7
Mariner 9 (1971)	398 601.2	±2.5
Venera (4–6)	398 600.37	±1.0
Laser ranging to the moon (Williams et al., 1987)	398 600.444	±0.010
Laser ranging to LAGEOS (Smith, et al., 1985)	398 600.434	±0.002
Laser ranging to the Moon (Dickey et al., 1994)	398 600.443	±0.004
Laser ranging to LAGEOS (Smith et al., 2000)	398 600.44187	±0.00020
Geodetic Reference System 1980	398 600.5	±0.5
MERIT Standards 1983	398 600.448	
IERS Conventions (McCarthy, 2000)	398 600.4418	±0.0008

12.3 Navigation and Marine Geodesy

12.3.1 Possible Applications and Accuracy Requirements in Marine Positioning

The accuracy requirements for marine positioning, on the sea surface and the ocean floor, are increasing, along with the growing interest in the ocean areas with respect to economy, sea traffic, mineral and living resources, their associated legal aspects, and global geodynamics. The following main fields of interest can be identified.

(a) *Law of the Seas and Marine Boundaries* Due to the new codified *Law of the Seas*, large parts of the open oceans are now characterized as *Exclusive Economic Zones* (EEZ) under the authority of the adjacent states. Boundaries have to be defined, fixed and set out, sometimes several hundred miles off the coast. In cases where the median line between two states runs through areas with oil and gas deposits, or other resources of economic interest (cf. Fig. 12.9), the accuracy requirements may be less than 5 meters. Delimitation of the outer edge of the continental shelf may require detailed mapping of sediment thickness, and inclination of the continental slope.
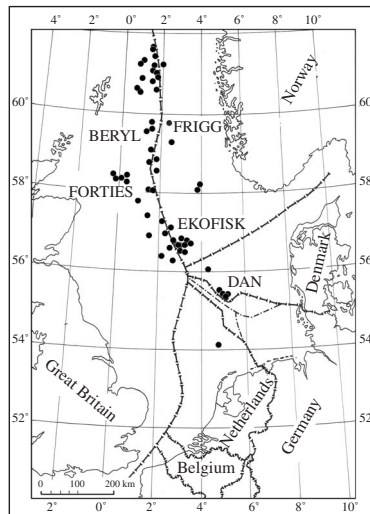
Figure 12.9. Median lines in the North Sea, and some oil and gas fields

(b) *Offshore Industry and Exploration*

This field of interest is concerned with such applications as: geological mapping of the seafloor with seismic techniques; installation of drilling sites and offshore structures; laying and control of pipelines; recovery of closed boreholes; and the exploration and extraction of mineral resources (such as manganese nodules or hydrothermal deposits). The most demanding requirements are related to detailed seismic surveying (3-D seismic), and to the exploration and exploitation of small-scale deposits in deep sea areas.

(c) *Sea Floor Mapping*

Detailed, high-resolution mapping of the ocean floor with multibeam sonar systems like SEABEAM or HYDROSWEEP has developed into one of the most powerful techniques in marine geoscience (Schenke, 1991; Schenke et al., 1998). An essential requirement is knowledge of the precise relationship between the ship's position and the multibeam data (Fig. 12.10) including the ship's attitude [7.6.2.7].
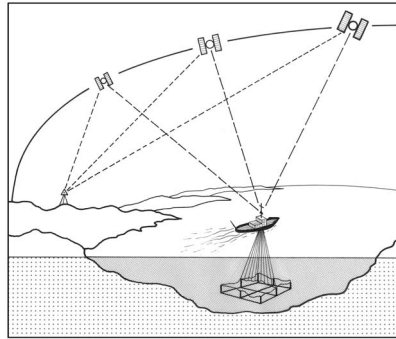
Figure 12.10. Sea bottom mapping with multibeam sonar systems and differential GPS

(d) *Geodynamic Research*

Tectonic plate boundaries run mainly through marine areas. In order to derive a representative pattern of recent global tectonics, it is necessary to include the ocean floor in motion and deformation studies. Deformation rates of 1 to 5 centimeters per year have to be determined. This very demanding task can only be solved by a combination of satellite techniques (e.g. GPS) with acoustic under-water range measurements [12.3.2].

(e) *Global Sea Level and the Marine Geoid*

For the establishment of a global height system differences in absolute sea level must be known over large oceanic areas (cf. Fig. 9.19, p. 468). In precise marine gravimetric surveying, for the contribution to a marine geoid, the height component must be known to about 0.1 m, and ship's velocity components to about 10^{-3} m/s. Gravimetric observations are also important in marine geophysics for the detailed study of features below the sea floor. The precise position of buoys is required for the calibration of altimeter missions [9.3.3].

12.3.2 Marine Positioning Techniques

The requirements in marine positioning, indicated above, can be met with various observation techniques, either alone or combined (Fig. 12.11). These are:

— shore based radio navigation techniques,

- satellite radio navigation techniques
- acoustic techniques,
- inertial techniques, and
- integrated techniques.

Shore-based radio navigation techniques (Egge, Seeber, 1979; Forsell, 1991) played an important role in marine positioning until about 1995, but are today mostly outdated. In most cases, they could only fulfill many accuracy requirements near the coast.The only precise positioning techniques available at a global scale are the satellite methods. Acoustic techniques are required for guidance, positioning and control of underwater objects, and in marine geodynamics. Inertial techniques have sel-

Figure 12.11. Different techniques in marine positioning

dom been used because of the high costs; they may be of more interest, in the future, as a supplement to GPS, in particular for attitude control and/or for bridging data gaps in hydrographic surveying (Böder, 2002). Integrated techniques have been, for many years, the only officially recognized and globally applicable methods for higher accuracy demands. Their significance has changed with the increasing availability of GPS. In the following some particular aspects of marine positioning will be addressed in short.

Use of the Satellite Doppler Technique

The TRANSIT system [6], using the Doppler effect for radio signals, was originally developed for marine navigation; its possible application to precise positioning on land was recognized and exploited much later. One of the characteristic features of this method is that positioning data can only be obtained for a limited period of 15 to 18 minutes during a particular *satellite pass*. Between two consecutive satellite passes the observer has to wait for up to several hours, depending on the satellite constellation and the geographic location. The following principal limitations are associated with the Doppler method for a moving user [6.6.2]:

- position information can only be derived from a single satellite pass,
- insufficiently known the observer's velocity severely corrupts the computed antenna position, and
- position information from other sources must be available for the time gaps between two satellite passes.

These limitations, together with the rather low accuracy, were the main reasons for a nearly complete substitution of the Doppler techniques by GPS after 1990. In principle, the DORIS system [6.7] can also be used for marine positioning and navigation. However, due to the limited real-time accuracy for moving platforms, the benefits of DORIS are preferably seen in the determination of satellite orbits and station locations.

Global Positioning System

The NAVSTAR GPS is the system "par excellence" for marine positioning and navigation because

- at least four, but in most cases more, satellites are visible from any position in the world, at any time, offering 24 hours of worldwide navigation capability,
- the positions are available in real-time,
- the achievable accuracy for a stand-alone user is, after the de-activation of SA, in many cases sufficient for general navigation,
- for precise positioning and navigation requirements it is possible to use correction services (DGPS) for most parts of the world, and
- if required, it is also possible to achieve an accuracy at centimeter level.

The conditions and possibilities of GPS use in the marine environment are discussed at length in chapters [7.5] and [7.6.2.7] and so they will not be repeated here. It can be expected that the Russian system GLONASS [7.7.1] and the European system GALILEO [7.7.3] will augment the constellation, and in particular will contribute to the availability and integrity of services [7.7.2].

Sea Floor Positioning

One of the most demanding tasks in marine geodesy is the determination of sea-bottom control points at centimeter level for monitoring of crustal deformation. Fig. 12.12 explains the situation. An array of control points is installed in active zones, each of which must be related to control points on land (case 2). In addition, the internal geometry between the sea-bottom points has to be monitored (case 1).

Case 1 is of interest at mid-ocean ridges, where control points can be established on either side of the diverging plate boundary. In general, it will be difficult to measure directly between the control points because of ray bending of acoustic waves in water. Hence it may be necessary to include a platform (moored buoy, ship) at the sea surface, or inside the water body as a relay station. Case 2 has the objective to determine motion

Figure 12.12. Seafloor positioning in geodynamic research, after Chadwell et al. (1998)

between an oceanic plate and a continental plate. The distance between the control points can reach several hundred or even several thousand kilometers. In both cases, the position and motion of a sea-surface platform has to be monitored.

The task includes two different surveying techniques. Only acoustic waves can propagate in water; hence *acoustic techniques* have to be applied to determine the position of sea bottom markers, and to control the relationship between the bottom

control points and the moving platform. The propagation of sound in seawater depends on many conditions (salinity, temperature, pressure), and is very difficult to model with the required high precision. The connection between the platform and a land-based reference station is via GPS, using long-range kinematic techniques [7.5]. For details and experiences see e.g. Chadwell et al. (1998).

12.4 Geodynamics

Geodynamics is an extremely broad field and it is developing fast (Turcotte, Schubert, 2002). Results from satellite geodesy are contributing considerably. For an overview of current developments see the results of the regular meetings (spring or autumn) of the *American Geophysical Union* (AGU), the series of IAG symposia proceedings, as well as the *Journal of Geophysical Research* (JGR) or *Journal of Geodesy*. Within this section only the topics *crustal motion*, *Earth rotation* and *reference frames* are briefly summarized. Another important field of application is the analysis of tides of the ocean and solid Earth, cf. [3.2.3.2] [8.5.6].

12.4.1 Recent Crustal Movements

The increasing accuracy of satellite-based geodetic positioning techniques makes it possible to derive information on the kinematics of tectonic plates from repeated or continuous observations. This can be regarded as one of the most important contributions of satellite geodesy to geodynamics. According to the model of *plate tectonics* (e.g. Le Pichon et al., 1973; Turcotte, Schubert, 2002), the crust of the Earth is divided in a number of thin, rigid plates moving with respect to each other. Plates are continuously being added to along the oceanic ridges from uprising material. In the collision zones between plates mountain chains, deep sea trenches, and island arcs form. At trenches one plate dives beneath the other in a process named subduction, forming *subduction zones*. Plate boundaries are defined by seismic activity and may also be characterized by volcanoes.

On this basis, several larger tectonic plates can be identified: the Pacific, North American, South American, Eurasian, Indian, African, Australian and Antarctic plates. In addition, several smaller plates are known, such as the Caribbean, Nacza, Cocos, and Arabian plate. The overall pattern of motion is fairly complicated, and can give rise to short lived *microplates*, especially in intraoceanic settings. Detailed knowledge of the kinematic behavior of these plates is fundamental to our understanding of its driving mechanism, and could contribute to a better understanding, and perhaps prediction, of seismic activity. It is also of importance for the maintenance of terrestrial reference frames [7.6.2.1].

Global models of plate tectonics can be established, based on geological, palaeomagnetic, and seismic investigations, and derived from accumulated motion rates over large periods of time. One generally recognized model was published by Minster, Jordan (1978) and has since been refined several times. The motion rates vary between 1

cm/year and about 10 cm/year. Only with modern satellite techniques has it become possible to prove these predictions, and to check whether the motions are continuing at the present time. Recent studies, since about 1980, indicate that tectonic plate motion, derived from geodetic measurements and averaged over several years is in general equal to that derived from geological observations and averaged over several million years. Suitable geodetic observation techniques are

- Very Long Baseline Interferometry (VLBI) [11.1],
- Satellite Laser Ranging (SLR) [8],
- GPS observations [7], and
- DORIS observations [6.7].

The motion of a plate can be described by a Cartesian rotation vector with the components $\Omega_x, \Omega_y, \Omega_z$. In order to derive a stable absolute reference frame for the plate motion, a condition can be introduced that the rotating lithospheric plates do not possess any angular momentum with respect to the rigid mesosphere. This condition is called *no-net rotation* (NNR). A generally used model based on this assumption is NNR-NUVEL-1A (DeMets et al., 1994). This model is also included in the IERS Conventions (McCarthy (2000), Table 12.3). The model can be used to compute new site positions from old site positions when no individual motion information (as e.g. for ITRF stations) is available.

Table 12.3. Cartesian rotation vector for each plate using the NNR-NUVEL-1A kinematic plate model (McCarthy, 2000)

Plate Name	Ω_x [rad/My.]	Ω_y [rad/My.]	Ω_z [rad/My.]
Pacific	-0.001510	0.004840	-0.009970
Cocos	-0.010425	0.021605	0.010925
Nazca	-0.001532	-0.008577	0.009609
Caribbean	-0.000178	-0.003385	0.001581
South America	-0.001038	-0.001515	-0.000870
Antarctica	-0.000821	-0.001701	0.003706
India	0.006670	0.000040	0.006790
Australia	0.007839	0.005124	0.006282
Africa	0.000891	-0.003099	0.003922
Arabia	0.006685	-0.000521	0.006760
Eurasia	-0.000981	-0.002395	0.003153
North America	0.000258	-0.003599	-0.000153
Juan de Fuca	0.005200	0.008610	-0.005820
Philippine	0.010090	-0.007160	-0.009670
Rivera	-0.009390	-0.030960	0.012050
Scotia	-0.000410	-0.002660	-0.001270

The NUVEL-1A model is based on magnetic seafloor-spreading anomalies of about the last three million years, and only refers to rigid plates. An alternative approach is

to derive an *Actual Plate Kinematic and Deformation Model* (APKIM) from present-day geodetic observations, such as VLBI, SLR and GPS. A series of such models has been developed (Drewes, 1999). In the latest solution, APKIM2000, about 280 site velocities were used to estimate 12 plate rotation vectors. The adjustment process allows to distinguish between rigid plates and deformation zones (Drewes, Angermann, 2001). In general the agreement between APKIM2000 and NUVEL-1A is very high. Significant differences, however, are visible in deformation zones. Fig. 12.13 gives an impression on the motion rates based on the geophysical model NNR-NUVEL1A and the actual kinematic plate model APKIM.
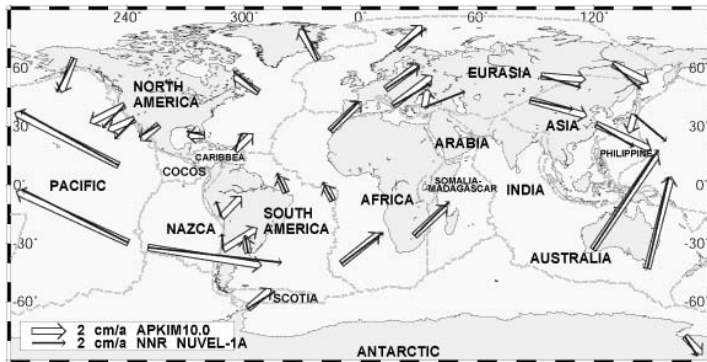
Figure 12.13. Main plate boundaries and expected motion rates (cm/year), source: DGFI

The available observations can be regarded as *zero observations*, but they already give promising results. As time goes on, and with continuing observations, the results will become more and more reliable. The longest series of observations was created within the framework of the NASA *Crustal Dynamics Program* (CDP) since 1979, based on repeated observations from approximately 35 Laser and 30 VLBI stations (status end 1991). The NASA CDP ended in 1991, and was replaced by the *Dynamics of the Solid Earth* (DOSE) program in 1992. Observations are now continued within the international services ILS, IVS, and IGS. By 2002, about 40 to 50 SLR stations and about 30 VLBI stations cooperated with regular observations within international projects. These stations form the backbone, to which are added more than 300 permanent GPS stations. As can be seen from the ITRF2000 network (Fig. 2.4), the global station distribution is not yet homogeneous. Nevertheless, the results from the regular solutions give an excellent insight into the detailed pattern of global plate motion and deformation. A number of regional projects and arrays, such as those in California, Iceland or Japan, and WEGENER in Europe, complete the picture [7.6.2.2], [8.5.4].

12.4.2   Earth Rotation, Reference Frames, IERS

The possible use of satellite observation techniques for the establishment of a terrestrial reference frame and models of the rotational behavior of the Earth has been discussed

in detail in previous sections [2.1.2], [7.6.2.1], [8.5.4], [8.5.5], [8.6], [11.1.2]. Several comments can be given to summarize.

Since about the beginning of the last century, polar motion and Earth rotation have been determined by international services through astronomical observations with fundamental instruments (e.g. photographic zenith tube, Danjon astrolabe). The responsible organizations were the *International Polar Motion Service* (IPMS), previously the *International Latitude Service* (ILS), and the *International Time Bureau* (*Bureau International de l'Heure*, BIH). About 50 globally distributed stations contributed with astrometric instruments on a regular basis. The accuracy was about $0.''1$ (corresponding to 3 m) for the pole coordinates and 0.7 ms for Earth's rotational velocity.

The related reference system was determined in 1968 through the adoption of the astronomical (geographical) coordinates of the total of 68 stations involved in the determination of Earth rotation (*BIH System 1968*).

Since 1967, almost as a by-product of the computation of precise ephemerides, pole coordinates have been determined by the Defense Mapping Agency from Doppler observations to TRANSIT satellites [6.6.2]. Since 1972, the results have been used by the BIH on a routine basis. This was the first operational use of satellite methods for the determination of Earth rotation parameters, and their capability and superiority was demonstrated. The coordinates of the pole were derived with an accuracy of ±40 cm from two-day solutions.

In the following years, the efficiency and capability of new space methods was demonstrated, in particular for VLBI and for laser ranging to satellites and to the Moon. Within the framework of several independent national and international projects, Earth rotation parameters were derived from observations with the new space techniques, and compared with the BIH products. The results were discussed at numerous international conferences (Gaposchkin et al., 1981), and led to the establishment of two working groups named

MERIT   *M*onitor *E*arth *R*otation and *I*ntercompare the *T*echniques of observation and analysis, and

COTES   *Co*nventional *T*errestrial *R*eference *S*ystem.

Within the scope of MERIT, the *MERIT Short Campaign* was observed from August to October 1980, and the *MERIT Main Campaign* for a period of 14 months (one *Chandler period*) from September 1, 1983 until October 31, 1984. Classical, as well as all available modern observation techniques for the determination of Earth rotation parameters and station coordinates were applied in a worldwide cooperative effort. Of particular interest was the participation of stations where different types of instruments were operated (collocation sites), which could be used to detect systematic differences between the techniques and their related reference frames. Observation stations were operated in 35 countries. During the *Main Campaign* the following types of observations were carried out:

Optical Astrometry (OA)   61 stations
Doppler Observations with TRANSIT (DTS)   20 stations

Satellite Laser Ranging (SLR)	27 stations
Lunar Laser Ranging (LLR)	3 stations
Connected Elements Interferometry (CEI)	1 station
Very Long Baseline Interferometry (VLBI)	8 stations.

A set of constants, parameters and correction models, the *MERIT Standards* (Melbourne et al., 1983), was defined to support the data analysis. These standards have since been widely used in the international scientific community and resulted later in the IERS standards and conventions.

The MERIT observations were of equal use for the objectives of COTES, namely the installation of a conventional terrestrial reference system. To support this objective, a 3-month *Intensive Campaign* was performed with (among others) daily VLBI observations, in order to study the relationship between the particular reference frames as materialized through the individual observation techniques.

The scientific results of MERIT and COTES have been published in the proceedings of the *International Conference on Earth Rotation and the Terrestrial Reference Frame*, July 31 to August 2, 1985, Columbus (Ohio, USA) (Mueller, Wei, 1985). The observation results are documented in Part III of this report (Feissel, 1986). They consist in each case of one or two lists:
  − Set of Station Coordinates (SSC), and
  − Earth Rotation Parameters (ERP).
The SSC-lists are simply the realization of a geocentric reference frame; the ERP results are compatible with the related SSC. It was demonstrated that the determination of polar motion and Earth rotation with SLR and VLBI was 5 to 6 times more accurate than the results of the existing service, based on astrometric and Doppler techniques (cf. [8.5.5], [11.1.2]).

Some of the analysis centers continued with their work after the end of the MERIT project, as the forthcoming International Earth Rotation Service was already in view. The results were published annually, until 1988, in part D (*Earth Rotation and Related Reference Systems*) of the BIH Annual Report.

In 1984, the BIH established a new reference system, based on the geocentric coordinates of those fundamental stations where geodetic space techniques were applied. The system was called the *BIH Terrestrial System* (BTS) and it coincided, within the related observational accuracy, with the former reference frame established by astronomical observations (Boucher, Feissel, 1984). The last realization, BTS 87, was derived from the SSCs of seven analysis centers contributing to the ERP series (BIH, 1988).

