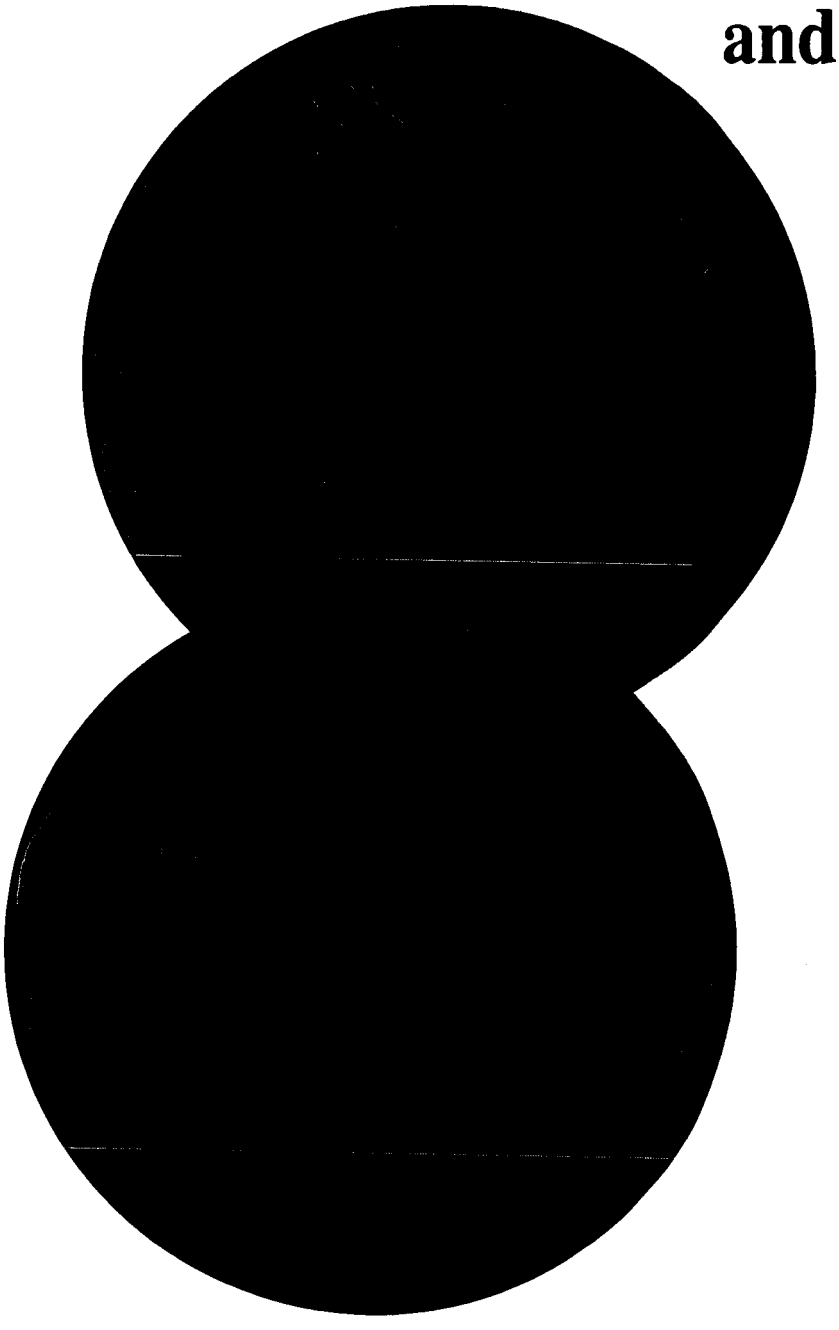
JPL Publication 94-18

165 P.

The Application of Spaceborne GPS to Atmospheric Limb Sounding and Global Change Monitoring



W. G. Melbourne
E. S. Davis
C. B. Duncan
G. A. Hajj
K. R. Hardy
E. R. Kursinski
T. K. Meehan
L. E. Young
T. P. Yunck

April 1994

NASA
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

CONTENTS

FIGURES

TABLES

PREFACE

Future commentators on the utilization of space may well refer to the 1990s as the decade of the microsats. In just a few years, we have witnessed a remarkable change in philosophy—a new paradigm in a sense—toward the capitalization, management, development, deployment, and operation of space assets for remote sensing, communications, and navigation. "Bigger is better" is out, at least in some circles, and "smaller is beautiful" is in.