Based on the experiences with MERIT and COTES, it was proposed (Wilkins, Mueller, 1986) that a new International Service be established for monitoring Earth rotation parameters and the maintenance of a conventional terrestrial reference frame. This service was to replace the IPMS and the Earth rotation section of the BIH.

The new *International Earth Rotation Service* (IERS) was established in 1987 by the International Astronomical Union (IAU) and the International Union of Geodesy and Geophysics (IUGG), and started operation on January 1, 1988 (cf. [2.1.2.3], Boucher, Altamini (1989)), based on the following space techniques:

− Very Long Baseline Interferometry (VLBI),
− Satellite Laser Ranging (SLR), and
− Lunar Laser Ranging (LLR).

At a later date, GPS data from the International GPS Geodynamics Service (IGS) and DORIS data were also included.

Following the terms of reference (IERS, 2001), the "primary objectives of the IERS are to serve the astronomical, geodetic and geophysical communities by providing the following:

− the International Celestial Reference System (ICRS) and its realization, the International Celestial Reference Frame (ICRF),
− the International Terrestrial Reference System (ITRS) and its realization, the International Terrestrial Reference Frame (ITRF),
− Earth orientation parameters required to study Earth orientation variations and to transform between the ICRF and the ITRF,
− geophysical data to interpret time/space variations in the ICRF, ITRF or Earth orientation parameters, and model such variations,
− standards, constants and models (i.e. conventions) encouraging international adherence."

Like the other services of interest to satellite geodesy, namely the IGS [7.8.1], the ILRS [8.5.1], and the IVS [11.1.3], the IERS fulfills its tasks within several components, namely:

− Technique Centers,
− Product Centers,
− Combination Centers,
− Analysis Coordinator,
− Central Bureau, and
− Directing Board.

For details, see the Annual Reports of the IERS. On January 1, 2001, the Central Bureau was transferred from the Paris Observatory to the *Bundesamt für Kartographie und Geodäsie* (BKG), Frankfurt, Germany. The products of the IERS include, among others, the following:

− International Celestial Reference Frame (ICRF),
− International Terrestrial Reference Frame (ITRF),
− monthly Earth orientation data,
− daily rapid service Earth orientation data and predictions,
− leap second announcements,
− long term Earth orientation information,
− annual reports, technical notes, and conventions.

The ICRF is realized through the coordinates of compact radio sources and is based on VLBI observations (cf. [2.1.2.1], [11.1.2]). The sky distribution of sources is depicted in Fig. (2.3). The latest ITRF realization is ITRF2000 (2.1.2.2). It is based on about 800 stations located at about 500 sites. The solution is based on observations by VLBI, LLR, SLR, GPS and DORIS. Fig. 2.4 shows the global distribution of the

primary sites and indicates the collocated techniques. The datum is defined as follows (IERS, 2001):

— the origin and its rate are derived from the most consistent SLR solutions,
— the scale and its rate are based on VLBI and SLR solutions,
— the orientation is aligned to that of ITRF97 and its rate is such that there is no-net-rotation with respect to NUVEL-1A.

The long-term stability of ITRF2000 is estimated to be better than 4 mm in origin and better than 0.5 ppb in scale. This last figure corresponds to a shift in station height of about 3 mm over Earth's surface.

The *Earth orientation parameters* (EOP) are based on VLBI, SLR and GPS. Table 12.4 shows the contribution of each technique. It becomes evident that polar motion and the variation of Earth rotation (LOD, length of day) are nearly exclusively determined by GPS. LOD is the difference between the astronomically determined duration of the day and 86400 seconds of TAI.

Table 12.4. Percentage of contribution in the final IERS EOP solution (IERS, 2001)

Technique	Polar Motion	UT1-UTC	LOD	Nutation offset
IERS VLBI	20	100	10	100
IERS SLR	10	–	–	–
IERS GPS	70	–	90	–

The polar motion between 1996 and 2000 and the mean pole displacement for the last century is illustrated in Fig. 12.14.
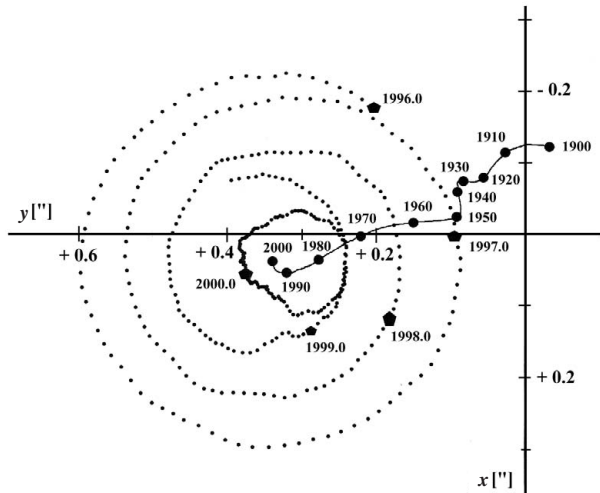
Figure 12.14. Polar motion 1996–2000 and mean pole displacement 1900–2000 (IERS, 2001)

As of April 2003 the name of the service was changed to *International Earth Rotation and Reference Systems Service* (IERS) in order to reflect the broader spectrum of generated products.

12.5 Combination of Geodetic Space Techniques

12.5.1 Basic Considerations

Geodetic space techniques can be combined to achieve more reliable and consistent results. The individual geodetic procedures have particular strengths and weaknesses; the objective of a combination is to compensate the shortcomings of one method with the strengths of the other. The strengths of the different observing techniques are, for example:

- SLR: relationship to the geocenter and to Earth's gravity field,
- VLBI: relationship to the inertial reference frame,
- DORIS: homogeneous global distribution of tracking stations,
- GPS: highly operational system for the densification of the terrestrial reference frame,
- CHAMP, GRACE, GOCE: high resolution Earth gravity field,
- Altimetry: structure and variation of sea level, connection of height systems.

Links between the different geodetic space techniques are possible (Rothacher, 2002)

- (a) at the level of stations,
- (b) at the level of satellites, and
- (c) at the level of parameters.

(a) Stations serve as a link when different types of observations are realized at the same location. Such sites are named *collocation sites*. The rationale behind the collocation of various techniques is that the same results are to be expected, for example station positions and velocities, and that the observations are not influenced by different atmospheric effects. It is essential to determine the local ties between the reference points of the participating instruments at the level of 1 to 2 mm. The combination can be realized by forming a simple weighted mean between the individual results, or by more sophisticated means using the complete variance-covariance matrix.

(b) Satellites serve as a link when a given satellite is tracked by different observation techniques. Precise orbit determination (POD) is supported by SLR, GPS, DORIS, and altimetry. The satellite "par excellence" is TOPEX/POSEIDON where all four techniques can be used. Two GPS satellites (SVN35 and SVN 36) and all GLONASS satellites carry SLR retro-reflectors. The observation of GPS or GLONASS signals with VLBI offers a direct link between the satellite orbits and the ICRF.

(c) The link at the level of parameters offers various possibilities because most parameters can be determined by different techniques. Table 12.5 gives an overview. The common parameters can be combined at the level of the normal equations or at the level of observations. The second approach is more flexible but also more demanding because only a few computer programs exist that are capable of treating all different data types in a consistent way (Rothacher, 2002). It is evident that consistent

Table 12.5. Common parameters for the combination of geodetic space techniques, after Rothacher (2002)

Parameter	VLBI	SLR	LLR	GPS/GLONASS	DORIS	Altimetry
ICRF Coordinates (Quasars)	X					
Nutation $\Delta\varepsilon$, $\Delta\psi$	X		X	(X)		
Pole x_P, y_P	X	X	X	X	X	
UT1	X					
Length of Day (LOD)		X	X	X	X	
Sub-daily ERPs	X			X		
ERP-Amplitudes (Ocean tides)	X	X		X		X
ITRF Coordinates and Velocities	X	X	X	X	X	(X)
Geocenter		X		X	X	X
Gravity Field		X	(X)	X	X	X
Orbits		X	X	X	X	X
LEO-POD		X		X	X	X
Troposphere	X	(X)		X	X	X
Ionosphere	X			X	X	X
Clocks (Time Transfer)	X	(X)		X		

standards and physical models (like the IERS Conventions) are required for a rigorous combination.

12.5.2 Fundamental Stations

In future, a number of fundamental stations with the most important geodetic space techniques will be required to maintain the fundamental reference frames. A geodetic observatory is called a *fundamental station for geodesy* if the following requirements a fulfilled (Schlüter et al., 2000):

- − permanent observation activities,
- − complimentary measuring techniques,
- − redundant observations for quality control, and
- − precise link between reference points from the different methods.

A homogeneous global distribution, with at least three stations on each major tectonic plate, is desirable. To fill the existing gaps, transportable systems are of interest that can be installed for a couple of years at sites of interest. Two examples are given, the permanent fundamental station *Wettzell* and the *Transportable Integrated Geodetic Observatory* (TIGO).

The station *Wettzell*, located in Bavaria in Southern Germany, includes all relevant geodetic space techniques. These are currently (Schlüter, 2002):

- a 20 m radio telescope, dedicated to geodetic VLBI [11.1.2],
- the Wettzell Laser Ranging System (WLRS), designed for measurements to satellites (SLR) and to the Moon (LLR) [8.3.3],
- several GPS receivers, Wettzell is an IGS "core-station" [7.8.1],
- a large number (45 in 2001) of remote controlled permanent GPS stations, contributing to IGS, EUREF, and the German reference network GREF [7.5.1.3],
- a time- and frequency laboratory (3 cesium oscillators and 3 hydrogen masers) [2.2.5],
- a super-conducting gravity meter for monitoring local gravity variations (e.g. due to tides),
- a ring laser gyro for the determination of Earth rotation, and
- meteorological sensors and a water vapor radiometer [2.3.3.2].

In addition, a high precision local network has been established to link all individual observing systems at an accuracy level of 1 to 2 mm. Seismic measurements are carried out for monitoring earthquakes. Fig. 12.15 gives an impression of the site.

Figure 12.15. Fundamental Station Wettzell; courtesy BKG, Frankfurt

The *Transportable Integrated Geodetic Observatory* (TIGO) is a transportable fundamental station built by the Bundesamt für Kartographie und Geodäsie (BKG) in Germany. Its purpose is to "provide observations for international services from a remote location in order to improve the realization and maintenance of global reference frames" (Schlüter et al., 2000). TIGO includes all relevant geodetic space techniques, namely

- Very Long Baseline Interferometry (6 m radio telescope),
- Satellite Laser Ranging (two color system, range up to 36 000 km), and

  − Microwave systems like GPS and GLONASS.

In addition, various sensors for local measurements are provided:
  − time and frequency to generate a local time scale (two cesium standards, two hydrogen masers, GPS time receivers) ,
  − gravity measurements for monitoring Earth tides (superconducting gravity meter),
  − seismic measurements for monitoring earthquakes (broad spectrum seismometer),
  − meteorological measurements (meteorological station and water vapor radiometer), and
  − local survey measurements for monitoring site stability and for linking the instrument reference points (high precision tacheometer and digital levelling instrument).

After an international request, a site in Concepción, Chile, was selected where TIGO is to be operated for several years, jointly by a Chilean consortium and BKG (Schlüter et al., 2002). Operations started in April 2002. Fig. 8.7 (p. 417) and Fig. 11.4 (p. 493) show some of the TIGO components.

12.5.3 Integrated Global Geodetic Observing System (IGGOS)

The integration of all geodetic space techniques and their services under one roof can be defined as an *Integrated Global Geodetic Observing System* (IGGOS). The concept of IGGOS has been widely discussed in the scientific geodetic community (Rummel et al., 2000; Drewes et al., 2002) and proposed as a *Project* in the new structure of the IAG, effective from 2003. IGGOS shall act as an interface between the IAG commissions, the IAG related services, and international non-geodetic organizations, for example the Committee on Space Research (COSPAR), the International Lithosphere Program (ILP), or the Scientific Committee on Antarctic Research (SCAR). The new IAG commissions are:
  1. Reference Frames,
  2. Gravity Field,
  3. Earth Rotation and Geodynamics,
  4. Positioning and Applications.

The existing services of IAG are:
  − International GPS Service,
  − International VLBI Service,
  − International Laser Ranging Service,
  − International Earth Rotation Service,
  − International Gravimetric Bureau,
  − International Geoid Service,
  − International Center of Earth Tides,
  − Permanent Service for Mean Sea Level,
  − International Service of Weights and Measures (Time Section).

Bibliography

Abbreviations frequently used:

AVN	Allgemeine Vermessungsnachrichten. H. Wichmann, Karlsruhe
Bull. Géod.	Bulletin Géodésique, Springer, Berlin-Heidelberg-New York
CSTG Bull.	International Coordination of Space Techniques for Geodesy and Geodynamics Bulletin, Columbus/Munich
DGK	Veröffentlichungen der Deutschen Geodätischen Kommission bei der Bayerischen Akademie der Wissenschaften, München
EOS	Transactions. American Geophysical Union, Washington, D.C.
ION-GPS	Proceedings of the Internat. Technical Meeting of the Satellite Division of The Institute of Navigation
J. Geod.	Journal of Geodesy, Springer, Berlin-Heidelberg-New York
JGR	Journal of Geophysical Research, American Geophysical Union, Washington D.C.
Man. geod.	manuscripta geodaetica, Springer, Berlin
Navigation	Journal of the Institute of Navigation
OSU Rep.	Reports of the Department of Geodetic Science, The Ohio State University, Columbus Ohio
SAO Spec. Rep.	SAO Special Reports, Smithsonian Institution, Astrophysical Observatory, Cambridge, Mass.
Wiss. Arb. Univ. Hannover	Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover
ZfV	Zeitschrift für Vermessungswesen. K. Wittwer, Stuttgart

Aardoom, L., Gelder, B. V., Vermaat, E. (1982): *Aspects of the analysis and utilization of satellite laser ranging at the Kootwijk observatory*. In: Feestbundel ter gelegenheid van de 65ste verjaardag van Professor Baarda, Delft, Vol. 2: 276–317.

Abidin, H., Wells, D. (1990): *Extra-wide-laning for "on the fly" ambiguity resolution: simulation of ionospheric effects*. Proc. Second Int. Symp. Precise Positioning with the Global Positioning System, GPS'90, Ottawa: 1217–1233.

Ádám, J. (1999): *Spaceborne VLBI beyond 2000*. In: Quo vadis geodesia...? Techn. Rep. Dep. Geod. GeoInform. Univ. Stuttgart, 1999.6: 1-3.

Ádám, J., Augath, W., Boucher, C., et al. (2000): *The European Reference System coming of age*. In: Schwarz (ed.) (2000): 47–54.

Agnieray, P. (1997): *The DORIS system performances and evolutions*. CSTG Bull. No. 14: 20–27.

Aguirre-Martinez, M., Cesare, S. (1999): *GOCE mission concept, error derivation and performances*. Boll. Geof. Teor. Appl. 40: 295–302.

Akos, D., Luo, M., Pullen, S., Enge, P. (2001): *Ultra-Wideband and GPS: Can They Co-exist?* GPS World 12 (9): 59–70.

Allan, D. (1987): *Time and Frequency (Time–Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators*. IEEE Transactions on Ultrasonic, Ferroelectrics, and Frequency Control, Vol. UFFC-34, No. 6: 647–654.

Allan, D., Ashby, N., Hodge, C. (1997): *The Science of Time Keeping*. Hewlett Packard Application Note 1289.

Altamimi, Z., Angermann, D., Argus, D., 22 more authors (2001): *The Terrestrial Reference Frame and the Dynamic Earth*. EOS 82, No.25.

Ambrosius, B., Noomen, R., Overgaanw, B., Wakker, K. (1991): *Geodetic coordinates solutions from SLR and GPS measurements*. IAG Symp. on Permanent Satellite Tracking Networks for Geodesy and Geodynamics, XX. General Assembly IUGG, Vienna.

Anderle, R. (1973): *Determination of polar motion from satellite observations*. Geophysical Surveys 1: 147–161.

Anderle, R. (1974): *Transformation of terrestral survey data to Doppler satellite datum*. JGR 79: 5319–5331.

Anderle, R. (1979): *Geodetic applications of the NAVSTAR Global Positioning System*. Proc. Int. Symp. Use of Artificial Satellites in Geodesy and Geodynamics II, 117ff, Athens.

Anderle, R. (1986): *Doppler satellite measurements and their interpretation*. In: A. J. Anderson, A. Cazenave (eds.), Space geodesy and geodynamics, Academic Press London.

Anderle, R., Malyevac, C. (1983): *Current plate motions based on Doppler satellite observations*. Geophysical Research Letters 10 (1): 67–90.

Angermann, D., Müller, H., Gerstl, M., Kelm, R., Seemüller, W., Vei, M. (2001): *Laserentfernungsmessungen zu LAGEOS-1 und LAGEOS-2 und ihr Beitrag zu globalen Referenzsystemen*. ZfV 126: 250–256.

Anodina, T., Prilepin, M. (1989): *The GLONASS System*. Proc. 5th Intern. Geod. Symp. Satellite Positioning, Las Cruzes, Vol. 1: 13–18,.

Aoki, S., Guinot, B., Kaplan, G., Kinoshita, H., McCarthy, D., Seidelmann, P. (1982): *The new definition of universal time*. Astron.Astrophys. 105: 359–361.

Argus, D., Gordon, R. (1991): *No-net rotation model of current plate velocities incorporating plate motion model NUVEL-1*. Geophys.Res.Lett. 18: 2039–2042.

Arnold, K. (1970): *Methoden der Satellitengeodäsie*. Akademie-Verlag Berlin.

Ashby, N., Spilker, J. (1996): *Introduction to Relativistic Effects on the Global Positioning System*. In: Parkinson, Spilker (eds.) (1996), Vol. 1, Chap. 18: 623–697.

Ashford, A. (2001): *Star Catalogues for the 21st Century*. Sky & Telescope, July: 65–67.

Ashjaee, J. (1989): *Precision survey with Ashtech XII, the all-in-one, all-in-view*. Proc. 5th Intern. Geod. Symp. Satellite Positioning, Las Cruzes, Vol. 1: 316–329.

Ashjaee, J., Lorenz, R. (1992): *Precision GPS surveying after Y-code*. Proc. ION GPS-92, Albuquerque: 657–659.

Augath, W. (1988): *Experiences with Trimble receivers in the control network of the F.G.R.*. In: Groten, Strauss (eds.) (1988): 131–143.

Ballani, L. (1982): *Die Bestimmung geodätisch-geodynamischer Parameter mit Hilfe von Laserdinstanzmessungen zum Mond*. Gerl. Beitr. Geophysik 91, 2: 121–140.

Balmino, G. (1973): *Analytical expressions for earth tide perturbations on close earth satellites*. Proc.Int.Symp. Use of Artificial Satellites for Geodesy and Geodynamics, Athens: 313–322.

Balmino, G. (1983): *The surface gravity data available for improvement of the global knowledge of the geopotential*. Proc. IAG Symp. Gen. Assembly Hamburg 1983, Columbus Ohio, Vol. 1: 365–387.

Balmino, G., Bernard, A. (1986): *Satellite gravity gradiometry for the determination of the geopotential*. Proc. ESA Workshop Solid Earth Science & Application Mission for Europe (SESAME), ESA SP-1080: 95–101.

Balmino, G., Perosanz, F., Rummel, R., Sneeuw, N., Sünkel, H. (1999): *CHAMP, GRACE and GOCE: mission concepts and simulations*. Boll. Geof. Teor. Appl. 40: 309–319.

Balmino, G., Reigber, C., Moynot, B. (1976a): *A geopotential model determined from recent satellite observing campaigns (GRIM1)*. Man. geod. 1.

Balmino, G., Reigber, C., Moynot, B. (1976b): *The GRIM2 earth gravity field model*. DGK, A 86, München.

Balmino, G., et al. (1984): *Le projet Gradio et la determination à haute resolution du geopotentiel*. Bull. Géod., 58 (2): 151–179.

Bamler, R. (1997): *Digital Terrain Models from Radar Interferometry*. Photogrammetrische Woche Stuttgart 1997: 93–105.

Bamler, R. (1998): *Synthetic aperture radar interferometry*. Inverse Problems, 14, R1–R54, IOP Publishing Ltd.

Bamler, R. (1999): *The SRTM Mission: A World-Wide 30 m Resolution DEM from SAR Interferometry in 11 Days*. Photogrammetrische Woche Stuttgart 1999: 145–154.

Bamler, R. (2000): *Principles of Synthetic Aperture Radar*. Surveys in Geophysics 21: 147–157.

Barrett, J. (1997): *Inventorying interstate pipelines with GPS*. GPS World 8 (10): 40–43.

Bastian, U. (1998): *The small astrometric interferometry satellite DIVA*. In: Brosche, P. (ed.): The Message of the Angles - Astrometry from 1798 to 1998, 207-212, Verlag H. Deutsch, Thun und Frankfurt.

Bastos, M., Landau, H. (1988): *Fixing cycle slips in dual-frequency kinematic GPS applications using Kahman filtering*. Man. geod. 13 (4): 249–256.

Bate, R., Mueller, D., White, J. (1971): *Fundamentals of Astrodynamics*. Dover Publications, New York.

Battin, R., Croopnick, S., Lenox, E. (1978): *The epoch state navigation filter*. J. of Guidance and Control, 1 (5): 359–364.

Bauch, A., Fischer, B., Heindorff, P., Hetzel, G., Petit, G., Schröder, R., Wolf, P. (2000): *Comparison of the PTB primary clocks with TAI*. Metrologia 37: 683–692.

Bedrich, S. (1998): *Hochgenaue satellitengestützte Zeitübertragung mit PRARE*. Scientific Technical Report STR98/24, GeoForschungsZentrum Potsdam.

Benedicto, J., Michel, P., Ventura-Traveset, J., Maier, T. (1999): *EGNOS, European Geostationary Navigation Overlay Service–Basic Technical Background*. Rep. Geod. 44, (3): 267–276.

Benning, W., Aussems, T. (1998): *Mobile mapping – digitale Datenerfassung mittels CDSS und automatisierte Auswertung von Videosequenzen*. ZfV 123: 202–209.

Benveniste, J., Roca, M., Levrini, G., 5 more authors (2001): *The Radar Altimeter Mission: RA-2, MWR, DORIS and LRR*. ESA Bull. 106: 67–76.

Berbert, J., Carney, D. (1979): *Intercomparison of GEOS-3 tracking system*. JGR 84, B8: 3933–3943.

Berroth, A., Hofmann, W. (1960): *Kosmische Geodäsie*. Verlag G. Braun, Karlsruhe.

Bertiger, W., Bar-Sever, Y., Haines, B., 11 more authors (1999): *A Prototype Real-Time Wide Area Differential GPS System*. Navigation, 44 (4): 433–447.

Beutler, G., Bauersima, I., Gurtner, W., Rothacker, M., Schildknecht, T. (1988): *Static positioning with the Global Positioning System (GPS): state of the art*. In: Groten, Strauss (eds.) (1988): 363–380.

Beutler, G., Eanes, R., Engeln, J., Herring, T., 6 more authors (1997): *SLR Review Committee Report*. International Laser Ranging Service.

Beutler, G., Gurtner, W., Bauersima, I., Rothacker, M. (1987): *Correlation between simultaneous GPS double difference carrier phase observations in the multistation mode. Implementation, considerations and first experiences*. Man. geod. 12: 40–44.

Beutler, G., Gurtner, W., Rothacher, M., Wild, U., Frei, E. (1990): *Relative static positioning with the Global Positioning System: basic technical considerations*. In: Bock, Leppard (eds.) (1990): 1–23.

Beutler, G., Hein, G., Melbourne, W., Seeber, G. (eds.) (1996): *GPS trends in precise terrestrial, airborne, and spaceborne applications*. IAG Symp. Proceed. 115, Springer-Verlag, Berlin-Heidelberg-New York.

Beutler, G., Weber, R., Hugentobler, U., Rothacher, M., Verdun, A. (1998): *GPS Satellite Orbits*. In: Teunissen, Kleusberg (1998), Chap. 2: 43–109.

Bianco, G., Selden, M., Bieneman, M., 9 more authors (1999): *Two-Color Ranging Upgrade for the MLRO System*. In: Schlüter, Schreiber, Dassing (1999): 346–352.

BIH (1988): *Annual Report*. Bureau International de l' Heure (until 1988), Paris.

BIPM (2002): *Annual Report*. Bureau International des Poids et Mesures, Sèvres. Published yearly.

Bisnath, S., Beran, T., Langley, R. (2002): *Precise Platform Positioning with a Single GPS Receiver*. GPS World 13 (4): 42–49.

Bisnath, S., Kim, D., Langley, R. (2001): *A New Approach to an Old Problem: Carrier-Phase Cycle Slips*. GPS World 12 (5): 46–51.

Black, H. (1978): *An easily implemented algorithm for the tropospehric range correction*. JGR, 83, B4.

Black, H. (1980): *The TRANSIT system 1977: Performance, plans, and potential*. Phil. Trans. Royal Soc. London, A 294: 217–236.

Blewitt, G. (1998): *IGS Densification Program*. In: IGS Annual Report 1997: 24–25.

Blewitt, G., et al. (1988): *GPS geodesy with centimeter accuracy*. In: Groten, Strauss (eds.) (1988): 30–40.

Bock, Y., Leppard, N. (eds.) (1990): *Global Positioning System. An Overview*. Symposium No. 102, Edinburgh, August 7-8, 1989. International Association of Geodesy. Symposia, Springer-Verlag, New York.

Bock, Y., Wdowinski, S., Fang, P., 18 more authors (1997): *Southern California Permanent GPS Geodetic Array*. JGR 102, B8: 18013–18033.

Bock, Y., Williams, S. (1997): *Integrated Satellite Interferometry in Southern California*. EOS 78 (29): 293, 299–300.

Bock, Y., et al. (1986): *Interferometric analysis of GPS phase observations*. Man. geod. 11 (4): 282–288.

Böder, V. (2002): *Zur hochpräzisen GPS Positions- und Lagebestimmung unter besonderer Berücksichtigung mariner Anwendungen*. Wiss. Arb. Univ. Hannover, Nr. 245.

Böder, V., Menge, F., Seeber, G., Wübbena, G., Schmitz, M. (2001): *How to deal with station-dependend errors - New developments of the absolute field calibration of PCV and phase-multipath*. ION-GPS 2001.

Böhler, W. (1972): *Über die Satellitenbeobachtungen des deutschen Meßtrupps im Weltnetz US C&GS*. DGK C 182, München.

Bossler, J., Goad, C., Bender, L. (1980): *Using the Global Positioning System (GPS) for geodetic positioning*. Bull. Géod. 54 (4): 553–563.

Boucher, C., Altamini, Z. (1989): *The initial IERS Terrestrial Reference Frame*. IERS Technical Note 1, Paris 1989.

Boucher, C., Altamini, Z., Sillard, P. (1999): *The 1997 International Terrestrial Reference Frame (ITRF97)*. IERS Technical Note 27.

Boucher, C., Feissel, M. (1984): *Realization of the BIH Terrestrial System*. Int. Symp. on Space Techniques for Geodynamics, Sopron, Vol. 1: 235–254.

Boucher, C., Feissel, M., Lestrade, J. (1988): *Concepts and methods of the Central Bureau of the International Earth Rotation Service*. Bull. Geod. 62 (4): 511–519.

Boucher, C., Paquet, P., Wilson, P. (1979): *The second European Doppler Observation Campaign (EDOC-2)*. Proc. 2nd Intern. Geodetic Symp. Satellite Doppler Positioning, Austin: 819–850.

Boucher, C., Wilkins, G. (eds.) (1990): *Earth rotation and coordinate reference frames.*. IAG Symposium No. 105, Edinburgh, August 10-11, 1989. Springer-Verlag, New York-Berlin.

Braasch, M. (1996): *Multipath Effects*. In: Parkinson, Spilker (eds.) (1996), Vol. 1, Chap. 14: 547–568.

Bretterbauer, K. (1997): *Eine Renaissance der Astrometrie in der Geodäsie*. Österr. Z. Vermessungswesen und Geoinformatik 1/97: 15–20.

Bretterbauer, K. (2001): *Die parallaktische Refraktion*. ZfV 126: 211–213.

Breuer, B., Campbell, J., Müller, A. (1993): *GPS Meß- und Auswerteverfahren unter operationellen GPS-Bedingungen*. J. Satellite Based Positioning, Navigation amd Communication (SPN), 2 (3): 82–90.

Brouwer, D. (1959): *Solution of the problem of artificial satellite theory without drag*. Astronomical Journal 64: 378–397.

Brouwer, D., Clemence, G. (1961): *Methods of Celestial Meachanics*. Academic Press, New York, London.

Brown, A. (1990): *A multi-sensor approach to assuring GPS integrity*. GPS World 1 (2): 44–48.

Brown, D. (1970): *Near term prospects for potential accuracies of 0.1 to 1.0 meters from satellite geodesy*. U.S. Airforce Cambridge Research Laboratories. Report AFCRL-70-0501, Bedford, Mass.

Brunner, F. (1992): *Refraction, refractive index, and dispersion: some relevant history*. In: DeMunck, Spoelstra (eds.) (1992): 3–9.

Brunner, F. (ed.) (1998): *Advances in Positioning and Reference Frames*. IAG Symp. Proceed. 118, Springer-Verlag, Berlin-Heidelberg-New York.

Bruns, H. (1878): *Die Figur der Erde*. Publ. Königl. Preuß. Geod. Inst., Berlin.

Bürki, B., Kahle, H.-G. (1995): *Mikrowellen-Wasserdampfradiometrie für hochgenaue Messungen mit GPS*. Vermessung, Photogrammetrie und Kulturtechnik, 1995: 232–240.

Campbell, J. (1979): *Die Radiointerferometrie auf langen Basen als geodätisches Meßprinzip hoher Genauigkeit*. DGK C 254, München.

544   Bibliography

Campbell, J. (2000a): *From Quasars to Benchmarks: VLBI Links Heaven and Earth*. In: Vandenberg, Baver (eds.) (2000): 19–34.

Campbell, J. (2000b): *Very Long Baseline Interferometry - a high precision tool for geodesy and astrometry*. C. R. Acad. Sci. Paris, t.1, Série IV: 1255–1265.

Campbell, J., Cloppenburg, H., Lohmar, F.-J. (1984): *Estimating the ionospheric refraction effect on interferometric GPS-measurements*. Proc. Int. Symp. Space Techniques for Geodyn., Sopron, Vol. 2: 196–206.

Campbell, J., Haas, R., Nothnagel, A. (2002): *Measurement of Vertical Crustal Motion in Europe by VLBI*. Scientific Report European Commission, TRM Network FMRX-CT96-0071, Geodetic Institute, University of Bonn.

Campbell, J., Nothnagel, A. (1998): *Erdkrustenbewegungen und Bezugssysteme - neuere Ergebnisse aus VLBI-Messungen*. ZfV 123: 42–48.

Campbell, J., Nothnagel, A., Schuh, H. (1992): *VLBI-Messungen für geodynamische Fragestellungen*. ZfV 117: 214–227.

Campbell, J., Seeber, G., Witte, B. (1973): *Kombination von Doppler-, Laser- und photographischen Beobachtungen in Satellitennetzen*. DGK B 200, München.

Campbell, J., Witte, B. (1978): *Grundlagen und geodätische Anwendung der Very Long Baseline Interferometry (VLBI)*. ZfV 103: 10–20.

Campos, M. (1987): *Controle da rede geodésica Brasileira por meio de satélites do sistema NNSS*. Ph.D. thesis. Curitiba.

Campos, M., Seeber, G., Wübbena, G. (1989): *Positioning with GPS in Brazil*. Proc. 5th Intern. Geod. Symp. Satellite Positioning, Las Cruces, Vol. 1: 526–535.

Capitaine, N. (2002): *The Essential Contribution of VLBI to Fundamental Astronomy*. In: IVS General Meeting Proceedings (2002): 14–23.

Capitaine, N., et al. (2002): *Proceedings of the IERS Workshop on the Implementation of the New IAU Resolutions*. IERS Technical Note No. 29, Frankfurt am Main.

Capitaine, N., Guinot, B., McCarthy, D. (2000): *Definition of the Celestial Ephemeris Origin and of UT1 in the International Celestial Reference Frame*. Astron. Astrophys. 355: 398–405.

Cappelari, J., Velez, C., Fuchs, A. (1976): *Mathematical Theory of the Goddard Trajectory Determination System*. GSFC Document X-582-76-77, Greenbelt.

Cazenave, A., Dominh, K., Soudarin, F. P. L., Cretaux, J., Provost, C. L. (1999): *Sea level changes from Topex-Poseidon altimetry and tide gauges, and vertical crustal motions from DORIS*. Geophys. Res. Letters, 26 (14): 2077–2080.

Cazenave, A., Royer, J. (2001): *Applications to Marine Geophysics*. In: Fu, Cazenave (eds.) (2001), Chap. 11: 407–439.

Chadwell, D., Spiess, F., Hildebrand, J., Young, L., Purcell, G., Dragert, H. (1998): *Deep-Sea Geodesy: Monitoring the Ocean Floor*. GPS World, 9 (9): 44–55.

Chelton, D., Ries, J. C., Haines, B., Fu, L., Callahan, P. (2001): *Satellite Altimetry*. In: Fu, Cazenave (eds.) (2001), Chap. 1: 1–131.

Chen, D., Lachapelle, G. (1995): *A comparison of the FASF and least-squares search algorithms for on-the-fly ambiguity resolution*. Navigation 42 (2): 371–390.

Cheney, R., Douglas, B., Green, R., Miller, L., Milbert, D., Porter, D. (1986): *The GEOSAT altimeter mission: a milestone in satellite oceanography*. EOS 67, Dec. 2.

Cheney, R., Douglas, B., Sandwell, D., Marsh, J., Martin, T., McCarthy, J. (1984): *Applications of satellite altimetry to oceanography and geophysics*. Marine Geophysical Researches 7: 17–32.

Chistyakov, V., Filatchenkov, S., Khimulin, V., Kornyenko, V. (1996): *Double Duty: Russia's DGPS/DGLONASS Maritime Service*. GPS World 7 (3): 59–62.

Chobotov, V. (1991): *Orbital Mechanics*. American Institute of Aeronautics and Astronautics, Washington D.C.

Chodota, M. (1987): *Present status of the African Doppler Survey Project - ADOS -*. CSTG Bull. No. 9, München.

Chodota, M. (1990): *Post ADOS strategy for unified networks for Africa*. Proc. 4th Symp. on Geodesy in Africa, Tunis.

CNES (1975): *STARLETTE* . Centre National d'Etudes Spatiales, Toulouse.

CNES (2000): *Proceedings DORIS DAYS*. Centre National d'Études Spatiales.

Cocard, M. (1995): *High Precision GPS Processing in Kinematic Mode*. Geod. Geophys. Arb. Schweiz, Vol. 52, Zürich.

Coco, D. (1991): *GPS-satellites of opportunity for ionospheric monitoring*. GPS World, 2 (9): 47–50.

Cohen, C. (1996): *Attitude Determination*. In: Parkinson, Spilker (eds.) (1996) Vol. 2, Chap.19: 519–538.

Cohen, S., Chinn, D., Dunn, P. (1990): *Geoscience laser ranging system (GLRS): characteristics and expected performance in geodynamic applications*. In: Vyskocil, Reigber, Cross (eds.) (1990): 56–63.

Collins, P., Langley, R. (1999): *Tropospheric Delay prediction for the WAAS User*. GPS World 10 (7): 52–58.

Colombo, O. (1984): *The global mapping of gravity with two satellites*. Netherlands Geodetic Commission, New Series 7, Delft.

Colombo, O. (1989): *Advanced techniques for high-resolution mapping of the gravitational field*. In: Sansò, Rummel (eds.) (1989): 335–372.

Colombo, O., Bhapkar, U., Evens, A. (1999): *Inertial-aided cycle-slip detection/correction for precise, long-baseline kinematic GPS*. ION GPS-99: 1915–1921.

Costes, M. (1997): *DORIS system: status and perspectives*. CSTG Bull. 13: 64–66.

Counselman, C., Shapiro, I. (1979): *Miniature interferometer terminals for earth surveying*. Bull. Géod. 53 (2): 139–163.

Counselman, C., Steinbrecher, D. (1982): *The Macrometer: A compact radio interferometry terminal for geodesy with GPS*. Proc. 3rd Int. Symp. on Satellite Doppler Positioning, Las Cruzes, Vol. 2: 1165–1172.

Cracknell, A., Hayes, L. (1991): *Introduction to remote sensing*. Taylor & Francis, London, New York.

Cramer, M. (2001): *Genauigkeitsuntersuchungen zur GPS/INS-Integration in der Aerophotogrammetrie*. DGK C 537, München.

Cross, P., Ahmad, N. (1988): *Field validation of GPS phase measurements*. In: Groten, Strauss (eds.) (1988): 349–360.

Cruddace, P. (2001): *Ordnance Survey's National GPS*. GEO Informatics, July/August: 12–13.

Cugusi, L., Proverbio, E. (1978): *Relativistic effects on the motion of Earth's artificial satellites*. Astron. Astrophys. 69: 321–325.

Cui, C., Mareyen, M. (1992): *Gauss's equations of motion in terms of Hill variables and first applications to numerical integration of satellite orbits*. Man. geod. 17: 155–163.

Dach, R. (1999): *Einfluss von Auflasteffekten auf präzise GPS-Messungen*. DGK C 519, München.

Daly, P., Misra, P. (1996): *GPS and Global Navigation Satellite System (GLONASS)*. In: Parkinson, Spilker (eds.) (1996) Vol. II, Chap. 9: 243–274.

Davies, K. (1990): *Ionospheric Radio*. IEEE Electromagnetic Waves Series 31. Peter Peregrinus, London.

Davis, J., Herring, T., Shapiro, I., Rogers, A., Elgered, G. (1985): *Geodesy by radio interferometry: Effects of atmospheric modelling errors on estimates of baseline length*. Radio Science 20 (6): 1593–1607.

Degnan, J. (2000): *An Overview of SLR2000 Engineering Progress and Potential Future Upgrades*. 12th Int. Laser Workshop on Laser Ranging, Matera, Italy.

Degnan, J., Pavlis, E. (1994): *Laser Ranging to GPS Satellites with Centimeter Accuracy*. GPS World 5 (9): 62–70.

Delikaraoglou, D. (1989): *On principles, methods, and recent advances in studies towards a GPS-based control system for geodesy and geodynamics*. NASA Techn. Memo. 100716. Greenbelt, Md.

DeMets, C., Gordon, R., Argus, D., Stein, S. (1994): *Effect of recent revisions to the geomagnetic reversal time scale on estimates of current plate motions*. Geophys. Res. Lett. 21: 2191–2194.

Demmel, M. (2000): *Satellitenortung in der Landwirtschaft – Möglichkeiten und Anforderungen*. Proc. 3. SAPOS Symp. München: 38–52.

DeMunck, J., Spoelstra, T. A. (eds.) (1992): *Refraction of Transatmospheric Signals in Geodesy*. Symp. Proceed., Publication on Geodesy No. 36, Netherlands Geodetic Commission, Delft.

Denker, H., Torge, W. (1998): *The European gravimetric quasigeoid EGG97 – and IAG supported continental enterprise*. In: Forsberg, Feissel, Dietrich (eds.) (1998): 249–254.

Desai, S., Wahr, J. (1995): *Empirical ocean tide models estimated from TOPEX/POSEIDON altimetry*. JGR 100 (C12): 25 205–25 228.

Devoti, R., Luceri, V., Sciarretta, C., Bianco, G., Donato, G. D., Vermeersen, L., Sabadini, R. (2001): *The SLR secular gravity variations and their impact on the interference of mantle rheology and lithospheric thickness*. Geophys. Res. Letters, 28 (5): 855–858.

DGFI (2001): *Deutsches Geodätisches Forschungsinstitut, Annual Report 2000/2001*. DGFI, München.

Dickey, J., Bender, P., Faller, J., Newhall, X., 8 more authors (1994): *Lunar Laser Ranging: A continuing Legacy of the Apollo Program*. Science, Vol. 265: 482–490.

Dickey, J., Williams, J., Newhall, X., Yoder, C. (1983): *Geophysical applications of Lunar Laser Ranging*. Proc. IAG Symp. Gen. Ass. Hamburg, Vol. 2: 509–521, Columbus, Ohio.

Dietrich, R., Dach, R., Engelhardt, G., et al., (2001): *ITRF coordinates and plate velocities from repeated GPS campaigns in Antarctica – an analysis based on different individual solutions*. Journal of Geodesy, 74, No. 11/12.

Dietrich, R. (ed.) (2000): *Deutsche Beiträge zu GPS-Kampagnen des Scientific Committee on Antarctic Research (SCAR) 1995–1998* . DGK B 310, München.