Among the reasons for the shift in our approach to space exploration, particularly for Earth orbiter applications, are (a) recent technological innovations and attendant cost reductions, notably in digital circuitry, small satellites, and launch systems, (b) an expanding global market for space-based services, (c) the marked reduction in Cold War tensions and the consequent accessibility to the commercial world of key Department of Defense (DoD)-developed technologies and space system elements, and (d) shrinking federal budgets.

Although the ramifications of this revolution are only now becoming discernible, the new millennium should be ushered in by a virtual explosion in the number and diversity of applications of Earth orbiters and supporting ground information systems that are being implemented by organizations within a number of countries. Within the United States, dedicated constellations of small and moderate-sized satellites are now being considered for deployment in the late 1990s: IRIDIUM, Globalstar, Odyssey, and the Teledesic Constellation. These should be contrasted with the very large and enormously expensive multipurpose platforms such as Space Station and Earth Observation System (EOS). For Earth orbiters, NASA's "faster-cheaper-better" aphorism aptly describes the ongoing restructuring of the utilization of space, except that it is no longer as centralized within the province of a few capital-intensive federal agencies and their defense/space contractors. Future diversification and expansion in space should involve a broad participation by industry in all aspects of space systems (and result in a corresponding increase in economic risks and returns for that sector). New space ventures should benefit from partnerships and consortia involving industry, federal agencies, and leading-edge research laboratories such as the Jet Propulsion Laboratory (JPL).

We describe in this monograph a particular remote-sensing application that is well suited to this new space milieu: the radio occultation technique. This technique was pioneered by JPL and Stanford University over 30 years ago for study of planetary atmospheres and ionospheres. We describe a satellite limb-sounding mission concept

that uses the Global Positioning System (GPS) to probe the Earth's atmosphere. This technique offers an innovative approach for monitoring global atmospheric temperatures, pressures and moisture distributions with high accuracy and spatial resolution and for studying the ionosphere.

The GPS receiver aboard a microsatellite in a low Earth orbit (LEO) tracks the high precision L-band navigation signals from the constellation of GPS satellites. The GPS satellites broadcast on two coherent radio frequency carriers that are separated in frequency by 350 MHz. The fundamental measurements made by the receiver are the phase and signal strength of each carrier, which are sampled at a rate up to 50 Hz.

When the LEO receiver tracks a GPS satellite that is observed to rise or set through the Earth's atmosphere (that is, a satellite being occulted by the Earth), the arrival time of the received radio signal is retarded because of the refractive bending and slowing of the signal as it traverses the atmosphere. By measuring the change in carrier phase over the entire occultation event, about 1 minute in duration for the neutral atmosphere, we can determine the atmospheric refractive index as a function of altitude. Pressure and temperature profiles can then be derived through a downward integration using the gas law, the known linear relationship between refractivity and density of dry air, and the assumption of hydrostatic equilibrium. This technique also requires (for satellite ephemeris control and elimination of oscillator errors) a globally distributed ground network of GPS receivers that are concurrently tracking all observable GPS satellites. The data streams from the ground network and the LEO are synchronously combined (by interpolation) at a central facility and processed in parallel to form high accuracy positions and velocities of all the satellites including the LEO, and generate highly precise differential phase residuals of the retarded signal arriving at the LEO from the occulted GPS satellite. All instrumental biases are eliminated because the concurrent and synchronous observation set enables us to remove the effect of instrumental (transmitter and receiver) phase delay and oscillator errors in the measurements and because signal processing circuitry in the receiver is digital.

A single LEO observing both rising and setting satellites would provide about 500 globally distributed temperature profiles daily with sub-kelvin accuracies over altitudes of ~8–45 km. The vertical resolutions would be comparable to those from the best limb sounders (<1 km). The technique is essentially all-weather and unaffected by clouds and particles.

Below the tropopause, the presence of water vapor limits the accuracy of temperature recovery because of uncertainty in the distribution and relative abundance of the water vapor in the troposphere. The refractivity of a mole of water vapor for microwaves is about 17 times that of dry air. On the other hand, we can exploit this high sensitivity to recover water vapor profiles in the troposphere by using meteorological models and other sensor information for temperature and pressure to separate the dry air and water vapor signatures on the tracking data. With this approach, the LEO can also chart much-needed water vapor distributions in the lower troposphere, particularly over tropical oceanic areas.