DOD (2001): *Global Positioning System Standard Positioning Service Performance Standard*. Department of Defense, Document, October 2001.

DOD/DOT (2001a): *Federal Radio Navigation Plan*. Department of Defense (DoD), Department of Transportation (DOT).

DOD/DOT (2001b): *Federal Radio Navigation Systems*. Department of Defense (DoD), Department of Transportation (DOT).

Douglas, B., Agreen, R. (1983): *The sea state correction for GEOS–3 and SEASAT satellite altimeter data*. JGR 88 (C3): 1655–1661.

Douglas, B., Cheney, R. (1990): *GEOSAT: Beginning a new era in satellite oceanography*. JGR 95 (C3), 2833–2836.

Douglas, B., Goad, C., Morrison, F. (1980): *Determination of the geopotential from satellite-to-satellite tracking data*. NOAA Techn. Memo. NOS NGS 24, Rockville.

Dow, J. (1988): *Ocean Tides and Plate Motions from LAGEOS*. DGK C 317, München.

Drew, A. (1983): *Glacial movements in Greenland from Doppler satellite observations*. Inst. of Polar Studies, Ohio State University Report No. 81, Columbus.

Drewes, H. (1998): *Time Evolution of the SIRGAS Reference Frame*. In: Brunner (ed.) (1998): 174–179.

Drewes, H. (1999): *Realisierung des kinematischen terrestrischen Referenzsystems ohne globale Rotation ("no net rotation")*. Mitt. Bundesamt f. Kart. u. Geod. 5, 120–125, Frankfurt a. M.

Drewes, H. (2000): *The Role of VLBI Among the Geodetic Space Techniques Within CSTG*. In: Vandenberg, Baver (eds.) (2000): 35–41.

Drewes, H., Angermann, D. (2001): *The Actual Plate Kinematic and Crustal Deformation Model 2000 (APKIM2000) as a Geodetic Reference System*. IAG 2001 Scientific Assembly, Budapest.

Drewes, H., Beutler, G., Rummel, R. (2002): *Challenges for VLBI Within an Integrated Global Geodetic Observing System*. Proceed. IVS 2002 General Meeting: 24–32.

Drewes, H., Reigber, C. (1982): *Untersuchungen zu einem satellitengetragenen Laser-Entfernungsmeßsystem zur präzisen Punktbestimmung*. ZfV 107: 396–405.

Eanes, R., Bettadpur, S. (1996): *The CSR3.0 global ocean tidal model: Diurnal and semidiurnal ocean tides from TOPEX/POSEIDON altimetry*. Techn. Rep. CSR-TM-96-05, University of Texas, Center for Space Research.

Eanes, R., Schutz, B., Tapley, B. (1983): *Earth and Ocean Tide Effects on Lageos and Starlette*. Proc. 9th Int. Symp. on Earth Tides: 239–249.

Egge, D. (1982): *Investigations on the effect of small antenna movements in Transit Doppler Positioning*. Proc. 3rd Int. Geod. Symp. Satellite Doppler Positioning, Las Cruces, Vol. 2: 953–968.

Egge, D. (1985): *Zur sequentiellen Auswertung von Doppler-Satellitenbeobachtungen*. Wiss. Arb. Univ. Hannover, No. 93.

Egge, D., Seeber, G. (1979): *Meßverfahren zur genauen Positionsbestimmung im Meeresbereich*. Wiss. Arb. Univ. Hannover, No.93.

Ehlert, D. (1991): *Differentielle Verschiebungen und Drehungen in dreidimensionalen Koordinatensystemen*. DGK B 295, München.

Eichhorn, H. (1974): *Astronomy of Star Positions*. F. Ungar Publ.Co., New York.

Eisfeller, B. (2002): *Das Europäische Satellitennavigationssystem GALILEO*. 4. SAPOS Symp. Hannover: 214–226.

Eisner, A., et al. (1982): *NOVA-1: The newest Transit satellite (a status report)*. Proc. 3rd Int. Geod. Symp. on Doppler Satellite Positioning, Las Cruces, Vol. 2: 843–862.

El-Sheimy, N. (2000): *Mobile multi-sensor system: The new trend in mapping and GIS applications*. In: Schwarz (ed.) (2000): 319–330.

Elgered, G., Jarlemark, P. (1998): *Ground-based microwave radiometry and long-term observations of atmospheric water vapor*. Radio Science, 33: 707–717.

Engelis, T. (1987): *Radial orbit error reduction and sea surface topography determination using satellite altimetry*. OSU Rep. 377.

Escobal, P. (1965): *Methods of orbit determination*. John Wiley, New York, London, Sydney.

Fagard, H., Orsoni, A. (2000): *Current status and evolution prospects of the DORIS network*. Proceed. DORIS days 2000.

Fashir, H., Abdalla, K. (1991): *Doppler geoid in Africa*. Survey Review 31: 175–180.

Fedrizzi, M., de Paula, E., Kantor, I., Langley, R., Santos, M., Komjathy, A. (2001): *The Low-Latitude Ionosphere: Monitoring its Behavior with GPS*. Proc. ION GPS 2001, 2468–2475, also: GPS World, 13 (2): 41–47.

Feissel, M. (1986): *Observational results on Earth rotation and reference systems*. Report on MERIT-COTES, Part. III. Bureau International de l'Heure, Paris.

Fejes, I. (1994): *Techniques and Methods in Geodynamical Research: Space VLBI*. Acta Geod. Geoph. Hung. 29: 443–456.

Fejes, I., Mihály, S. (1991): *Application of Space VLBI to satellite dynamics*. Adv. Space Res. 11: 429–437.

Fell, P. (1980): *Geodetic Positioning using a Global Positioning System*. OSU Rep. 299.

Flechtner, F. (2000): *Bestimmung des Gesamtelektroneninhalts der Ionosphäre aus PRARE-Entfernungs- und Dopplerbeobachtungen*. Scientific Technical Report STR00/02, Geo-ForschungsZentrum Potsdam.

Fliegel, H., Gallini, T. (1989): *Radiaton pressure models for Block II GPS satellites*. Proceed. 5th Int. Geod. Symp. Sat. Positioning, Las Cruces: 789–798.

Fliegel, H., Gallini, T., Swift, E. (1992): *Global Positioning System Radiation Force Model for Geodetic Applications*. JGR, 97 (B1): 559–568.

Fontana, R., Cheung, W., Stansell, T. (2001): *The Modernized L2 Civil Signal*. GPS World 12 (9): 28–32.

Forrest, W. (2002): *GALILEO*. Proc. 39th CGSIC Conference, Springfield, Virginia, 18th April 2002.

Forsberg, R., Feissel, M., Dietrich, R. (eds.) (1998): *Geodesy on the Move: Gravity, Geoid, Geodynamics, and Antarctica* . IAG Symp. Proceed. 119, Springer-Verlag, Berlin-Heidelberg-New York.

Forsell, B. (1991): *Radionavigation Systems. Prentice Hall, New York*.

Fortes, L., Luz, R., Pereira, K., Costa, S., Blitzkow, D. (1997): *The Brazilian Network for Continuous Monitoring of GPS (RBMC): Operation and Products*. In: Brunner (ed.) (1998): 73–78.

Forward, R. (1974): *Review of artificial satellite gravity gradiometer techniques for geodesy*. Proc. Symp. on The Use of Artificial Satellites for Geodesy and Geodynamics, Athens: 157–192.

Fosu, C. (1998): *Astrogeodetic Levelling by the Combination of GPS and CCD Zenith Camera*. Studiengang Vermessungswesen, Schriftenreihe UniBw München 62.

Frei, E., Beutler, G. (1989): *Some considerations concerning an adaptive optimized technique to resolve the initial phase ambiguities for static and kinematik GPS surveying techniques.* Proceed. 5th Int. Symp. Satellite Positioning, Las Cruces, Vol. 2: 671–686.

Frei, E., Beutler, G. (1990): *Rapid static positioning based on the fast ambiguity resolution approach: the alternative to kinematic positioning.* Proceed. GPS '90, Ottawa: 1196–1216.

Fricke, W., Schwan, H., Lederle, T. (1988): *Fifth Fundamental Catalogue (FK5), Part I: The Basic Fundamental Stars.* Astronomisches Recheninstitut Heidelberg, Veröffentlichungen, Nr. 32. Verlag G. Braun, Karlsruhe.

FRNP (2001): *Federal Radio Navigation Plan 2001.* Rep. No. DOT-VNTSC-RSPA-01-3/DOD-4650.5, Nat.Techn.Inf.Service, Springfield Virginia USA.

Fu, L., Cazenave, A. (eds.) (2001): *Satellite Altimetry and Earth Sciences.* Academic Press, San Diego-London.

Fu, L., Chelton, D. (2001): *Large-Scale Ocean Circulation.* In: Fu, Cazenave (eds.) (2001), Chap. 2: 133–169.

Gabriel, A., Goldstein, R. M., Zebker, H. A. (1989): *Mapping Small Elevation Changes Over Large Areas: Differential Radar Interferometry.* JGR 94 (B7): 9183–9191.

Gaignebet, J. (ed.) (1985): *5th Int. Workshop on Laser Ranging Instrumentation (Herstmonceux 1984).* Vol. 1 and Vol. 2, Bonn.

Ganin, A. (1995): *Differential GLONASS in Russia: The Ways of Development .* Proceed. ION GPS-95: 1049–1052.

Gao, Y., Zuofa, L. (1999): *Cycle Slip Detection and Ambiguity Resolution Algorithms for Dual-Frequency GPS Data Processing.* Marine Geodesy 22 (3): 169–181.

Gaposchkin, E. (1973): *Smithsonian Standard Earth (III).* SAO Spec. Rep. 353, Cambridge, Mass.

Gaposchkin, E., Kolaczek, B. (eds.) (1981): *Reference coordinate systems for Earth dynamics.* Proc. 56th Colloquium IAU (Warschau 1980). Astrophysics and Space Science Library Vol. 56, D. Reidel, Dordrecht-Boston-London.

Gelb, A. (ed.) (1974): *Applied Optimal Estimation.* M.I.T. Press, Cambridge Mass.

Gendt, G. (2000): *IGS Tropospheric Products.* IGS Annual Report 2000.

Gens, R. (1998): *Quality assessment of SAR data.* Wiss. Arb. Univ. Hannover, Nr. 226.

Georgiadou, Y., Kleusberg, A. (1988): *On carrier signal multipath effects in relative GPS positioning.* Man. geod. 13: 172–179.

Georgiadou, Y., Kleusberg, A. (1990): *Multipath effects in static and kinematic GPS surveying.* In: Bock, Leppard (eds.) (1990): 82–89.

Gerstbach, G. (1999): *CCD und Astro-Geodäsie. Unterwegs zur automatischen Lotrichtungsmessung.* Geowiss. Mitt. 50: 45–58.

Giacaglia, G. (1973): *Lunar perturbations on artificial satellites of the Earth.* SAO Special Report 352, Cambridge, Mass.

Gibbons, G. (1991): *GPS and friends: integrating technology.* GPS World 2 (5): 10.

Gibbons, G., Maynard, G. (1990): *Odysseys of geodesy.* GPS World, 1 (2): 20–28.

Goad, C. (1977): *Application of digital filtering to satellite geodesy.* NOAA Techn. Report NOS71 NGS 6, Rockville, Md.

Goad, C. (1985): *Precise relative position determination using Global Positioning System carrier phase measurements in a non-difference mode.* Proceed. 1st Int. Symp. Precise Positioning with GPS, Rockville, Vol. 1: 347–356.

Goad, C. (1998): *Short Distance GPS Models*. In: Teunissen, Kleusberg (1998), Chap. 11, 457–481.

Goad, C. C., Müller, A. (1988): *An automated procedure for generating an optimum set of independent double difference observables using GPS carrier phase measurement*. Man. geod. 13: 365–369.

Goffinet, P. (2000): *Qualitätssteigerung der Seevermessung und Navigation durch neuartige Beschickungsverfahren*. Wiss. Arb. Univ. Hannover Nr. 235.

Goldan, H. (1996): *Beiträge zur GPS gestützten Höhenbestimmung im Küstenbereich*. Wiss. Arb. Univ. Hannover Nr. 215.

Görres, B. (1996): *Bestimmung von Höhenänderungen in regionalen Netzen mit dem Global Positioning System*. DGK C 461, München.

Görres, B., Campbell, J. (1998): *Bestimmung vertikaler Punktbewegungen mit GPS*. ZfV 123: 222–230.

Green, R. (1985): *Spherical Astronomy*. Cambridge University Press, Cambridge.

Großmann, W. (1976): *Geodätische Rechnungen und Abbildungen in der Landesvermessung*. 3. Aufl., Wittwer, Stuttgart.

Groten, E. (1979, 1980): *Geodesy and the earth's gravity field*. Vol. I: Principles and conventional methods; Vol. II: Geodynamics and advanced methods. Dümmler, Bonn.

Groten, E., Mathes, A., Becker, M., Sauermann, K. (2001): *GNSS Information*. AVN 4/2001, p. 154.

Groten, E., Strauss, R. (eds.) (1988): *GPS-techniques applied to geodesy and surveying*. Lecture Notes in Earth Sciences 19, Springer-Verlag, Berlin-Heidelberg.

Gruber, T., Reigber, C., Schwintzer, P. (2000): *The 1999 pre-CHAMP high resolution gravity model*. In: Schwarz (ed.) (2000): 89–95.

Guier, W. (1962): *Navigation using artificial satellites-the TRANSIT System*. Proceed. 1st Int. Symp. Use of Art. Satellites for Geodesy, Washington D.C.: 261–277.

Guier, W., Weiffenbach, G. (1960): *A satellite Doppler navigation system*. Proceed. IRE: 48.

Guier, W., Weiffenbach, G. (1998): *Genesis of Satellite Navigation*. Johns Hopkins APL Technical Digest, Vol. 19, No. 1: 14–17.

Guinot, B. (1979): *Basic Problems in the Kinematics of the Rotation of the Earth*. In: McCarthy and Pilkington (eds.): Time and the Earth's Rotation, IAU Symp. No. 82: 7–16.

Guinot, B. (1995): *Scales of Time*. Metrologia 31: 431–440.

Gurtner, W. (1994): *RINEX: The Receiver Independent Exchange Format*. GPS World 5 (7): 48–52.

Gurtner, W. (2001): *RINEX: The Receiver Independent Exchange Format Version 2.10*. Available via Internet from IGS.

Gurtner, W., Beutler, G., Bauersima, I., Schildtknecht, T. (1985): *Evaluation of GPS carrier difference observations: the Bernese second generation software package*. Proc. 1st Int. Symp. Precise Positioning with GPS, Rockville, Vol. I: 363–372.

Gurtner, W., Mader, G., Macarthur, D. (1989): *A common exchange format for GPS data*. Proc. 5th Int. Geod. Symp. Satellite Positioning, Las Cruces, USA, Vol. 2: 920–931.

Habrich, H. (2000): *Geodetic Applications of the Global Navigation Satellite System (GLONASS) and of GLONASS/GPS Combinations*. Mitt. Bundesamt für Kartographie und Geodäsie, Bd. 15, Frankfurt a. M.

Hahn, J. (1999): *Aspekte der Uhrensynchronisation, Zeithaltung und -verteilung bei der Konzeption zukünftiger Satelliten-Navigationssysteme*. Forschungsbericht 1999-46, Deutsches Zentrum für Luft- und Raumfahrt, Oberpfaffenhofen.

Hamal, K., Prochazka, I. (1989): *Prospect of laser ranging using a semitrain versus a single pulse*. In: Veillet (ed.) (1989): 209–214.

Han, S., Rizos, C. (1997): *Comparing GPS ambiguity Resolution Techniques*. GPS World 8 (5): 54–61.

Hankemeier, P. (1996): *The DGPS Service for the FRG: Concept and Status*. In: Beutler, Hein, Melbourne, Seeber (eds.) (1996): 75–84.

Hartl, P., Thiel, K.-H., Wu, X., Xia, Y. (1995): *Earth Observation by Means of SAR: Present State and Future Possibilities* . Photogrammetrische Woche Stuttgart 1995: 127–137.

Hartmann, G., Leitinger, R. (1984): *Range errors due to ionospheric and tropospheric effects for signal frequencies above 100 MHz*. Bull. Geod. 58: 109–136.

Hatch, R. (1982a): *It's about time - TRANSIT time*. Proc. 3rd Int. Geod. Symp. Sat. Doppler Pos., Las Cruces, Vol. 2: 903–915.

Hatch, R. (1982b): *The synergism of GPS code and carrier measurements*. Proc. 3rd Int. Geod. Symp. on Doppler Positioning, Las Cruces, Vol. 2: 1213–1232.

Hatch, R. (1991): *Instantaneous ambiguity resolution*. In: Schwarz, Lachapelle (eds.) (1991): 299–308.

Hatch, R., Jung, J., Enge, P., Pervan, B. (2000): *Civilian GPS: The Benefits of Three Frequencies*. GPS Solutions 3 (4): 1–9.

Hatch, R., Keegan, R., Stansell, T. A. (1992): *Kinematic receiver technology from Magnavox*. Magnavox Electronic Systems Company, Torrance, CA.

Hay, C. (2000): *The GPS Accuracy Improvement Initiative*. GPS World 11 (6): 56–61.

Hay, C., Wong, J. (2000): *Tropospheric Delay Prediction at the Master Control Station*. GPS World, 11 (6): 56–62.

Heidland, C. (1994): *Satellitenaltimetrie über Eis*. Ber. Polarforsch. 141, Bremerhaven.

Hein, G. (1983): *Erdmessungen als Teil einer integrierten Geodäsie - Begründung Stand und Entwicklungstendenzen*. ZfV 108: 93–104.

Hein, G. (2000): *From GPS and GLONASS via EGNOS to GALILEO – Positioning and Navigation in the Third Millennium*. GPS Solutions 3 (4): 39–47.

Heiskanen, W. A., Moritz, H. (1967): *Physical Geodesy*. W.H. Freeman, San Francisco, London.

Helmert, F. R. (1880/1884): *Die mathematischen und physikalischen Theorien der höheren Geodäsie*. Teubner, Leipzig. Reprint Minerva GmbH, Frankfurt a.M. 1961.

Helwig, A., Offerans, G., van Willigen, D. (1997): *Eurofix DGPS Service Through the Sylt Loran-C Transmitter: Static and Dynamic Test Results*. Proc. 26th Annual Convention and Technical Symposium International Loran Association, Ottawa.

Henriksen, S. (1977): *National Geodetic Satellite Program-Introduction*. In: NGSP, Washington, Vol. I: 3–85.

Herring, T. (1992): *Modeling atmospheric delays in the analysis of space geodetic data*. In: DeMunck, Spoelstra (eds.) (1992): 157–164.

Hey, J. (1984): *Das Radiouniversum - Einführung in die Radioastronomie*. Weinheim 1974. The Radio Universe, 3. ed., Oxford.

Hieber, S. (1983): *Satellite systems for geodesy and geodynamics*. CSTG Bull. No. 6: 253–261.

Hilla, S. (2002): *Extending the Standard Product 3 (SP3) Orbit Format*. IGS Network, Data and Analysis Center Workshop, Ottawa 2002.

Hinze, H. (1990): *Zum Einsatz von Satelliten-Positionierungsverfahren für glaziologische Aufgaben in der Antarktis*. Wiss. Arb. Univ. Hannover, Nr. 163.

Hirt, C. (2001): *Automatic Determination of Vertical Deflections in Real-Time by Combining GPS and Digital Zenith Camera for Solving the "GPS-Height-Problem"*. Proc. ION GPS-2001, Salt Lake City.

Hocke, K., Tsuda, T. (2001): *Using GPS Satellites to Study Plasma Irregularities*. GPS World 12 (7): 34–36.

Hofmann-Wellenhof, B., Lichtenegger, H., Collins, J. (2001): *Global Positioning System: Theory and Practice*. 6th ed., Springer-Verlag, Wien-New York.

Hofmann-Wellenhof, B., Remondi, B. (1988): *The antenna exchange: one aspect of high precision GPS kinematic survey*. In: Groten, Strauß (eds.), 261–277, Berlin.

Hofton, M., Foulger, G. (1996): *Postrifting unelastic deformation around the spreading plate boundary , North Iceland*. JGR 101: 25403–25421.

Hog, E., Fabricius, C., others (2000): *Construction and Verification of the Tycho-2 Catalogue*. Astron. Astrophys. 357 (1): 367–386.

Hopfield, H. (1969): *Two-quartic tropospheric refractivity profile for correcting satellite data*. JGR 74 (18).

Hopfield, H. (1971): *Tropospheric effect on electromagnetically measured range: prediction from surface weather data*. Radio Science 6, (3): 357–367.

Hoyer, M. (1982): *Satellitendopplermessungen als unterstützende Beobachtungen bei der Kontrolle und Verbesserung eines geodätischen Netzes in Venezuela*. Wiss. Arb. Univ. Hannover, Nr. 111.

Hoyer, M., Arciniegas, S., Pereira, K., Fagard, H., Maturana, R., Torchetti, R., Drewes, H., Kumar, M., Seeber, G. (1998): *The definition and realization of the reference system in the SIRGAS project*. In: Brunner (ed.) (1998): 168–173.

Hugentobler, U. (1998): *Astrometry and Satellite Orbits: Theoretical Considerations and Typical Applications*. Geod. Geophys. Arb. Schweiz Vol. 57.

Hugentobler, U., Schaer, S., Fridez, P. (2001): *Bernese GPS Software Version 4.2*. University of Berne, Switzerland.

Husson, S. (1999): *SLR Global Performance Evaluation*. In: Schlüter, Schreiber, Dassing (1999), Vol. 2: 645–649.

ICD (1993): *Interface Control Document – Navstar GPS Space Segment/Navigation User Interfaces (ICD-GPS-200C)*. ARINC Research Corporation, Revision IRN-200C-004, 12 April 2000.

IERS (2001): *International Earth Rotation Service*. Annual Report 2000, BKG Frankfurt.

IGS (1996): *SINEX – Solution Independent Exchange Format Version 1.00*. International GPS Service.

IGS (2000): *IGS Annual Report 2000 *. International GPS Service, Central Bureau, NASA-JPL.

IGS (2002a): *Homepage of the Central Bureau in the World Wide Web*. Http://igscb.jpl.nasa.gov.

IGS (2002b): *Towards Real Time, Network, Data and Analysis Center Workshop 2002*. Proceedings, Ottawa Canada, IGS Website.

Ilk, K. (1990): *Zukünftige Möglichkeiten der globalen hochauflösenden Schwerefeldbestimmung*. In: Schneider (ed.) (1990): 213–255.

ILRS (2000): *ILRS Annual Report*. International Laser Ranging Service.

ILRS (2001): *Proceed. 12th International Workshop on Laser Ranging, Matera, Italy, 13–17 November 2000* . ILRS Website.

ILRS (2002): *International Laser Ranging Service*. ILRS Website.

Jacobsen, K. (1992): *Recent experiences in combined block adjustment with kinematic GPS data*. XVII Congress International Society for Photogrammetry and Remote Sensing, IUSM Session on GPS, Washington D.C., USA.

Jacobsen, K. (1997): *Operational Block Adjustment without Control Points*. ASPRS Annual Convention, Seattle 1997.

Jacobsen, K. (2000): *Potential and Limitation of Direct Sensor Orientation*. IntArchPhRS (33), ISPRS Amsterdam 2000, B3a: 429–435.

Jäger, R. (1990): *Optimum positions for GPS points and supporting fix-points in geodetic networks*. In: Bock, Leppard (eds.) (1990): 254–261.

Jäger, R., Schneid, S. (2002): *Passpunktfreie direkte Höhenbestimmung mittels DFHBF - ein Konzept für Positionierungsdienste wie SAPOPS*. Proceed. 4. SAPOS Symposium Hannover: 149–166.

Jahn, C.-H., Seeber, G., Foulger, G., Einarsson, P. (1991): *GPS epoch measurements across the Mid-Atlantic plate boundary in Northern Iceland 1987-1990*. IUGG Symp. U5 "Application of Gravimetry and Space Techniques for Geodynamics and Ocean Dynamics", Vienna.

Jahn, C.-H., Winter, R. (2002): *SAPOS-HEPS in Niedersachsen*. 4. SAPOS Symposium Hannover, 22–36.

Janiczek, P. (ed.) (1986): *Global Positioning System*. Papers published in Navigation, Washington D.C., Vol. I (1980), Vol. II (1984), Vol. III (1986).

Jayles, J., Berthias, J.-P., Laurichesse, D. (2000): *DORIS-DIODE: From SPOT–4 to JASON–1*. Proceed. DORIS days 2000.

Jefferson, D., Bar-Sever, Y. (2000): *Accuracy and Consistency of Broadcast GPS Ephemeris Data*. Proceed. ION GPS-2000: 391–395.

Jekeli, C. (2000): *Calibration/validation methods for GRACE*. In: Schwarz (ed.) (2000): 83–88.

Johannessen, R. (1997): *Interference: Sources and Symptoms*. GPS World, 8 (11): 45–48.

Johnson, N. (1994): *GLONASS Spacecraft*. GPS World 5 (11): 51–58.

de Jong, K. (2001): *GPS and the Internet*. GEO Informatics, January/February: 12–13.

Joosten, P., Tiberius, C. (2000): *Fixing the Ambiguities: Are you Sure They're Right?* GPS World 11 (5): 46–51.

Kahmen, H. (1978): *Elektronische Messverfahren in der Geodäsie*. 2. Aufl., Wichmann Karlsruhe.

Kahmen, H., Feig, W. (1988): *Surveying*. W. de Gruyter, Berlin-New York.

Kahn, W., Vonbun, F., Smith, D., Englar, T., Gibbs, B. (1980): *Performance analysis of the spaceborne laser ranging system*. Bull. Géod. 54 (2): 165–180.

Kakkuri, J. (1995): *Final Results of the Baltic Sea Level 1993 GPS Campain*. Rep. Finnish Geod. Inst. 94:2.

Kang, Z. (1998): *Präzise Bahnbestimmung niedrigfliegender Satelliten mittels GPS und die Nutzung für die globale Schwerefeldmodellierung*. Scientific Technical Report STR98/25, GeoForschungsZentrum, Potsdam.

Kaniuth, K., Drewes, H., Stuber, K., Tremel, H. (2001): *Bestimmung rezenter Krustendeformationen im zentralen Mittelmeer mit GPS*. ZfV 126: 256–262.

Kaniuth, K., Kleuren, D., Tremel, H. (1998): *Sensitivity of GPS Height Estimates to Tropospheric Delay Modelling*. AVN 105: 200–207.

Kaniuth, K., Stuber, K. (1999): *Einfluss von Antennen-Radomen auf die GPS-Höhenbestimmung*. AVN, 106: 234–238.

Kaplan, E. (1996): *Understanding GPS: Principles and Applications*. Artech House, Boston, London.

Kaplan, M. (1976): *Modern spacecraft dynamics & control*. New York.

Karslioglu, M. (2000): *Probleme der Mess- und Regeltechnik bei der geodätischen Nutzung künstlicher Erdsatelliten*. Scientific Technical Report STR00/06, GeoForschungsZentrum Potsdam.

Kaula, W. (1966): *Theory of satellite geodesy*. Blaisdell Publ. Comp., London.

Keating, T., Taylor, P., Kahn, W., Lerch, F. (1986): *Geopotential Research Mission, science, engineering and program summary*. NASA Technical Memorandum 86240.

Kee, C. (1996): *Wide Area Differential GPS*. In: Parkinson, Spilker (eds.) (1996), Vol. 2, Chap. 3: 81–115.

Kee, C., Parkinson, B., Axelrad, P. (1999): *Wide Area Differential GPS*. Navigation Vol. 38 No.2.

Keller, W., Heß, D. (1999): *Gradiometrie mit GRACE*. ZfV 124: 137–144; 205–211.

Khan, M. (1983): *Satellite contributions to geophysical exploration at sea. In: Geyer, R. A. (ed.): CRC Handbook of Geophysical Exploration at Sea, 3-68*. Boca Raton, Florida.

Kim, D., Langley, R. (1999): *An optimized least-squares technique for improving ambiguity resolution performance and computational efficiency*. Proceed. ION GPS-99: 1579–1588.

Kim, D., Langley, R. (2000): *GPS Ambiguity Resolution and Validation: methodologies, Trends and Issues*. Int. Symp. GPS/GNSS, Seoul, Korea.

Kim, M. (1997): *Theory of satellite ground-track crossovers*. J. Geod. 71: 749–767.

King, R., Masters, R., Rizos, C., Stolz, A. (1987): *Surveying with GPS*. University of New South Wales, Australia 1985, Verlag F. Dümmler Bonn 1987.

King-Hele, D. (1958): *The effect of the Earth's oblateness on the orbit of a near satellite*. Proceed. Roy. Soc., Ser.A, 247: 49–72.

King-Hele, D., Merson, R. H. (1958): *Use of artificial satellites to explore the Earth's gravity field: Results from Sputnik 2*. Nature 182: 640–641.

Kirchner, G., Koidl, F., Prochazka, I., Hamal, K. (1999): *SPAD Time Walk Compensation and Return Energy Dependent Ranging*. In: Schlüter, Schreiber, Dassing (1999), Vol. 2: 521–525.

Klees, R., Koop, R., Visser, P., van den IJssel, J. (2000): *Data analysis for the GOCE mission*. In: Schwarz (ed.) (2000): 68–74.

Kleusberg, A. (1984): *Diagnose und Therapie von Geodätischen Satellitennetzen vom Typ Doppler*. DGK C 293, München.

Kleusberg, A. (1990): *A review of kinematic and static GPS surveying procedures*. Proceed. GPS'90, Ottawa: 1102–1113.

Kleusberg, A. (1995): *Mathematics of attitude determination with GPS*. GPS World 6 (9): 72–78.

Kleusberg, A. (1998): *Atmospheric Models from GPS*. In: Teunissen, Kleusberg (1998), Chap. 15: 599–623.

Klobuchar, J. (1987): *Ionospheric time-delay algorithm for single-frequency GPS users*. IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-23, No.3.

Klobuchar, J. (1991): *Ionospheric Effects on GPS*. GPS World, 2 (4): 48–51.

Klobuchar, J. (1996): *Ionospheric Effects on GPS*. In: Parkinson, Spilker (eds.) (1996), Vol. 1, Chap. 12: 485–515.

Klokočník (1982): *A review of methods of comparison of "resonant" lumped coefficients*. Bull. Astron. Inst. Czechl. 33: 89–104.

Klosko, S. (1999): *SLR Contributions to Determining the Gravitational Field and its Variations*. In: Schlüter, Schreiber, Dassing (1999), Vol. 1: 28–40.

Klotz, J., Reinking, D., Angermann, D. (1996): *Die Vermessung der Deformation der Erdoberfläche*. Geowissenschaften 14: 389–394.

Knudsen, P., Andersen, O. (1997): *Improved Recovery of the Global Marine Gravity Field from the GEOSAT and the ERS-1 Geodetic Mission Altimetry*. In: Segawa, Fujimoto, Okubo (eds.) (1997): 429–436.

Koch, K. (1990): *Parameter estimation and hypothesis testing in linear models*. Dümmler, Bonn.

Kolb, W. (1999): *Interferenzprobleme bei GPS-Signalen*. DVW Schriftenreihe 35: 161–198.

Kolenkiewicz, R., Martin, C. (1982): *SEASAT altimeter height calibration*. JGR 87 (C5): 3189–3197.

Komaki, K., Sasaki, M., Hashimoto, H. (1985): *Plan for satellite geodesy in Japan*. CSTG Bull. No. 8: 58–63, München.

König, R., Chen, Z., Reigber, C., Schwintzer, P. (1999): *Improvement in global gravity field recovery using GFZ-1 satellite laser tracking data*. J. Geod. 73: 398–406.

Kouba, J. (1983a): *A review of geodetic and geodynamic satellite Doppler positioning*. Review Geophys. Space Phys. 21 (1): 27–40.

Kouba, J. (1983b): *An efficient short-arc orbit computation*. Bull. Géod. 57 (2): 138–145.

Kouba, J. (1983c): *GEODOP V–Geodetic Doppler Positioning Programs (Version V)* . Earth Physics Branch, Ottawa.

Kovalevsky, J. (1971): *Mécanique Céleste. In: Levallois, J.J., Géodésie Générale, Tome IV, 85–268* . Editions Eyrolles, Paris.

Kovalevsky, J. (1990): *Astrométrie moderne*. Lecture Notes in Physics 358, Springer-Verlag, Berlin-Heidelberg-New York.

Kovalevsky, J. (1995): *Modern Astrometry*. Springer-Verlag, Berlin-Heidelberg-New York.

Kovalevsky, J., Lindegren, L., Perryman, M., 22 more authors (1997): *The Hipparcos Catalogue as a realisation of the extragalactic reference system*. Astron. Astrophys. 323: 620–633.

Kovalevsky, J., Mueller, I., Kolaczek, B. (eds.) (1989): *Reference frames in astronomy and geophysics*. Astrophysics and Space Science Library, Vol. 154. D. Reidel, Dordrecht/Boston/London.

Kozai, Y. (1959): *On the effects of the sun and the moon upon the motion of a close Earth satellite*. SAO Special Report 22, Cambridge, Mass.

Kozai, Y. (1966): *Lunisolar perturbations with short periods*. SAO Special Report 235, Cambridge, Mass.

Krynski, J. (1983): *Study on satellite-to-satellite data processing*. In: Schwarz, Lachapelle (eds.), Geodesy in Transition, 313–326, Calgary.

Kumar, M., Maul, G. (eds.) (1991): *Proc. Int. Symp. on Marine Positioning INSMAP 90*. University of Miami, Oct. 15–19, 1990, Marine Technology Society, Washington D.C., USA.

Lachapelle, G. (1990): *GPS in the marine environment: status, trends and prospects*. FIG XIX Int. Congress Helsinki, Comm.4, 108–118.

Lachapelle, G. (1991): *GPS observables and error sources for kinematic positioning*. In: Schwarz, Lachapelle (eds.) (1991): 17–26.

Lachapelle, G., Hagglund, J., Falkenberg, W., Bellemare, P., Eaton, M., Casey, M. (1986): *GPS land kinematic positioning experiments*. Proceed. 4th Int. Geod. Symp. on Satellite Positioning, Austin, Vol. 2: 1327–1344.

Ladd, J., Counselmann, C., Gourewitsch, S. (1985): *The Macrometer II Dual-Band Interferometric Surveyor*. Proc. 1st Int. Symp. Precise Positioning with GPS, Rockville, Vol. 1: 175–180.

Ladd, J., Welshe, R., Brown, A., Sturza, M. (1986): *Mini-Mac - A new generation dual-band surveyor*. Bull. Géod., 60 (3): 221–228.

Lambeck, K., Cazenave, A., Balmino, G. (1975): *Solid earth and fluid tides from satellite orbit analysis*. In: Proceed. Int. Symp. "Use of Artificial Satellites for Geodesy and Geodynamics", Athens, 353–393.

Landau, H. (1988): *Zur Nutzung des Global Positioning Systems in Geodäsie und Geodynamik*. Studiengang Vermessungswesen, Schriftenreihe UniBw München 36.

Landau, H. (1990): *Precise GPS positioning with the multistation-multisession software TOPAS*. In: GPS for Geodesy and Geodynamics, Luxembourg, 61–68.

Landau, H., Hagmeier, D. (1986): *Analysis of the required force modeling for NAVSTAR GPS satellites*. Studiengang Vermessungswesen, Schriftenreihe UniBw 19: 193–208.

Langley, R. (1990): *Why is the GPS signal so complex?* GPS World 1 (3): 56–59.

Langley, R. (1991a): *The GPS receiver: an introduction*. GPS World 2 (1): 50–53.

Langley, R. (1991b): *Time, Clocks, and GPS*. GPS World 2 (11): 38–42.

Langley, R. (1997a): *GLONASS: Review and Update*. GPS World 8 (7): 46–51.

Langley, R. (1997b): *GPS Receiver System Noise*. GPS World 8 (6): 40–45.

Langley, R. (1998a): *A Primer on GPS Antennas*. GPS World 9 (7): 50–54.

Langley, R. (1998b): *Propagation of the GPS Signals*. In: Teunissen, Kleusberg (1998), Chap. 3: 111–149.

Langley, R. (1998c): *RTK GPS*. GPS World 9 (9): 70–76.

Langley, R. (1999a): *The GPS End-of-Week-Rollover*. GPS World 9 (11): 40–47.

Langley, R. (1999b): *The Integrity of GPS*. GPS World 10 (3): 60–63.

Langley, R. (2000a): *GPS, the Ionosphere, and the Solar Maximum*. GPS World 11 (7): 44–49.

Langley, R. (2000b): *The Evolution of the GPS Receiver*. GPS World 11 (4): 54–58.

Le Pichon, X., Francheteau, J., Bonnin, J. (1973): *Plate Tectonics. Development in Geotectonics*. Vol. 6, Amsterdam-London-New York.

Leberl, F. (1990): *Radargrammetric Image Processing*. Artech House, Norwood, MA, USA.

Lee, J. (1996): *Untersuchung von Verfahren zur kombinierten Aerotriangulation mittels integriertem GPS/INS*. Wiss. Arb. Univ. Hannover, Nr. 220.

Lefebvre, M., Cazenave, A., Escudier, P., Biancale, R., Cretaux, J., Soudarin, L., Valette, J. (1995): *Space Tracking System Improves Accuracy of Geodetic Measurements*. EOS 77 (4): 25–29.

Leick, A. (1995): *GPS satellite surveying*. 2nd. Edition, John Wiley, New York.

Lelgemann, D. (1979): *Eine lineare Lösung der Hillschen kanonischen Gleichungen für die Bewegung eines Satelliten im Gravitationsfeld der Erde*. Astr. Geod. Arb. 39: 351–359, München.

Lemoine, F., Smith, D., et al. (1998): *The development of the joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) Geopotential Model EGM96*. NASA/TP-1998-206861. Goddard Space Flight Center, Greenbelt, Maryland.

Leppard, N. (1980): *Satellite Doppler fixation and international boundaries*. Phil. Trans. Roy. Soc. London, Ser. A, Vol. 294: 289–298.

LeProvost, C. (2001): *Ocean Tides*. In: Fu, Cazenave (eds.) (2001), Chap. 6: 267–303.

LeProvost, C., Lyard, L., Kolines, J., Genco, M., Rabilloud, F. (1998): *A hydrodynamic ocean tide model improved by assimilating a satellite altimeter derived dataset*. JGR 103 (C3): 5513–5529.

Lerch, F., Klosko, S., Laubscher, R., Wagner, C. (1979): *Gravity model improvement using GEOS-3 (GEM9 and 10)*. JGR 84 (B 8): 3897–3916.

Lerch, F., Klosko, S., Patel, G. (1983): *A refined gravity model from LAGEOS (GEM-L2)*. NASA Techn. Mem. 84986.

Lerch, F., Wagner, C., Klosko, S., Belott, R. (1978): *Goddard Earth Model development for oceanographic applications*. Proceed. Marine Geodesy Symposium, Miami.

Lerch, F. J. (1992): *Geopotential models of the Earth from satellite tracking, altimeter and surface gravity observations, GEM-T3 and GEM-T3S*. NASA Tech. Mem. 104555.

Lewandowski, W., Azpubib, J., Klepszynski, W. (1999): *GPS: Primary Tool for Time Transfer*. Proc. IEEE, Vol. 87 (1): 163–172.

Li, K. (1992): *Empirische Untersuchungen zur GPS-gestützten kombinierten Blockausgleichung*. Wiss. Arb. Univ. Hannover, Nr. 175.

Lichten, S. (1990): *High accuracy Global Positioning System orbit determination. Progress and prospects*. In: Bock, Leppard (eds.) (1990): 146–164.

Lichtenegger, H., Hofmann-Wellenhof, B. (1990): *GPS-Data preprocessing for cycle slip detection*. In: Bock, Leppard (eds.) (1990): 57–68.

Liebsch, G. (1996): *Aufbereitung und Nutzung von Pegelmessungen für geodätische und godynamische Zielstellungen*. Doctor Dissertation, University Dresden.

Lillesand, T., Kiefer, R. (2000): *Remote Sensing and Image Interpretation*. 4th Edition, J. Wiley & Sons, New York.

Logsdon, T. (1998): *Orbital Mechanics: Theory and Applications*. John Wiley, & Sons, New York.

Lohmar, F. (1985): *Zur Berechnung ionosphärischer Refraktionskorrekturen für VLBI-Beobachtungen aus simultanen Dopplermessungen nach Satelliten*. Mitt. Geod. Inst. Univ. Bonn, Nr.67.

Lombardi, M., Nelson, L., Novick, A., Zhang, V. (2001): *Time and Frequency Measurement Using the Global Positioning System (GPS)*. Int. J. of Metrology 8: 26–33.