From the time series of the difference in phase on the two L-band carriers that are measured by the LEO receiver, the change in the integrated ionospheric electron content along the line between the GPS and LEO satellites can be precisely determined (to less than 0.1% of the total). With this technique, tomographic studies of ionospheric electron distributions can also be undertaken with

the combined LEO space platforms and GPS ground tracking network.

Because the GPS radio occultation technique is free of instrumental biases, it can also serve as a calibration system. An operational radio occultation system that provides long-term coverage will furnish an absolute standard for calibrating other global change monitoring instruments to ensure long-term stability in their atmospheric measurement series.

A virtue of this mission concept that sets it apart from most remote-sensing systems under consideration for global change monitoring is its potential for very low cost (less than approximately $5M) per satellite. This is achieved through the use of small, dedicated satellites and through the exploitation of small payload launch opportunities. The low cost and short deployment cycle for this class of mission arise from several factors: (a) the maturity and low cost of microsat technology, (b) the proliferation of commercial enterprises engaged in microsat and launch systems, (c) the enormous progress in GPS-specific very large-scale integration (VLSI) circuitry, and (d) the high degree of heritage and synergy in GPS and radio occultation data acquisition and processing systems provided by related programs such as NASA's Geodynamics program. Because of low cost and modest payload requirements, a number of innovative risk management strategies are economically feasible, making inexpensive (as low as $200k) launch opportunities possible. Costs of operations and occultation data processing systems are also low because of modest data rates (~65 Mbytes/day/LEO) and because these systems build on an existing 30-year heritage in radio occultation data processing and use established automation techniques.

A single LEO continuously operating over a nominal 5-year lifetime would return about 1 million occultations, which works out to roughly $10/occultation, including the costs of operations and data reduction. This cost figure can be compared to the average cost of a radiosonde, which is about $200, including associated labor costs. There are about 365,000 radiosonde launches made by the United States over a 5-year period.

What about disadvantages of the GPS radio occultation technique? A limb-sounder's accuracy depends, to some degree, on the atmospheric refractivity being invariant along equipotential surfaces,* at least over distances comparable to the horizontal resolution, roughly 300 km. Because water vapor is not easily modeled, recovery of accurate temperature and pressure profiles with the technique is essentially limited to altitudes at and above the upper troposphere, although in very cold geographical regions devoid of water vapor the technique should yield good results down to the Earth's surface. Certainly the lower troposphere in temperate and tropical regions is a formidable medium for a microwave limb-sounder to operate in because of the propensity for severe spatial inhomogeneities in atmospheric refractivity. During an occultation, the amplitude of the received signal is altered through defocusing and multipath effects induced by gradients in atmospheric refractivity, particularly in the lower troposphere. Both the amplitude and phase measurements would be used in this region where spectral analyses can be performed to separate multipath signals and where Fresnel diffraction deconvolution techniques can be applied to achieve in some cases vertical resolutions of tropospheric features that are a small fraction of the first

* For the Earth's atmosphere, the effect of along-path variability in refractivity that is antisymmetric about the geographical point of tangency of the signal path essentially cancels in recovering the refractivity profile; only the symmetric component affects recovery.

Fresnel zone. Although data processing to recover the profiles of atmospheric parameters can be largely automated over most altitudes, the additional processing and analyses required to deal with adverse signal conditions in the lower troposphere may prove to be labor-intensive and therefore relatively expensive. We do not know how the GPS technique will fare (in accuracy, resolution, and labor cost) at these lower altitudes, although we are encouraged by recent results from observations made by a JPL group of some setting GPS satellites from the summit of Mauna Kea in Hawaii.

One of the exciting prospects about the radio occultation technique, if it can be shown to work well in the lower troposphere, is its potential for recovery of water vapor distributions, especially in tropical regions. For example, a constellation of 20 GPS limb sounders would return roughly 10,000 globally distributed vertical water vapor profiles on a daily basis with a 250-km mean geographical spacing between the profiles. This product alone could significantly enhance the accuracy and reliability of weather forecasting.