Lorrain, P., Corson, D., Lorrain, F. (1988): *Electromagnetic fields and waves: including electric circuits*. 3rd. edition. Freeman, New York.

Lundquist, C., Veis, G. (1966): *Geodetic parameters for a 1966 Smithsonian Institution Standard Earth*. SAO Spec.Rep. 200.

Lutgens, F., Tarbuck, E. (1998): *The atmosphere: an introduction to meteorology*. 7. ed., Prentice Hall, Upper Saddle River, N.J.

Ma, C., Arias, E., Eubanks, T. a. (1998): *The International Celestial Reference Frame based on VLBI Observations of Extragalactic Radio Sources*. Astron. J. 116: 516–546.

Ma, C., Feissel, M. (eds.) (1997): *Definition and Realization of the International Celestial Reference System by VLBI Astrometry of Extragalactic Objects*. IERS Technical Note 23, Observatoire de Paris, Paris.

Mac Arthur, J., Marth, P., Wall, J. (1987): *The GEOSAT radar altimeter*. Johns Hopkins APL Technical Digest, Laurel Md., 8 (2): 176–181.

Mac Doran, P. (1979): *Satellite emission radio interferometric earth surveying SERIES-GPS geodetic system*. Bull. Géod. 53: 117.

Mac Doran, P. (1983): *SERIES-GPS codeless pseudo-ranging positioning technology*. CSTG Bull. No.5: 46–55.

Mackie, J. (1985): *The elements of astronomy for surveyors*. 9th edition, Charles Griffin & Company, High Wycombe.

Mader, G. (1990): *Ambiguity function techniques for GPS phase initialization and kinematic solutions*. In: GPS'90, Ottawa, 1248–1256.

Mader, G. (1999): *GPS Antenna Calibration at the National Geodetic Survey*. GPS Solutions 3 (1): 50–58.

Mader, G., Czopek, F. (2002): *The Block IIa Satellite: Calibrating Antenna Phase Center*. GPS World 13 (5): 40–46.

Maenpa, J., Balodis, M., Sandholzer, J., Walter, G. (1997): *New Interference Rejection technology from Leica*. Proceed. ION GPS-97, Kansas City.

Malys, S., Slater, J. (1994): *Maintenance and Enhancement of the World Geodetic System 1984*. Proceed. ION GPS-94, Salt Lake City, Utah, 17–24.

Mangs, G., Mittal, S., Stansell, T. (2001): *Worldwide Beacon DGPS*. Sea Technology, May 2001, 39–45.

Maniatis, A., Campbell, J., Müller, A., Schuh, H., Seeger, H. (1987): *Zur Auswertung von geodätischen GPS-Beobachtungen*. AVN 94: 9–19.

Maral, G., Bousquet, P. (1986): *Satellite communication systems*. John Wiley & Sons, Chichester-New York.

Marini, J. (1972): *Correction of satellite tracking data for an arbitrary atmospheric profile*. Radio Science 7 (2): 223–231.

Markovitz, W. (1954): *Photographic determination of the moon's position and applications to the measure of time, rotation of the earth, and geodesy*. Astron. J. 59: 69–73.

Marks, K., D.C., M. A., Sandwell, D. (1991): *GEOSAT GM data reveal new details of ocean floor*. EOS 72 (14): 145.

Marquis, W., Krier, C. (2000): *Examination of the GPS Block IIR Solar Pressure Model*. Proceed. ION GPS-2000: 407–416.

Marsh, J., Lerch, F. (1985): *Precision geodesy and geodynamics using STARLETTE laser ranging*. JGR 90 (B11): 9335–9345.

Marsh, J., Lerch, F., et al. (1990): *The GEM-T2 gravitational model*. JGR 95 (B13): 22 043–22 071.

Marsh, J., Martin, T. (1982): *The SEASAT altimeter mean sea surface model*. JGR 87 (C5): 3269–3280.

Marsh, J., Vincent, S. (1975): *Gravimetric geoid computations and comparisons with Skylab altimeter data in the GEOS-C calibration area*. Gen. Ass. IUGG, Grenoble.

Marsh, J., et al. (1987): *An improved model of the Earth's gravitational field*. GEM-T1. NASA Technical Memorandum 4019.

Martin, D., Oh, I. (1979): *Utilization of satellite-to-satellite tracking data for determination of the geocentric gravitational constant GM*. JGR 84 (B8): 3944–3950.

Massonet, D. (1997): *Radar-Interferometrie zur Messung der Erdkrustendynamik*. Spektrum der Wissenschaft, H. 9: 56–65.

Massonet, D., et al. (1993): *The Displacement Field of the Landers Earthquake Mapped by Radar Interferometry*. Nature 364: 138–142.

Maul, G. (1985): *Introduction to satellite oceanography*. Martinus Nijhoff Publ., Dordrecht.

Mayer, M., Lindner, K., Kutterer, H., Heck, B. (2000): *Deformationsanalyse zur Aufdeckung von Punkt- und Blockbewegungen im Bereich der Antarktischen Halbinsel*. In: Dietrich (ed.) (2000): 127–144.

McCarthy, D. (2000): *IERS Conventions (2000)*. IERS Technical Note, International Earth Rotation Service.

McConathy, D., Kilgus, C. (1987): *The Navy GEOSAT mission: an overview*. Johns Hopkins APL Technical Digest, 8 (2): 170–175, Laurel Md.

McDonald, K. (1999): *Opportunity Knocks. Will GPS Modernization Open Doors?* GPS World 10 (9): 36–46.

Meehan, T., Srinivasan, J., Spitzmesser, D., Dunn, C., J.Y. TEN, J. T., MUNSON, T., DUNCAN, C. (1992): *The TurboRogue receiver*. 6th Int. Geod. Symp. Satellite Positioning, Columbus Ohio, USA.

Melbourne, W., Davis, E., Duncan, C., 6 more authors (1994a): *The applications of spaceborne GPS to atmospheric limb sounding and global change monitoring*. JPL Publication 94-18, NASA Jet Propulsion Laboratory, Pasadena.

Melbourne, W., Dixon, T., Thorton, C. (1985): *GPS-based geodesy in Mexico and the Caribean*. CSTG Bull. 8: 182–190, München.

Melbourne, W., Tapley, B., Davis, E., Yunck, T. (1994b): *The GPS flight experiment on TOPEX/Poseidon*. Geophys. Res. Lett. 21 (7).

Melbourne, W., et al. (1983): *Project Merit Standards*. U.S. Naval Observatory Circular No.167, Dec. 27.

Mendes, V., Langley, R. (1994): *A Comprehensive Analysis of Mapping Functions Used in Modelling Tropospheric Propagation Delay in Space Geodetic Data*. Proceed. KIS 94: 87–92.

Mendes, V., Langley, R. (1999): *Tropospheric Zenith Delay Prediction Accuracy for High-Precision GPS Positioning and Navigation*. Navigation: J. of the Institute of Navigation 46 (1): 25–34.

Mendes, V., Prates, G., Pavlis, E., Pavlis, D., Langley, R. (2001): *Improved Mapping Functions for Atmospheric Refraction Correction in SLR*. Geophys. Res. Lett.

Menge, F., Seeber, G. (2000): *Untersuchungen und Beiträge zur Problematik der Phasenzentrumsvariationen von GPS Antennen*. In: Dietrich (ed.) (2000): 181–194.

Merrigan, M., Swift, E., Wong, R., Saffel, J. (2002): *A Refinement to the World Geodetic System 1984 Reference Frame*. ION-GPS 2002, Portland, OR, Sept. 2002.

Meyer, U., Charlot, P., Biancale, R. (2000): *GINS: A New Multi-Technique Software for VLBI Analysis*. In: Vandenberg, Baver (eds.) (2000): 324–328.

Milani, A., Nobili, A., Farinella, P. (1987): *Non-gravitational perturbations and satellite geodesy*. Adam Hilger Ltd., Bristol.

Milliken, R., Zoller, C. (1980): *Principles of NAVSTAR and system characteristics*. Navigation 25 (2), 1978. In: Janiczek (ed.) (1986), Vol. 1, 3–14.

Minster, J., Jordan, T. (1978): *Present-day plate motions*. JGR 83 (B11): 5331–5354.

Misra, P., Enge, P. (2001): *Global Positioning System, Signals, Measurement, and Performance*. Ganga-Jamuna Press, Lincoln, Massachusetts.

Misra, P., Pratt, M., Muchnik, R., Burke, B., Hall, T. (1996): *GLONASS Performance: Mesurement Data Quality and System Upkeep*. Proceed. ION GPS-96, Kansas City, 261–270.

Monka, F. (1984): *Zur Aufbereitung und Auswertung von GEOS-3 Satellitenradaraltimetermessungen*. Wiss. Arb. Univ. Hannover, Nr.134.

Montag, H. (1984): *Zur Untersuchung des Erdrotationsvektors mit Hilfe von Laserentfernungsmessungen zu künstlichen Erdsatelliten*. Veröff. Zentralinstitut Physik der Erde, Nr. 80, Berlin.

Montenbruck, O., Gill, E. (2000): *Satellite Orbits - Models, Methods, and Applications*. Springer-Verlag, Berlin-Heidelberg.

Moritz, H. (1970): *Methoden zur Berechnung von Satellitentriangulationen*. AVN 77: 353–359.

Moritz, H. (1990): *The figure of the Earth. Theoretical geodesy and the Earth's interior*. H. Wichmann, Karlsruhe.

Moritz, H. (2000): *Geodetic Reference System 1980*. J. Geod. 74: 128–133.

Moritz, H., Mueller, I. (1987): *Earth Rotation, Theory and Observation*. The Ungar Publ. Comp., New York.

Mueller, I. (1964): *Introduction to satellite geodesy*. The Ungar Publ. Comp., New York.

Mueller, I. (1969): *Spherical and practical astronomy as applied to geodesy*. The Ungar Publ. Comp., New York.

Mueller, I. (1975): *Tracking station positioning from artificial satellite observations*. Geophysical Surveys 2: 243–276.

Mueller, I., Wei, Z. (1985): *MERIT Main Campaign: Reference frame intercomparisons*. Geodetic and Geophysical Research Institute Hungarian Academy of Sciences. Papers in Honour of Antal Tarzy-Hornoch 85: 110–117, Sopron.

Mueller, T. (1994): *Wide Area Differential GPS*. GPS World 5 (6): 36–44.

Muellerschoen, R., Bar-Sever, Y., Bertiger, W., Stowers, D. (2001): *NASA's Global DGPS for High Precision Users*. GPS World 12 (1): 14–20.

Muellerschoen, R. W. B., Lough, M., Stowers, D., Dong, D. (2000): *An Internet-Based Global Differential GPS System, Initial Results*. ION National Technical Meeting, Anaheim, CA, Jan. 2000.

Müller, J. (1991): *Analyse von Lasermessungen zum Mond im Rahmen einer post-Newton'schen Theorie*. DGK C 383, München.

Müller, J. (2001): *Die Satellitengradiometriemission GOCE - Theorie, technische Realisierung und wissenschaftliche Bedeutung*. DGK C 541, München.

Nakiboglu, S., Krakiwski, E., Schwarz, K., Buffet, B., Wanless, A. (1985): *A multi-station, multi-pass approach to Global Positioning System improvement and precise positioning*. Geod.Surv.Canada, Ottawa, Report 85-003.

NGS (1986): *Geodetic Glossary*. National Geodetic Survey, Rockville.

NGSP (1977): *National Geodetic Satellite Program*. 2 Volumes. Washington D.C.

Niell, A. (1996): *Global mapping functions for the atmospheric delay at radio wavelengths*. JGR 101 (B2): 3227–3246.

Niemeier, W. (2002): *Ausgleichungsrechnung*. W. de Gruyter, Berlin-New York.

Niemeier, W., Rennen, M., Salbach, H. (2000): *Bestimmung regionaler und globaler Deformationen im Bereich der Antarktischen Halbinsel*. In: Dietrich (ed.) (2000): 109–126.

NIMA (2000): *Department of Defense World Geodetic System 1984*. Technical Report TR8350.2, 3.ed., National Imagery and Mapping Agency, Washington D.C., USA.

NOAA (1997): *The GEOSAT Altimeter JGM-3 GDRs on CD-ROM*. NODC Laboratory for Satellite Altimetry, Silver Spring, MD 20910, USA.

Nordtvedt, K. (2001): *Lunar Laser Ranging – A Comprehensive Probe of the Post-Newtonian Long Range Interaction*. In: C. Lämmerzahl, C. W. F. Francis, and F. W. Hehl (eds.): Lecture Notes in Physics 526: 317–329.

Nothnagel, A. (2000): *Der Einfluss des Wasserdampfes auf die modernen raumgestützten Messverfahren*. Mitt. Bd. 6, Bundesamt für Kartographie und Geodäsie, Frankfurt.

Nothnagel, A. (2002): *Third IVS Analysis Workshop*. IVS Newsletter, Issue 2, April 2002.

Noton, M. (1998): *Spacecraft Navigation and Guidance*. Springer-Verlag, Berlin-Heidelberg-New York.

O'Brien, D., Prata, A. (1990): *Navigation of ERS-1 along-track scanning radiometer (ATSR) images*. ESA Journal, 14 (4): 447–466.

O'Keefe, J. (1958): *Oblateness of the Earth by artificial satellites*. Harvard Coll. Obs. Announcement Card 1408.

Otsubo, T., Amagai, J., Kunimori, H. (1999): *Measuring Asijai's spin motion*. In: Schlüter, Schreiber, Dassing (1999) Vol. 2: 674–677.

Parkinson, B., Enge, P. (1996): *Differential GPS*. In: Parkinson, Spilker (eds.) (1996) Vol. 2, Chap. 1: 3–50.

Parkinson, B., Spilker, J. (eds.) (1996): *Global Positioning System: Theory and Applications*. American Institute of Aeronautics and Astronautics, 2 volumes, Washington D.C.

Paunonen, M. (1999): *The new satellite laser ranging system at Metsähovi*. In: Schlüter, Schreiber, Dassing (1999) Vol. 1: 139–144.

Pavlis, E., Smith, D., Torrence, M. (1991): *A global reference frame for geodynamics SL7.1*. XX. IUGG General Assembly, Vienna.

Pearlman, M., et al. (1983): *Early experience of the SAO Satellite Tracking Program*. EOS 64, June 21, 417.

Pearlman, M. R. (1977): *National Geodetic Satellite Program – Smithsonian Astrophysical Observatory, Instrumentation*. In: NGSP (1977) Vol.2: 797–811.

Pelzer, H. (1985): *Grundlagen der mathematischen Statistik und der Ausgleichsrechnung*. In: Pelzer H. (ed.): Geodätische Netze in Landes- und Ingenieurvermessung II, 3–120, Konrad Wittwer, Stuttgart.

Perryman, M., Pace, O. (2000): *GAIA – Unravelling the Origin and Evolution of Our Galaxy*. Esa bulletin 103: 26–35.

Pierron, F., Nicolas, J., Kasser, M., Haase, J. (1999): *Status and new capabilities of the French Transportable Laser Ranging System*. In: Schlüter, Schreiber, Dassing (1999) Vol. 1: 104–112.

Pisacane, V. (ed.) (1998): *The Legacy of Transit*. Johns Hopkins APL Technical Digest, Vol. 19, No. 1.

Ploner, M. (1996): *CCD - Astrometrie von Objekten des geostationären Ringes*. Geow. Mitt. TU Wien H. 46.

Ploner, M., Jackson, P. (1999): *Satellitenastrometrie mit CCD*. In: Festschrift K. Bretterbauer, Geow. Mitt. TU Wien 50: 29–44.

Pospelov, S., Botchkovski, A. (2000): *GNSS Software Receivers*. GPS Solutions 4 (1): 48–55.

Powers, E., et al. (2002): *Real-Time Ultra-Precise Time Transfer to UTC Using the NASA Differntial GPS System*. IGS Workshop "Towards Real Time", Ottawa, April 8–11, 2002.

Preuß, E., Campbell, J. (1992): *Very-long-baseline interferometry in astro-, geo- and gravitational physics*. In: Ehlers, Schäfer (eds.), Proc. of Heraeus Seminar "Recent Developments in Relativistic Gravity Research", Bad Honnef 2–6 Sept. 1991. Lecture Notes in Physics, Springer-Verlag, Berlin.

Priester, W., Hergenhahn, G. (1958): *Bahnbestimmung von Erdsatelliten aus Doppler-Effekt-Messungen*. Wiss. Arb. AGF Nordrhein Westfalen, Vol. 82.

Prochazka, I., Hamal, K., Blazej, J. (1999): *Millimetre Ranging and Echo Signal Strength Monitoring with SPAD Detectors*. In: Schlüter, Schreiber, Dassing (1999) Vol. 2: 452–457.

PTTI (2000): *Proceedings of the Annual Precise Time and Time Interval (PTTI), Applications and Planning Meeting* . U.S.Naval Observatory, Washington, D.C., published yearly.

Purivigraipong, S., Unwin, M. (2001): *Determining the Attitude of a Minisatellite by GPS*. GPS World 13 (6): 60–66.

Quine, B. (1996): *Spacecraft Guidance Systems, Attitude Determination using Star Camera Data*. PhD Dissertation, University of Oxford.

Rabe, E. (1962): *Comparison of terrestrial and astronomical results for GM*. Proc. 1st Int. Symp. Use of Artificial Satellites for Geodesy, 383-386, Washington D.C.

Rapp, R. (1998): *Past and future developments in geopotential modeling*. In: Forsberg, Feissel, Dietrich (eds.) (1998): 58–78.

Rapp, R., Cazenave, A., Nerem, R. (eds.) (1996): *Global Gravity Field and Its Temporal Variations*. IAG Symp. Proceed. 116, Springer-Verlag, Berlin-Heidelberg-New York.

Raquet, J., Lachapelle, G. (2001): *RTK Positioning with Multiple Reference Stations*. GPS World 12 (4): 48–53.

Reigber, C. (1981): *Representation of orbital element variations and force function with respect to various reference systems*. Bull. Géod. 55 (1): 111–131.

Reigber, C., Balmino, G., Moynot, B., Mueller, H. (1983): *The GRIM 3 Earth Gravity Field Model*. Man. geod. 8: 93–138.

Reigber, C., Gendt, G., Dick, G., Tomassini, M. (2002): *Near real-time water vapor monitoring for weather forcasts* . GPS World 13 (1): 18–27.

Reigber, C., Hartl, P. (1989): *PRARE/PRAREE system development status*. CSTG Bull. 11: 157–164, München.

Reigber, C., Hartl, P. (1990): *Das Satellitenbeobachtungssystem PRARE*. ZfV 115: 512–519.

Reigber, C., Schwintzer, P., Hartl, P., Ilk, K., Rummel, R., M. van Gelder, E. S., Wakker, K., Ambrosius, B., Leemann, H. (1987): *Study of a Satellite-to-Satellite Tracking Gravity Mission*. ESTEC/Contract No. 6557/85/NL PP(SC), München.

Reigber, C., Schwintzer, P., Lühr, H. (1999): *The CHAMP geopotential mission*. Boll. Geof. Teor. Appl. 40: 285–289.

Remondi, B. (1984): *Using the Global Positioning System phase observable for relative geodesy: Modelling, processing and results*. PhD Dissertation, University of Texas, Austin.

Remondi, B. (1985): *Performing centimeters accuracy relative surveys in seconds using GPS carrier phase*. Proc. 1st Int. Symp. Precise Positioning with GPS, Rockville, Vol. 2: 789–797.

Remondi, B. (1986): *Performing centimeter-level surveys in seconds with GPS carrier phase: Initial results*. Proc. 4th Int. Geodetic Symposium on Satellite Positioning, Austin, Vol. 1: 1229–1250.

Remondi, B. (1990): *Recent advances in pseudo-kinematic surveying*. In: ION GPS-90: 106–114.

Remondi, B. (1991): *NGS second Generation ASCII and Binary Orbit Formats and Associated Interpolated Studies*. Proc. IUGG XXth Gen.Assembly Vienna, August1991.

Rémy, F., Legresy, B., Testut, L. (2001): *Ice sheet and satellite altimetry*. Surveys in Geophysics 22: 1–29.

Renken, H. (1999): *Ein Verfahren der bildverarbeitenden Erkennung von unbekannten Sternmustern zur autonomen und 3-axialen Lagebestimmung von Raumflugkörpern*. Shaker Verlag, Aachen.

Richardus, P. (1984): *Project Surveying*. 2. Edition, Balkema, Rotterdam-Boston.

Riepl, S. (1998): *Lasermessungen nach Erdsatelliten auf zwei Wellenlängen unter Verwendung einer Streak-Kamera*. DGK C 495, München.

Ries, J. (1997): *Non-gravitational force modeling effects on satellite orbits*. CSTG Bull. 14: 73–80, München.

Ries, J., Eanes, R., Shum, C., Watkins, M. (1992): *Progress in the determination of the gravitational coefficient of the Earth*. Geophys. Res. Lett. 19 (6): 529–531.

Ries, J., Tapley, B. (1999): *Centimeter level orbit determination for the TOPEX/POSEIDON altimeter satellite*. Adv. Astronautical Sci. 102: 583–598.

Rim, H., Schutz, B. (1999): *Precision Orbit Determination (POD)*. Geoscience Laser Altimeter System (GLAS)–Algorithm Theoretical Basis Document, Center of Space Research, University of Texas, Austin.

Rinner, K., Killian, K., Meissl, P. (1969): *Beiträge zur Theorie der geodätischen Netze im Raum*. DGK A 61, München.

Rinner, K., Seeber, G., Seeger, H. (eds.) (1982): *Die Deutsch-Österreichische Dopplerkampagne*. DGK B 269, München.

Rizos, C., Stolz, A. (1985): *Force modelling for GPS satellites orbits*. 1st. Int. Symp. Precise Positioning with GPS, Rockville, Md., Vol. 1: 87–98.

Robertson, D., Carter, W. (1985): *Earth orientation determined from VLBI observations*. Proceed. Int. Conf. Earth Rotation and Terrestrial Reference Frame, Columbus, Vol. 1: 296–306.

Robinson, I. (1995): *Satellite Oceanography*. Wiley, Chichester, New York.

Rohlfs, K., Wilson, T. (1996): *Tools of Radio Astronomy*. 3rd ed. Springer-Verlag, Berlin-New York.

Roßbach, U. (2001): *Positioning and Navigation Using the Russian Satellite System GLONASS*. Studiengang Geodäsie und Geoinformation, Schriftenreihe HsBw, 70, München.

Roßbach, U., Habrich, H., Zarraoa, N. (1996): *Transformation Parameters Between PZ-90 and WGS 84*. Proceed. ION GPS-90, Kansas City: 279–285.

Rothacher, M. (1992): *Orbits of Satellite Systems in Space Geodesy*. Geod. Geophys. Arb. Schweiz, Schweizerische Geodätische Kommission, Zürich, Vol. 46.

Rothacher, M. (2000a): *Comparisons of Absolute and Relative Antenna Phase Center Variations*. GPS Solutions 4 (4).

Rothacher, M. (2000b): *Toward an Integrated Global Geodetic Observing System*. In: Rummel, Drewes, Bosch, Hornik (2000): 41–52.

Rothacher, M. (2002): *Combination of Space-Geodetic Techniques*. Proceed. IVS 2002 General Meeting, 33–43.

Rothacher, M., Mervart, L. (1996): *Bernese GPS Software Version 4.0*. University of Berne, Switzerland.

Rothacher, M., Schaer, S., Mervart, L., Beutler, G. (1995): *Determination of Antenna Phase Center Variations Using GPS Data*. Proceed. IGS Workshop, May 15–17, Potsdam.

Rothacher, M., Springer, T., Schaer, S., Beutler, G. (1998): *Processing strategies for regional GPS networks* . In: Brunner (ed.) (1998): 93–100.

Roy, A. (1978): *Orbital Motion*. Adam Hilger, Bristol.

RTCM (2001): *RTCM Recommended Standards for Differential GNSS (Global Navigation Satellite Systems) Service, Version 2.3*. Radio Technical Commission for Marine Services, 1800 Diagonal Road, Suite 600, Alexandria VA 22314-2840 USA.

Rubicam, D. (1986): *Lageos orbit decay due to infrared radiation from Earth*. EOS 67 (16).

Rummel, R. (1979): *Determination of short-wavelength components of the gravity field from satellite-to-satellite tracking or satellite gradiometry - an attempt to an idenfication of problem areas*. Man. geod. 4: 107–148.

Rummel, R. (1985): *Satellitengradiometrie*. ZfV 110: 242–257.

Rummel, R. (1986): *The Earth's Gravity Field*. Proceed. ESA Special Workshop on "Solid Earth Science & Application Mission for Europe (SESAME), ESA SP-1080, 69–76.

Rummel, R., Balmino, G., Johannessen, J., Visser, P., Woodworth, P. (2002): *Dedicated gravity field missions – principles and aims*. J. of Geodynamics 33: 3–20.

Rummel, R., Drewes, H., Bosch, W., Hornik, H. (2000): *Towards an Integrated Global Geodetic Observing System (IGGOS)*. IAG Symp. Proceed. 120, Springer-Verlag, Berlin-Heidelberg-New York.

Rummel, R., Reigber, C., Ilk, K. (1978): *The use of satellite-to-satellite tracking for gravity parameter recovery*. Proc. ESA Workshop on Space Oceanography, Navigation and Geodynamics (SONG), ESA-SP-137, 151–161.

Rummel, R., Sanso, F. (1993): *Satellite altimetry in geodesy and oceanography*. Lecture notes in Earth sciences, Vol. 50, Springer-Verlag, Berlin.

Rummel, R., Schrama, E. (1991): *Two complementary systems on-board ARISTOTELES, Gradio and GPS*. ESA Journal, 15 (1): 135–139.

Russel, S., Schaibly, J. (1980): *Control segment and user performance*. Navigation 25 (2) (1978). In: Janiczek (ed.) (1986) Vol. 1: 74–80.

Rutscheid, E. (1972): *Preliminary results of the SECOR Equatorial Network*. In: Henriksen (ed.): The Use of Artificial Satellites for Geodesy, Washington D.C.

Saastamoinen, J. (1973): *Contributions to the theory of atmospheric refraction*. Bull. Géod. 107 (1): 13–34.

Samain, E., Fridelance, P. (1998): *Time Transfer by Laser Link (T2L2) experiment on Mir*. Metrologia, 35: 151–159.

Sandwell, D., Milbert, D., Douglas, B. (1986): *Global nondynamic orbit improvement for altimetric satellites*. JGR 91 (B9): 9447–9451.

Sansò, F., Rummel, R. (eds.) (1989): *Theory of satellite geodesy and gravity field determination*. Lecture Notes in Earth Sciences 25, Springer-Verlag, Berlin-New York.

Santerre, R. (1991): *Impact of GPS satellite sky distribution*. Man. geod. 16 (1): 28–53.

Sasaki, M. (1983): *On a laser ranging system at Simosato Hydrographic Observatory*. CSTG Bull. 5, 80–86.

Schaer, S., Gurtner, W., Feltens, J. (1998): *IONEX: The IONosphere Map EXchnge Format Version I.*. Proceed. 1998 IGS Analysis Center Workshop, Darmstadt: 233–247.

Scharroo, R., Visser, P. (1998): *Precise orbit determination and gravity field improvement for the ERS satellites*. JGR 103: 8113–8127.

Scheneweck, M., et al. (1990): *Status of CIGNET and orbit determination at the National Geodetic Survey*. In: GPS'90: 179–189.

Schenke, H. (1984): *Untersuchungen zur Genauigkeit von Doppler-Satellitenbeobachtungen im Testnetz Westharz*. Wiss. Arb. Univ. Hannover, Nr. 129.

Schenke, H. (1991): *Impact of GPS on the accuracy of bathymetric charts from multi beam sonar surveys*. In: Kumar, Maul (eds.) (1991): 147–170.

Schenke, H., Dijkstra, S., Niederjasper, F., Hinze, H., Hoppmann, B., Schöne, T. (1998): *The New Bathymetric Charts of the Weddell Sea: AWI BCWS*. AGU, Antarctic Research Series, Vol. 75, 371–380.

Schildknecht, T. (1994): *Optical Astrometry of Fast Moving Objects Using CCD Detectors*. Geod. Geophys. Arb. Schweiz Vol. 49.

Schildknecht, T., Dudle, G. (2000): *Time and Frequency Transfer: High Precision Using GPS Phase Measurements*. GPS World 11 (2): 48–52.

Schlüter, W. (2002): *Die Aktivitäten der Fundamentalstation Wettzell*. DGK A 118: 148–151, München.

Schlüter, W., Hase, H., Böer, A., Riepl, S., Alarez, J., Cecioni, A. (2000): *Progress of the TIGO-project*. CSTG Bull. 16, 77–84.

Schlüter, W., Hase, H., Boer, A., Riepl, S., Cecioni, A. (2002): *TIGO starts its operations in Concepción in 2002*. CSTG Bull. 17: 92–93, München.

Schlüter, W., Pesec, P. (1982): *Mathematisches Modell zur Auswertung von Dopplermessungen*. DGK B 260, 37–54, München.

Schlüter, W., Schreiber, U., Dassing, R. (1999): *Proceedings 11th International Workshop on Laser Ranging, Deggendorf, Germany, Sept. 20–25, 1998*. Mitt. BKG Bd. 10, 1999.

Schmid, H. (1974): *Worldwide geometric satellite triangulation*. JGR 79 (35): 5349–5476.

Schmid, H. (1977): *National Geodetic Satellite Program-National Geodetic Survey*. In: NGSP (1977), Vol. 2, 525–644.

Schmidt, M., Moore, A. (2002): *Network Issues*. Position Paper, IGS Network Data and Analysis Center Workshop 2002 Ottawa, Canada.

Schmitz, M. (1998): *Untersuchungen zur strengen GPS Parametrisierung in der gemeinsamen Ausgleichung von kinematischem GPS und Aerotriangulation*. Wiss. Arb. Univ. Hannover, Nr.225.

Schneider, M. (1981): *Himmelsmechanik*. B.I. Wissenschaftsverlag, Mannheim, Wien, Zürich.

Schneider, M. (1988): *Satellitengeodäsie*. Mannheim/Wien/Zürich.

Schneider, M. (1993): *Himmelsmechanik, Bd. 2*. B.I. Wissenschaftsverlag, Mannheim, Wien, Zürich.

Schneider, M. (ed.) (1990): *Satellitengeodäsie. Ergebnisse aus dem gleichnamigen Sonderforschungsbereich*. VCH Verlagsgesellschaft, Weinheim.

Schödlbauer, A. (2000): *Geodätische Astronomie*. W. de Gruyter, Berlin-New York.

Schöne, T. (1997): *The Gravity Field in the Weddell Sea, Antarctica, by Radar Altimetry from GEOSAT and ERS-1 (in German)*. Ber. Polarforsch. 220, AWI Bremerhaven.

Schuh, H. (1987): *Die Radiointerferometrie auf langen Basen zur Bestimmung von Punktverschiebungen und Erdrotationsparametern*. DGK C 328, München.

Schuh, H. (2000): *Contribution of VLBI to Space Geodesy*. In: Rummel, Drewes, Bosch, Hornik (2000): 33–40.

Schuh, H., Charlot, P., Hase, H., 10 more authors (2002): *IVS Working Group 2 for Product Specification and Observing Programs - Final Report*. In: IVS Annual Report 2001 ed. by N. R. Vandenberg and K. D. Baver, NASA, 2002.

Schüler, T. (2001): *On Ground-Based GPS Tropospheric Delay Estimation*. Studiengang Geodäsie und Geoinformation, Schriftenreihe UniBw 73, München.

Schupler, B. (1994): *Signal Characteristics of GPS User Antennas*. Navigation 41 (3).

Schupler, B., Clark, T. (2001): *Characterizing the Behavior of Geodetic GPS Antennas*. GPS World 12 (2): 48–55.

Schutz, B. (1998): *Spaceborne Laser Altimetry: 2001 and beyond*. In: Plag, H. P. (ed.), WEGENER-98, Norwegian Mapping Authority, Honefoss, Norway.

Schuyer, M. (1997): *European Capability and Prospects for a Spaceborne Gravimetric Mission*. In: Lecture Notes in Earth Sciences, Vol. 65, 569–588, Springer-Verlag, Berlin-Heidelberg.

Schwarz, K. (1980): *Inertial surveying systems – experience and prognosis*. The Canadian Surveyor, 34.

Schwarz, K. (1983): *Inertial surveying and geodesy*. Reviews of Geophysics and Space Physics, 21 (4): 878–890.

Schwarz, K. (1991): *Kinematic modelling-progress and problems*. In: Schwarz, Lachapelle (eds.) (1991): 17–26.

Schwarz, K. (ed.) (2000): *Geodesy Beyond 2000: The Challenges of the First Decade*. IAG Symp. Proceed. 121, Springer-Verlag, Berlin-Heidelberg-New York.

Schwarz, K., Glennie, C., Bruton, A. (1997): *Improving accuracy and reliability of airborne gravimetry by multiple system configurations*. In: Forsberg, Feissel, Dietrich (eds.) (1998): 11–17.

Schwarz, K., Lachapelle, G. (eds.) (1991): *Kinematic systems in geodesy, surveying and remote sensing*. Papers from a Symp. held in Bauff, Canada, Sept. 1990. IAG Symp. 107, New York-Berlin.

Schwarz, K., Sideris, M. (1993): *Heights and GPS*. GPS World 4 (2): 50–56.

Schwiderski, H. (1984): *Combined hydrodynamical and empirical modeling of ocean tides*. In: Seeber, Apel (eds.): Geodetic Features of the Ocean Surface and their Implications, 215-229. D. Reidel, Dordrecht.

Schwintzer, P., et al. (1997): *Long Wavelength Global Gravity Models - GRIM4-S4, GRIM4-C4*. J. Geod. 71, 189–208.

Schwintzer, P., et al. (2000): *A new global Earth's gravity field model from satellite orbit perturbations: GRIM5-S1*. Geophys. Research Letters 27: 3611–3614.

Schwintzer, P., Kang, Z., Reigber, C., Zhu, S. (1995): *GPS satellite-to-satellite tracking for TOPEX/Poseidon precise orbit determination and gravity field model improvement*. J. Geodynamics 20 (2): 155–166.

Schwintzer, P., Reigber, C., Biancale, R., Balmino, G., et al. (1992): *GRIM4 – Globale Erdschwerefeldmodelle*. ZfV 117: 227–247.

Seeber, G. (1972): *Über das stochastische Verhalten von photographisch bestimmten Stern- und Satelliten- Koordinaten*. DGK C 178, München.

Seeber, G. (1979): *Inertiale Vermessungssysteme und ihre Anwendungsmöglichkeiten in der Geodäsie*. ZfV 104: 460–471.

Seeber, G. (1983): *The current role of Transit Doppler measurements for precise marine positioning*. Proceed. Symp. Point Positioning in Marine Geodesy, Maracaibo: 119–128.

Seeber, G. (1989a): *Anwendungsmöglichkeiten von GPS in Geodäsie und Nachbargebieten - Erfahrungen mit dem TI 4100 NAVSTAR Navigator*. Verm. Phot. Kulturtechnik 87: 479–489.

Seeber, G. (1989b): *Satellitengeodäsie*. W. de Gruyter, Berlin-New York.

Seeber, G. (1993): *Satellite Geodesy. Foundations, Methods, and Applications*. W. de Gruyter, Berlin-New York.

Seeber, G., Böder, V. (1998): *Precise Real-Time Positioning in Hydrography with DGPS and INS–Developments and Results*. Proceed. 4th Int. Symp. Marine Positioning (INSMAP 98), Melbourne, Florida: 265–275.

Seeber, G., Egge, D. (1981): *Positionsbestimmung mit Satelliten-Dopplerverfahren in bewegter Situation*. ZfV 106: 123–131.

Seeber, G., Egge, D., Schuchardt, A., Siebold, J., Wübbena, G. (1985): *Experiences with the TI4100 NAVSTAR Navigator at the University of Hannover*. Proceed. 1st Int. Symp. Precise Positioning with GPS, Rockville, Vol. 1: 215–226.

Seeber, G., et al. (1987): *Status Report on DÖNAV*. Symp. "Impact of Global Positioning System on Geophysics". IUGG Gen. Assembly, Vancouver.

Seeber, G., Hinze, H. (1984): *Bestimmung von Gletscherbewegungen mit Doppler-Satellitenmessungen in der Antaktis*. In: Schödlbauer, Welsch (eds.): Satelliten-Dopplermessungen. Studiengang Vermessungswesen, Schriftenreihe HSBw, Nr. 16, München.

Seeber, G., Menge, F., Völksen, C., Wübbena, G., Schmitz, M. (1997a): *Precise GPS Positioning Improvements by Reducing Antenna and Site Dependent Effects*. In: Brunner (ed.) (1998): 237–244.

Seeber, G., Seeger, H. (1984): *Satelliten-Dopplermessungen im Deutschen Hauptdreiecksnetz*. In: Schödlbauer, Welsch (eds.): Satelliten-Dopplermessungen. Studiengang Vermessungswesen, Schriftenreihe HSBw, Nr. 16, München.

Seeber, G., Torge, W., Denker, H., Goldan, H.-J. (1997b): *Präzise Höhenbestimmung des Helgoländer Pegels*. Die Küste, 59: 39–61.

Seeber, G., Wübbena, G. (1989): *Kinematic positioning with carrier phases and "on the way" ambiguity solution*. Proceed. 5th Int. Geod. Symp. on Satellite Positioning, Las Cruces, Vol. 2: 606–609.

Seeger, H. (1984): *Zur Geoidbestimmung im Alpenraum mit Hilfe von Doppler-Satellitenmessungen im NNSS (ALGEDOP)*. In: Schödlbauer, Welsch (eds.): Satelliten-Dopplermessungen, Studiengang Vermessungswesen, Schriftenreihe HSBw, Nr. 15, 249–266.

Seemüller, W., Drewes, H. (2000): *Annual Report 1999 of RNAAC SIRGAS*. 1999 Technical Reports, IGS Central Bureau.

Segawa, J., Fujimoto, H., Okubo, S. (eds.) (1997): *Gravity, Geoid and Marine Geodesy*. IAG Symp. Proceed. 117, Springer-Verlag, Berlin-Heidelberg-New York.

Seidelmann, P. (ed.) (1992): *Explanatory Supplement to the Astronomical Almanac*. U.S. Naval Observatory, Washington, D.C.

Seidelmann, P., Fukushima, T. (1992): *Why new time scales?* Astron. Astrophys. 265: 833–838.

Seidelmann, P., Guinot, B., Doggelt, L. (1992): *Time*. In: Seidelmann (ed.) (1992) Chap. 2: 39–94.

Seidelmann, P., Johnston, K., et al. (1998): *Fizeau Astrometric Mapping Explorer (FAME) Satellite*. AAS 98-154: 763–782.

Shapiro, I. (1978): *Principles of Very-Long-Baseline Interferometry*. Proceed. 9th GEOP Conf., OSU Rep. 280, 29–33.

Shelus, P., Veillet, C., Dickey, J. (1996): *Recent Developments in Lunar Laser Ranging*. CSTG Bull. 12, 47–52, Munich.

Shum, C., Yuan, D., Ries, J., Smith, J., Schutz, B., Tapley, B. (1990): *Precision orbit determination for the GEOSAT Exact Repeat Mission*. JGR 95 (C3): 2887–2898.

Sidi, M. (1997): *Spacecraft Dynamics and Control*. Cambridge Aerospace Series 7, Cambridge University Press.

Sigl, R. (1984): *The contribution of satellite geodesy to the geosciences*. Geo Journal 8: 341–362.

Sinclair, A. (1999): *Data Screening and Normal Point Formation*. International Laser Ranging Service.

Sinnott, R. (2001): *The Best Star Catalogue Ever*. Sky & Telescope July 2001, 22–23.

Sjöberg, L. (1999): *An efficient iterative solution to transform rectangular geocentric coordinates to geodetic coordinates*. ZfV 124: 295–297.

Sjögren, D., Gottlieb, P., Muller, P., Wollenhaupt, W. (1972): *Lunar gravity via Apollo 14 Doppler radio tracking*. Science 175: 165–168.

Slater, J., Malys, S. (1997): *WGS 84 - Past, Present and Future*. In: Brunner (ed.) (1998): 1–7.

Slater, J., Noll, C., Gowey, K. (eds.) (1999): *International GLONASS Experiment IGEX-98*. Proceed. IGEX-98 Workshop, IGS Central Bureau.

Smart, W. (1977): *Textbook on Spherical Astronomy*. 6th ed., University Press, Cambridge.

Smith, D., et al. (1985): *A global geodetic reference frame from LAGEOS ranging (SL5.1AP)*. JGR 90: 9221–9233.

Smith, D., Kolenkiewicz, R., Dunn, P., Torrence, M. (2000): *Earth scale below a part per billion from Satellite Laser Ranging*. In: Schwarz (ed.) (2000): 3–12.