The "GPS/MET" mission, which is primarily sponsored by the National Science Foundation (NSF) and managed by the University Corporation for Atmospheric Research (UCAR), promises to be the first demonstration mission. It features a short (3-year), end-to-end proof-of-concept deployment cycle with a direct cost of less than $3M to the NSF and its federal partners (the National Oceanic and Atmospheric Administration and the Federal Aviation Administration). Orbital Sciences Corporation (with NASA support) is providing the spacecraft, the Pegasus launch vehicle, and the mission operations. The currently planned launch date is fall of 1994. The radio occultation concepts in the GPS/MET mission are largely based on the proposal presented by JPL in 1988 to NASA for the GPS Geosciences Instrument (GGI) for flight on EOS. The mission will use the TurboRogue GPS receiver, which was developed by JPL under sponsorship by NASA's Geodynamics Program. The receiver hardware has been modified by its commercial manufacturer, Allen Osborne Associates Inc., to improve its reliability (to Class D) as a flight instrument. The receiver software has been upgraded by JPL under contract to UCAR to make it more suitable for space operations. NASA (the Solid Earth Science Branch and the Physical Climate Branch, respectively) is sponsoring JPL's deployment and operation of the ground network, and the development of the data reduction software for recovery of atmospheric profiles. Although scheduled to operate for only a few weeks to a few months at most, GPS/MET offers a potentially high science payoff by enabling atmospheric and radio scientists to gauge the scientific value and operational utility of radio occultation measurements that might be forthcoming from a permanent constellation of continuously operating limb-sounding LEOs.

Over the next few years other promising missions for demonstrating the GPS radio occultation technique are likely. These include Ørsted, a Danish/French magnetometer mission with a launch planned for late 1995, which will also include an upgraded TurboRogue GPS receiver suitable for occultation measurements; SAC-C, a joint Argentine/U.S. mission planned for 1998. In Europe the Centre National d'Etudes Spatiales (CNES), which sponsored the GLIMPSE concept, and the European Space Agency (ESA) continue to support studies and technology development programs that may lead to the use of the radio occultation technique on small satellites.

It is interesting to speculate why a low-cost remote-sensing system for the study of atmospheric thermodynamic processes, such as the

GPS-based radio occultation technique, has been so long in coming. The history of science is replete with examples of delayed development of key technologies and measurement systems that are somewhat paradoxical. For example, the discovery in 1964 of the cosmic background radiation at 2.7 K occurred by chance, decades after its detection was first technologically feasible and at least a decade after its existence could have been inferred from the Big Bang hypothesis. Radio occultation investigations conducted within the U.S. planetary space program have certainly proved their scientific worth over the past three decades, and yet relatively few atmospheric scientists have embraced the technique and even fewer will be found in the vanguard ranks striving for its early deployment. Our opinion is that the contributions that the radio occultation technique can make to atmospheric science, most notably its potential for achieving extraordinary accuracies in recovery of temperature profiles from the middle stratosphere down to the upper troposphere and the very precise determination of the altitude of the tropopause on each profile, have largely gone unrecognized by the scientific community.

With the benefit of 30 years of hindsight, the root causes for the relatively small constituency among atmospheric scientists can probably be traced to the historical development of the radio occultation technique itself and the nature of its products. NASA's planetary missions of the past have tended to be singular and infrequent events with narrow participation from the scientific community compared to the size of that community at large. Also, and without pejorative intent, the radio occultation investigations on planetary missions principally have been conducted by radio scientists, not atmospheric scientists. Although world-class scientific achievements have resulted from these investigations, the benefits that might have accrued by bringing to bear other cross-cutting disciplines and measurement techniques from atmospheric science and meteorology probably have been limited. Most atmospheric remote-sensing instruments involve power measurements of a radiating or scattering source and usually require some kind of calibration system. Although some interferometric techniques are used in atmospheric science, instruments that measure phase coherent signals from artificial electromagnetic sources, such as lidars and radars, are relatively less familiar to atmospheric scientists. Another reason the radio occultation technique has not been embraced by the atmospheric science community is the technique's perceived character as a "one-product" limb-sounder; it mainly provides vertical profiles of thermodynamic properties of the atmosphere (although other atmospheric processes, for example, acoustic gravity waves, can in fact be indirectly studied with the technique). In contrast, many atmospheric sensors, particularly those providing spectral information from both limb-sounding and nadir-looking devices, recover not only thermodynamic information (albeit not as accurately over most altitudes as those from the radio occultation technique) but also many other important properties such as relative abundances of the chemical constituents, aerosol and cloud particle sizes and distributions, atmospheric dynamics, and so on. On the other hand, these sensor systems tend to be substantially more costly than the radio occultation technique. Perhaps a cost/benefit approach would be a better way to compare these systems. Of course in the case of the Earth there are logistical challenges imposed by the satellite-to-satellite (SST) tracking technique itself, which requires deployment of substantial space assets before a radio occultation experiment is really

feasible. The GPS constellation has only recently become operational, although opportunities for an experiment have existed for well over a decade.

In this monograph my co-authors and I have assumed that our readership primarily comprises atmospheric scientists, related program managers, and newcomers who have little or no background in radio occultation concepts but wish to learn about them. We address most of the salient concepts of a radio occultation mission, including estimated costs for a proof-of-concept mission and for a full-up operational program. The occultation methodology is described along with its major error sources, its accuracy and resolution in recovering temperature and pressure profiles, and its potential application for recovering water vapor distributions in the lower troposphere.

Although our original intent in this monograph was to give a broad and relatively complete treatment of radio occultation concepts that have been developed over the last 30 years—including new techniques suited to the GPS—we are not unmindful of a number of important operational and technical areas that deserve more depth than we were able to provide—for example, data acquisition techniques under adverse signal conditions including multipath situations, scintillations, along-path inhomogeneities in refractivity, advanced Fresnel deconvolution topics, atmospheric dynamics, ray propagation in three dimensions, data processing with constraints from meteorological models and data sets, validation strategies, and mission operation topics.

With regard to the atmospheric and meteorological sciences, it is noted that our underlying philosophy in this monograph toward recovery of atmospheric profiles with the occultation technique has mainly assumed no ancillary information from the other (and extensive) Earth- and space-based programs in these sciences. However, almost certainly many of these programs will provide strong constraints for occultations occuring within their domain. It is likely that a preferable data processing strategy for certain objectives will treat the radio occultation tracking data as new and complementary inputs to existing meteorological and atmospheric dynamical models and data sets and will recover those occultation-related data products within this framework; enhanced weather forecasting could be one of the benefits of this approach.

Limitations in our available time and the need for a timely publication effectively controlled the scope and depth of the subject matter. Perhaps someday we will be able to revise and update this monograph.

William G. Melbourne
Co-Author and Editor-in-Chief
April 11, 1994

ACKNOWLEDGMENTS

We wish to acknowledge the work done by the early researchers, principally at Stanford University and JPL, in developing radio occultation techniques to study the solar system. These researchers include D. L. Cain, V. R. Eshelman, G. Lindal (aka Fjeldbo), A. J. Kliore, and G. L. Tyler, as well as their many collaborators. A significant proportion of the contents of this monograph is based on the work of these individuals and their collaborators. We also wish to acknowledge the help of others who have consulted with us: J. C. Breidenthal, Y. Chiu, C. E. Dunn, M. L. Exner, T. W. Hamilton, A. Hauchecorne, J. A. Klobuchar, R. P. Linfield, C.-H. Liu, D. J. McCleese, R. K. Melbourne, G. M. Resch, L. J. Romans, and J. B. Thomas.

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. A portion of the research appearing in this monograph was done by K. R. Hardy with support from the RDD155 of the Lockheed Independent Research Program, sponsored by Lockheed Missiles and Space Company, Inc. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government, or the Jet Propulsion Laboratory, California Institute of Technology.

CHAPTER

1 OVERVIEW

This monograph is intended for readers with minimal background in radio science who seek a relatively comprehensive treatment of the mission and technical aspects of an Earth-orbiting radio occultation satellite. Part I (Chapters 1–6) describes mission concepts and programmatic information; Part II (Chapters 7–12) deals with the theoretical aspects of analyzing and interpreting radio occultation measurements.

In this mission concept the navigation signals from a Global Positioning System (GPS) satellite that is being occulted by the Earth's limb are observed by a GPS flight receiver on board a low Earth orbiter (LEO) satellite. As the received signal traverses deeper layers of the atmosphere, its amplitude and phase are progressively altered through atmospheric refraction and possibly interference arising from spatial irregularities in refractivity. The signal delay and amplitude profiles recorded by the LEO during the occultation, which typically lasts about 1 min, are used to recover profiles of atmospheric refractivity, pressure, and temperature from the surface up to the upper stratosphere at heights of 50 to 60 km. Below the tropopause, water vapor sharply limits the accuracy of temperature and pressure recovery; on the other hand, nominal profiles for temperature and pressure, particularly over tropical oceanic regions, enable researchers to recover valuable information about water vapor distribution. Radio occultation techniques have been used to probe planetary atmospheres for over three decades; however, a unique application of this technique is used to study the Earth using small, dedicated, and relatively low-cost space systems to carry out the measurement programs.