Smith, D., Milbert, D. (1999): *The GEOID96 high-resolution geoid height model for the United States*. J. Geod. 73: 219–236.

Snay, R. (2000): *The National and Cooperative CORS System in 2000 and Beyond*. Proceed. ION GPS-2000: 55–58.

Sneeuw, N., Doborantu, R., Gerlach, C., Müller, J., Oberndorfer, H., Rummel, R., 5 more authors (2001): *Simulation of the GOCE gravity field mission*. In: Proceed. IV. Hotine-Marussi Symposium on Mathematical Geodesy, IAG Symp. Vol. 122, 14–32, Springer-Verlag, Berlin-Heidelberg-New York.

Sneeuw, N., Ilk, K. (1997): *The status of spaceborne gravity field mission conceps: a comparative simulation study*. In: Segawa, Fujimoto, Okubo (eds.) (1997): 171–178.

Soddu, C., Razumovsky, O. (2001): *GPS Augmentation – Inmarsat's New Navigation Payload*. GPS World, 12 (11): 24–32.

Soffel, M., Müller, J. (1997): *Lasermessungen der Monddistanz*. Sterne und Weltraum, 36: 646–651.

Soop, E. (1983): *Introduction to geostationary orbits*. ESA SP-1053.

Sovers, O., Fanselow, J., C.S.Jacobs (1998): *Astrometry and geodesy with radio interferometry: experiments, models, results*. Reviews of Modern Physics, 70 (4): 1393–1454.

Spilker, J. (1980): *GPS signal structure and performance characteristicas*. Navigation (USA) 25 (1978). In: Janiczek (ed.) (1986), Vol. 1: 29–54.

Spilker, J. (1996a): *Foliage Attenuation for Land Mobile Users*. In: Parkinson, Spilker (eds.) (1996), Vol. 1, Chap. 15: 569–583.

Spilker, J. (1996b): *GPS Signal Structure and Theoretical Performance*. In: Parkinson, Spilker (eds.) (1996), Vol. 1, Chap. 3: 57–119.

Spilker, J. (1996c): *Satellite Constellation and Geometric Dilution of Precision*. In: Parkinson, Spilker (eds.) (1996), Vol. 1, Chap. 5: 177–208.

Spilker, J. (1996d): *Tropospheric Effects on GPS*. In: Parkinson, Spilker (eds.) (1996), Vol. 1, Chap. 13: 517–546.

Stanley, H. (1979): *The Geos-3 project*. JGR 84 (B8): 3779–3783.

Stansell, T. (1979): *The MX 1502 Satellite Surveyor*. 2nd Int. Geod. Symp. on Satellite Doppler Positioning, Austin, Vol. 1: 497–534.

Stewart, M., Tsakari, M. (1998): *GLONASS Broadcast Orbit Computation*. GPS Solutions 2: 16–27.

Stolz, A. (1979): *Precise modeling aspects of lunar measurements and their use for the improvement of geodetic parameters*. DGK A 90, München.

Strang, G., Borre, K. (1997): *Linear Algebra, Geodesy, and GPS*. Wellesley-Cambridge Press.

Stumpff, K. (1959/1965/1974): *Himmelsmechnik I–III*. 3 Vol., VEB Deutscher Verlag der Wissenschaften, Berlin.

Swift, E. (1985): *NSWC's GPS Orbit/Clock determination system*. Proceed. 1st Int. Symp. on Precise Positioning with the GPS System, Rockville, Vol. 1, 59–62.

Taff, L. (1985): *Celestial Mechanics, a computational guide for the practitioner*. John Wiley & Sons, New York.

Tapley, B., Kim, M. (2001): *Applications to Geodesy*. In: Fu, Cazenave (eds.) (2001), Chap. 10: 371–406.

Tapley, B., Watkins, M., Ries, J., et al. (1996): *The Joint Gravity Model 3*. JGR 101 (B12): 28 029–28 049.

Tavernier, G., Larson, K., Noll, C., Ries, J., Soudarin, L., Willis, P. (2000): *Current status of the DORIS pilot experiment*. Proceed. DORIS days 2000.

Tetewsky, A., Mullen, F. (1997): *Carrier Phase Wrap-up Induced by Rotating GPS Antennas*. GPS World 8 (2): 51–57.

Teunissen, P. (1998): *GPS Carrier Phase Ambiguity Fixing*. In: Teunissen, Kleusberg (1998), Chap. 8: 319–388.

Teunissen, P., de Jonge, P., Tiberius, C. (1995): *A New Way to Fix Carrier Phase Ambiguities*. GPS World 6 (4): 58–61.

Teunissen, P. J. G., Kleusberg, A. (1998): *GPS for Geodesy*. Second Edition. Springer-Verlag, Berlin-Heidelberg-New York.

Torge, W. (1980): *Geodätisches Datum und Datumstransformation*. In: Pelzer (ed.), Geodätische Netze in Ingenieur- und Landesvermessung, 113–130, K. Wittwer, Stuttgart.

Torge, W. (1989): *Gravimetry*. W. de Gruyter, Berlin-New York.

Torge, W. (1991): *Geodesy*. 2nd. ed., W. de Gruyter, Berlin-New York.

Torge, W. (2001): *Geodesy*. 3rd. ed., W. de Gruyter, Berlin-New York.

Torge, W. (2003): *Geodäsie*. 2nd. ed., W. de Gruyter, Berlin-New York.

Torrence, M. (1999): *Realization of the EGM96 Reference Frame*. In: Schlüter, Schreiber, Dassing (1999), Vol. 2: 777–785.

Touboul, P., Willemenot, E., Josselin, B. F. V. (1998): *Accelerometers for CHAMP, GRACE and GOCE space missions: synergy and evolution*. Boll. Geof. Teor. Appl. 40: 321–327.

Trimmer, R., Manning, D. (1996): *The altimetry derived gravity anomalies to be used in computing the DMA/NASA Earth gravity model*. In: Rapp, Cazenave, Nerem (eds.) (1996): 71–81.

Tscherning, C., Rapp, R., Goad, C. (1983): *A Comparison of Methods for Computing Gravimetric Quantities from High Degree Spherical Harmonic Expansions*. Man. geod. 8: 249–272.

Tsui, J. B.-Y. (2000): *Fundamentals of Global Positioning System Receivers: a Software Approach*. John Wiley & Sons.

Tucker, A. (1998): *Computerized Ionospheric Tomography*. Johns Hopkins APL Technical Digest, Vol. 19, No. 1: 66–71.

Turcotte, D., Schubert, G. (2002): *Geodynamics*. 2nd ed., Cambridge University Press, Cambridge UK.

Väisälä, Y. (1946): *An astronomical method of triangulation*. Sitz. Finn. Akad. Wiss., 99–107.

Van Dierendonck, A. (1995): *Understanding GPS Receiver Terminology: A Tutorial*. GPS World 6 (1): 34–44.

Van Dierendonck, A., Hegarty, C. (2000): *The New L5 Civil GPS Signal*. GPS World 11 (9): 64–71.

Van Dierendonck, A., Russel, S., Kopitzke, E., Birnbaum, M. (1980): *The GPS Navigation Message*. Navigation 25 (2) (1979) also in: Janiczek (ed.) (1986), Vol. 1: 55–73.

Van Diggelen, F. (1998): *GPS Accuracy: Lies, Damn Lies, and Statistics*. GPS World 9 (1): 41–48.

Van Melle, M. (1990): *Cesium and rubidium standards and performance on the GPS program*. In: ION GPS-90: 123–130.

Vandenberg, N., Baver, K. (eds.) (2000): *International VLBI Service 2000 General Meeting Proceedings*. NASA/CP-2000-209893.

Vandenberg, N., Baver, K. (eds.) (2002): *International VLBI Service for geodesy and astrometry 2001 Annual Report*. NASA/TP-2002-210001.

Vaníček, P., Krakiwsky, E. (1986): *Geodesy: The Concepts*. 2. rev. ed., Elsevier Science Publ., Amsterdam-New York.

de Vegt, C. (1999): *The New Astronomical Reference Frame*. Mitt. Bundesamt Kart. Geod. 5, 65–69, Frankfurt a. M.

Veillet, C. (ed.) (1989): *Seventh International Workshop on Laser Ranging Instrumentation*. Matera, Italy, 2-8 Oct. 1989. OCA/CERGA Av. Copernic-F06130, Grasse.

Vinti, J. (1998): *Orbital and Celestial Mechanics*. Edited by G. J. Der and N. L. Bonavito. American Institute of Aeronautics and Astronautics, Reston, Virginia.

Visser, P., Wakker, K., Ambrosius, B. (1994): *Global gravity field recovery from the ARISTOTE-LES satellite mission*. JGR 99 (B2): 2841–2851.

Völksen, C. (2000): *Die Nutzung von GPS für die Deformationsanalyse in regionalen Netzen am Beispiel Islands*. Wiss. Arb. Univ. Hannover, No.237.

Völksen, C., Seeber, G. (1998): *Nachweis von rezenten Krustendeformationen in Nordisland mit GPS*. ZfV 123: 68–75.

Vollath, U., Birnbach, S., Landau, H., Fraile-Ordonez, J., Martin-Neira, M. (1999): *Analysis of three-carrier ambiguity resolutiontechnique for precise relative positioning in GNSS-2*. Navigation 46 (1): 13–24.

Vollath, U., Buecherl, A., Landau, H., Pagels, C., Wagner, B. (2000): *Multi-Base RTK Posiitoning Using Virtual Reference Stations*. Proceed. ION GPS-2000: 123–131.

Vonbun, F. (1977a): *Goddard laser systems and their accuracies*. Phil. Trans. Roy. Soc. London, A 284: 443–450.

Vonbun, F. (1977b): *Probing the Earth's gravity field by means of satellite-to-satellite tracking*. Phil. Trans. Roy. Soc. London, A 284: 475–483.

Vyskocil, P., Reigber, C., Cross, P. (eds.) (1990): *Global and Regional Geodynamics*. IAG Symp. No.101, Edinburgh 1989, Springer-Verlag, New York-Berlin.

Wahr, J. (1981): *The forced nutation of an elliptical, rotating elastic and oceanless Earth*. Geophys. J. R. Astr. Soc. 64: 705–727.

Wakker, K., Ambrosius, B., Zandbergen, R., Geldrop, G. V. (1987): *Precise orbit computation gravity model adjustment and altimeter data processing for the ERS-1 altimetry mission*. ESA Report, Delft.

Walker, R. (2000): *Astronomical VLBI: Comparison and Contrast with Geodetic/Astrometric VLBI*. In: Vandenberg, Baver (eds.) (2000): 42–51.

Walker, R., Kubik, K. (1996): *Numerical Modelling of GPS Signal Propagation*. Proceed. ION GPS-96: 709–717.

Walter, H., Sovers, J. (2000): *Astrometry of Fundamental Catalogues*. Springer-Verlag, Berlin-Heidelberg-New York.

Wanninger, L. (1992): *Monitoring total electron content and ionospheric irregularities with GPS*. In: DeMunck, Spoelstra (eds.) (1992): 141–146.

Wanninger, L. (1994): *Der Einfluss der Ionosphäre auf die Positionierung mit GPS*. Wiss. Arb. Univ. Hannover, Nr. 201.

Wanninger, L. (1998): *Real-time Differential GPS Error Modelling in Regional Reference Station Networks*. In: Brunner (ed.) (1998): 86–92.

Wanninger, L. (2000): *Präzise Positionierung in regionalen GPS-Referenznetzen*. DGK C 508, München.

Wanninger, L., Jahn, C.-H. (1991): *Effects of severe ionospheric conditions on GPS data processing*. IAG Symp. G-2 "Permanent Satellite Tracking Networks for Geodesy and Geodynamics", IUGG General Assembly, Vienna.

Watkins, M., Eanes, R., Tapley, B., Schutz, B. (1990): *Station positions and plate motion from LAGEOS long arc LLA8903*. In: Vyskocil, Reigber, Cross (eds.) (1990): 1–10.

Webb, F., Zumberge, J. (1993): *An Introduction to GIPSY/OASIS-II*. Jet Propulsion Laboratory, Pasadena, USA.

Webb, J., Penney, R. (1981): *Six years of inertial surveying at the Geodetic Survey in Canada*. Proc. 2nd Int. Symp. of Inertial Technology for Surveying and Geodesy, 325–342, Banff.

Weber, G. (2002): *DGPS/RTK-Dienste über Internet*. Proceed. 4. SAPOS Symposium, Hannover 2002: 54–61.

Weber, L., Tiwari, A. (1995): *DGPS Architecture Based on Separating Error Components, Virtual Reference Stations, and FM Subcarrier Broadcast*. Proceed. 51st Annual Meeting of ION, Colorado Springs: 191–200.

Webster, I., Kleusberg, A. (1992): *Regional modelling of the ionosphere for single frequency users of the Global Positioning System*. Proceed. 6th Int. Symp. Satellite Positioning, Columbus, Ohio, Vol. 1: 230–239.

Weill, L. (1997): *Conquering Multipath: The GPS Accuracy Battle*. GPS World 8 (4): 59–66.

Wells, D. (1974): *Doppler Satellite Control*. UNB Techn.Rep. 29, Fredericton N.B., Canada.

Wells, D. (ed.) (1986): *Guide to GPS Positioning*. Fredericton N.B., Canada.

Wenzel, H. (1985): *Hochauflösende Kugelfunktionsmodelle für das Gravitationspotential der Erde*. Wiss. Arb. Univ. Hannover, Nr. 137.

Wenzel, H. (1999): *Schwerefeldmodellierung durch ultra-hochauflösende Kugelfunktionsmodelle*. ZfV 124: 144–154.

Westrop, J., Napier, M., Ashkenazi, V. (1990): *The use of phase for kinematic positioning by GPS*. In: Bock, Leppard (eds.) (1990): 334–339.

White, L. (1980): *Spheroidal and Cartesian Coordinates*. Survey Review: 285–286.

Wickert, J., Reigber, C., Beyerle, G., 8 more authors (2001): *Atmospheric sounding by GPS radio occultation: First results from CHAMP*. Geophys. Res. Lett.

Wild, U. (1994): *Ionospheric and Geodetic Satellite Systems: Permanent GPS Tracking Data for Modelling and Monitoring*. Geod. Geophys. Arb. Schweiz, Schweizerische Geodätische Kommission, Zürich, Vol. 48.

Wilkins, G., Mueller, I. (1986): *On the rotation of the Earth and the terrestrial reference system*. Bull. Géod. 60: 85–100.

Willgalis, S., Seeber, G., Krueger, C., Romao, V. (2002): *A Real Time GPS Reference Network for Recife, Brazil, Enabling Pricise and Reliable Cadastral Surveys*. FIG XXII International Congress, Washington D.C.

Williams, J., Newhall, X., Dickey, J. (1987): *GM (Earth) from Lunar Laser Ranging (LLR)*. EOS 68 (16): 281.

Williamson, R., Marsh, J. (1985): *Starlette Geodynamics: The Earth's tidal response*. JGR 90 (B11): 9346–9352.

Wilson, P., Michel, G. (eds.) (1998): *The Geodynamics of S and SE Asia (GEODYSSEA) Project*. Scientific Technical Report STR 98/14, GeoForschungsZentrum Potsdam 1998.

Wohlleben, R., Mattes, H., Krichbaum, T. (1991): *Interferometry in radioastrometry and radar techniques*. Kluwer Academic Publishers, Dordrecht.

Wolf, H. (1967): *Computation of satellite triangulation by spatial fitting of the orbit*. DGK B 153, München.

Wolf, H. (1970): *Short Arc Methode und räumliche Bahnglättung in der Satellitentriangulation*. ZfV 94: 37–41.

Wolf, H. (1975): *Ausgleichungsrechnung – Formeln zur praktischen Anwendung*. Dümmler, Bonn.

Wolf, H. (1984): *Satelliten-Dopplermessungen in der Berechnung des Europäischen Dreiecksnetzes (RETrig)*. In: Schödlbauer, Welsch (ed.): Satelliten-Dopplermessungen, Studiengang Vermessungswesen, Schriftenreihe HSBw, Nr. 15: 221–230.

Wolf, H. (1985): *Fortschritte der Geodäsie: Satelliten- und terrestrische Methoden mit ihren Möglichkeiten*. Vorträge Rhein.Westf. Akad. Wiss., Nr.338, Opladen.

Wood, R. (1999): *Improving Orbit Predictions*. In: Schlüter, Schreiber, Dassing (1999) Vol. 2: 615–620.

Wübbena, G. (1985): *Model and program developments for cm-positioning with GPS*. Proceed. FIG Joint Meeting Inertial, Doppler and GPS Measurements for National and Engineering Surveys, 553–565, München.

Wübbena, G. (1988): *GPS carrier phase and clock modeling*. In: Groten, Strauss (eds.) (1988): 381–392.

Wübbena, G. (1989): *The GPS adjustment software package GEONAP, Concepts and models*. Proceed. 5th Int. Geod. Symp. Satellite Positioning, Las Cruces, Vol. 1: 452–461.

Wübbena, G. (1991): *Zur Modellierung von GPS Beobachtungen für die hochgenaue Positionsbestimmung*. Wiss. Arb. Univ. Hannover, Nr. 168.

Wübbena, G. (2002): *Zu großräumigen Vernetzung von GNSS Referenzstationen*. Proceed. 4. SAPOS Symposium, Hannover 2002: 139–148.

Wübbena, G., Bagge, A., Boettcher, G., Schmitz, M., Andree, P. (2001a): *Permanent Object Monitoring with GPS with 1 Millimeter Accuracy*. Proceed. ION GPS-2001, Salt Lake City.

Wübbena, G., Bagge, A., Schmitz, M. (2001b): *RTK Networks based on Geo++ GNSMART – Concepts, Implementation, Results*. Proc. ION GPS-2001, Salt Lake City.

Wübbena, G., Bagge, A., Seeber, G., Böder, V., Hankemeier, P. (1996): *Reducing Distance Dependent errors for Real-Time Precise DGPS-Applications by Establishing Reference Station Networks*. Proceed. ION GPS-1996: 1845–1852.

Wübbena, G., Menge, F., Schmitz, M., Seeber, G., Völksen, C. (1997): *A New Approach for Field Calibation of Absolute Antenna Phase Center Variations*. Journal Navigation, Vol. 44 (2): 247–255.

Wübbena, G., Schmitz, M., Menge, F., Böder, V., Seeber, G. (2000): *Automated Absolute Field Calibration of GPS Antennas in Real-Time*. Proceed. ION GPS-2000: 2514–2522.

Wübbena, G., Willgalis, S. (2001): *State Space Approach for Precise Real Time Positioning in GPS Reference Networks*. Proceed. KIS 2001, Banff, Canada.

Wunsch, C., Stammer, D. (1997): *Atmospheric loading and the oceanic "inverted barometer" effect*. Rev. Geophys. 35: 79–107.

Young, C., Neilan, R., Bletzacker, F. (1985): *GPS satellite multipath: An experimental investigation*. Proceed. 1st Int. Symp. Precise Positioning with GPS, Rockville: 423–432.

Yunck, T. (1996): *Orbit Determination*. In: Parkinson, Spilker (eds.) (1996), Vol. 2, Chap. 21: 559–592.

Yunck, T., Melbourne, W. (1990): *Geoscience from GPS tracking by Earth satellite*. In: Bock, Leppard (eds.) (1990): 351–369.

Zacharias, N., Urban, S., Zacharias, M., Hall, D., Wycoff, G., Rafferty, T., Germain, M., Holdenried, E., J.Pohlmann, Gauss, F., Monet, D., Winter, L. (2000): *The First United States Naval Observatory CCD Astrograph Catalog*. The Astronomical Journal, Vol. 120.

Zahran, K. (2000): *Accuracy Assesment of Ocean Tide Loading Computations for precise Geodetic Observations*. Wiss. Arb. Univ. Hannover, Nr. 238.

Zhang, S. (2000): *Interpolation of Geoidal/Quasigeoidal Surfaces for Height Determination with GPS*. Wiss. Arb. Univ. Hannover, Nr. 236.

Ziebart, M., Cross, P., Adhya, S. (2002): *Modeling Photon Pressure – The Key to High-Precision GPS Satellite Orbits*. GPS World 13 (1): 43–50.

Zielinski, J. (1989): *GPS baseline error caused by the orbit uncertainty*. Man. geod. 14: 117–124.

Zwally, H., Brenner, A. (2001): *Ice Sheet Dynamics and Mass Balance*. In: Fu, Cazenave (eds.) (2001), Chap. 9: 351–369.

Index