Chapter 2 summarizes the basic space system concepts of the limb-sounding technique and describes a low-cost strawman demonstration mission. The GPS/MET (meteorological) mission proposed by the University Navstar Consortium (UNAVCO) (Ware 1992) is one realization of this low-cost concept. The proposed GLIMPSE mission (Bauer 1992) is another. Chapter 2 also summarizes the expected performance of the system, including accuracies and resolution. A very brief summary of the heritage in radio science from NASA's planetary program is included.

Chapter 3 discusses some of the scientific benefits of using radio occultation on a suite of small satellites. These benefits include enhancements to short-term weather forecasting by monitoring and charting water vapor distributions in the lower troposphere over tropical oceanic regions, as well as studies of global change by monitoring long-term trends in stratospheric temperatures and correlating these trends with results from other remote sensing systems.

Chapter 4 provides a more detailed discussion of several system elements in a radio occultation

mission, including (1) the launch system for small payloads, (2) the LEO microsat, (3) the GPS constellation, (4) the GPS flight receiver payload, (5) the mission operations ground control and data receiving system, (6) the ground-based GPS global tracking network for precision orbit determination, and (7) the central data processing and archive system.

Chapter 5 addresses the various technology readiness questions that invariably arise. This includes discussions of small payload opportunities, the maturity and commercial viability of the microsat industry, the readiness of flight-rated GPS receivers, and Department of Defense (DoD)-specific issues associated with GPS operations.

Chapter 6 discusses the overall costs of a demonstration mission such as GPS/MET. It also addresses the various synergistic benefits from related programs that could help contain cost growth. Finally, it presents a cost model for an operational constellation of radio occultation satellites and its associated ground system.

Part II explains the theory of the radio occultation technique. Chapter 7 describes a geometrical optics approach to coplanar atmospheric occultation. It discusses the measurements made by the LEO in terms of the refractive properties of the atmosphere and the inversion of these measurement profiles to obtain the refractivity profile. The recovery of the various atmospheric thermodynamic properties including water vapor from the refractivity profile is also discussed.

Chapter 8 addresses major questions regarding accuracy of the occultation technique. It discusses the principal error sources, including signal-to-noise ratio (SNR), multipath, the effects of water vapor on temperature recovery and the converse problem, the limitations of the ionosphere for upper stratospheric observations, and the effects of departures of the atmospheric refractivity from local horizontal isotropy.

Chapter 9 describes some simulations that have been performed to evaluate the sensitivity of the recovered profiles of atmospheric parameters to different error sources, such as departure from spherical symmetry, water vapor, etc.

Chapter 10 discusses horizontal and vertical resolution associated with limb sounders in general.

Chapter 11 treats selected Fresnel diffraction techniques that can be used in radio occultation measurements to sharpen resolution.

Chapter 12 provides brief discussions on selected special topics, such as strategies for handling interference and multipath processes that may arise for rays traveling in the lower troposphere.

It should be noted that a number of important topics in radio occultation are not included in this document. For example, ray-tracing techniques in three dimensions and in multipath situations have been omitted. Also, Fresnel diffraction in three dimensions has been omitted.

CHAPTER

2 INTRODUCTION

Background

Figure 2-1 shows the variation of atmospheric temperature with altitude and pressure for an idealized standard atmosphere without inversion layers, and also displays the boundaries between the principal strata of the atmosphere.

Two key variables in the study of global change processes are global atmospheric temperature and moisture distributions. Although satellite-based radiometry and spectrometry have been used for more than two decades to obtain vertical profiles of atmospheric temperature, their impact on models of global change has been limited. Long-term observations leading to better vertical resolution of temperature and moisture distributions, and to better determination of their temporal and spatial variability, are urgently needed to improve our understanding of the fundamental Earth processes involved in global change. Accordingly, international efforts by the space agencies of the

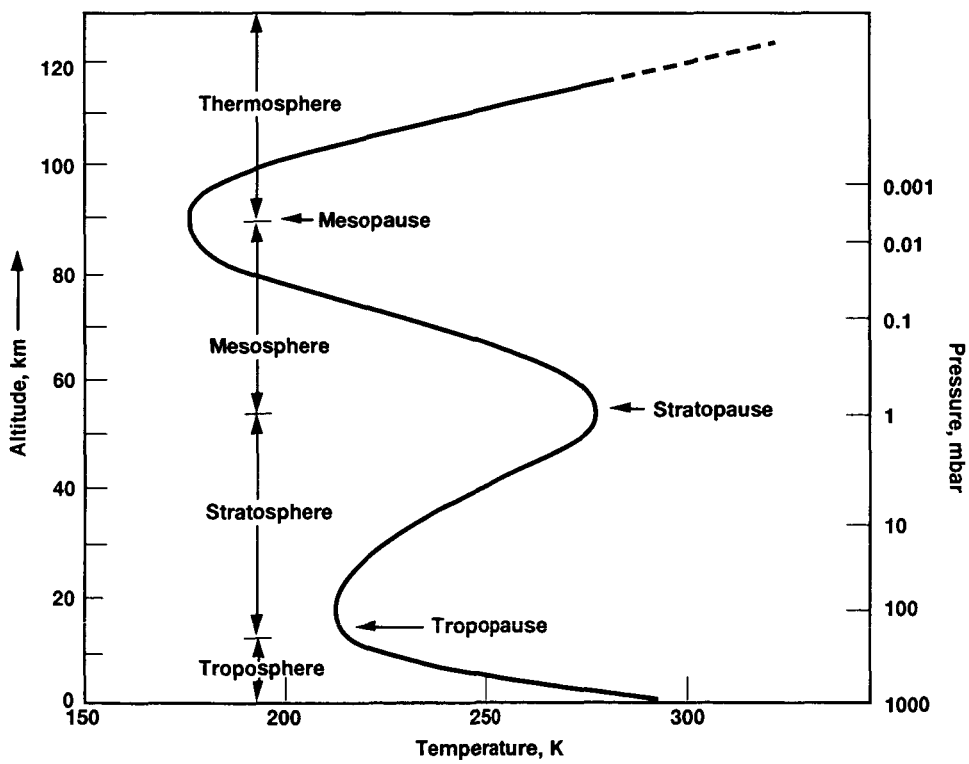


Figure 2-1. Standard atmosphere.

developed nations are now under way to mount comprehensive and long-term remote-sensing and data information system programs for monitoring global change processes relating to atmospheric chemistry, biogeochemistry, energy and mass transport, climate, and solid Earth dynamics. The European Space Agency's (ESA's) Earth Research Satellite (ERS-1) and NASA and the Centre National d'Etudes Spatiales' (CNES's) TOPEX/POSEIDON are current examples of such missions. NASA's Earth Observing System (EOS) and Earth Probe programs, leading to missions in the latter part of the decade and into the twenty-first century, are coordinated approaches to the global change problem.

Big remote-sensing platforms like EOS, carrying a broad suite of sensors, require considerable time for development and deployment—at least a decade before observations begin to flow. The scientific community has pressed for supplementary space and ground programs that would partially fill our information gap in critical areas several years earlier. Some proposals involve "small satellite" missions that target key atmospheric parameters that can be deployed and completed well within the decade and that are relatively "low cost," i.e., less than $100M. The GPS/MET concept is one example of this class of mission; the mission would monitor atmospheric temperature and moisture distributions on a global basis at a very low cost (less than $10M for an end-to-end proof-of-concept mission), and development, deployment, and science demonstration would be completed within three years.

The GPS/MET Mission

Figure 2-2, which is taken from the GPS geoscience instrument (GGI) proposal for NASA's EOS platforms (Melbourne et al. 1988), shows a radio occultation mission that uses the DoD GPS to probe the Earth's atmosphere. Here, vertical profiles of atmospheric refractivity through limb soundings are recovered using the radio occultation technique. GPS/MET is the name for a

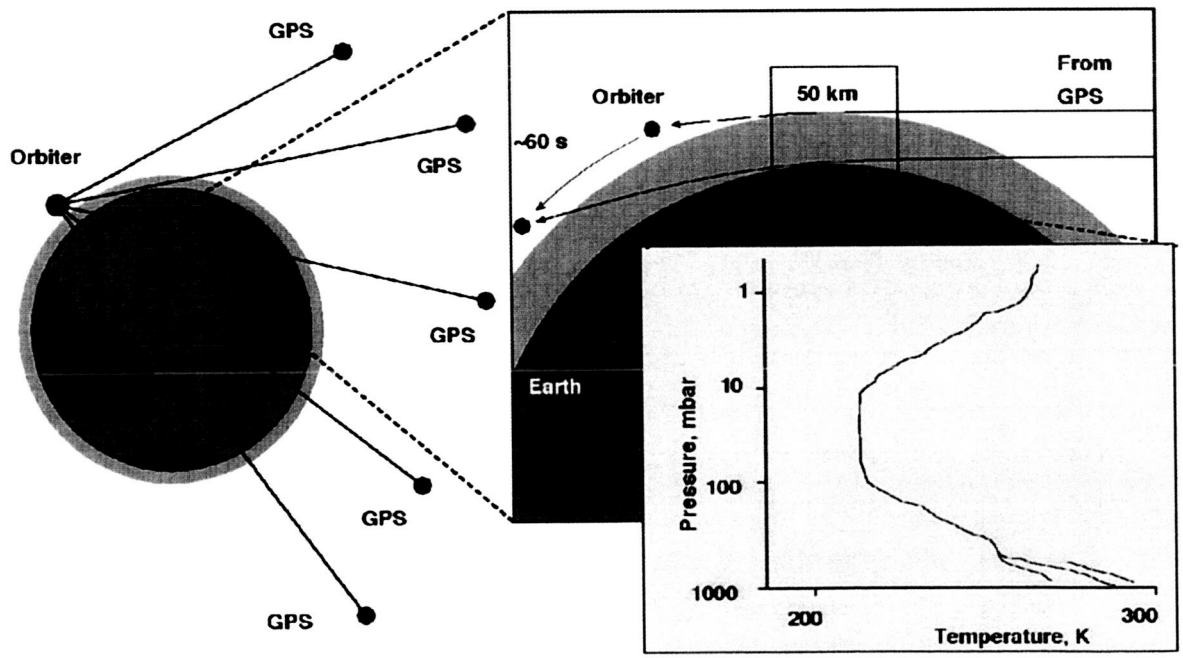


Figure 2-2. Schematic of the satellite-to-satellite geometry for probing the atmosphere of the Earth using the radio occultation technique.

particular limb-sounding mission (Ware 1992) that uses a small dedicated satellite approach but is based upon the GGI concepts. It is noted that the CNES-sponsored GLIMPSE concept discussed in Bauer et al. (1992) possesses essentially the same features as GPS/MET and is also based upon the GGI radio occultation proposal.

Either of these missions as well as GGI on EOS would recover density; pressure and temperature profiles would follow using the known linear relationship between refractivity and the density of dry air, the gas law, and the fact that the atmosphere is essentially in hydrostatic equilibrium. For the lower troposphere, both dry air and water vapor contribute to the refractivity. Water vapor profiles in the lower troposphere would be obtained by using ancillary temperature and pressure information and models to separate water vapor and dry air signatures in the tracking data. Significant temperature information up to altitudes of 60 km would be recovered and sub-kelvin accuracies would be achieved over altitudes of about 8 to 45 km (Figure 2-3) with vertical resolutions of better than 1 km.

Figure 2-3, which characterizes the error in temperature recovery in terms of the limiting error sources, is based on error analyses explained in Chapters 8 and 9. For the lower troposphere, water vapor will mask the dry error refractivity signature except in dry polar regions with cold temperatures. For the upper stratosphere, the signal-to-noise ratio limitations will prevail; in addition, possibly higher order ionospheric effects, which are not completely removed by modeling and the application of the dual-frequency correction, will also limit accuracy.

The mission uses a dual-band GPS receiver on board a small dedicated satellite in an LEO that tracks the navigation signals from all visible GPS satellites, including those that are observed to rise or set through the atmosphere. When the path of the GPS signal begins to transect the mesopause at about a 90-km altitude, it is sufficiently retarded so that a measurable (100 μm) residual in the dual-frequency carrier phase observations is obtained. As the signal path descends through successively thicker layers of the atmosphere, the residual increases to roughly

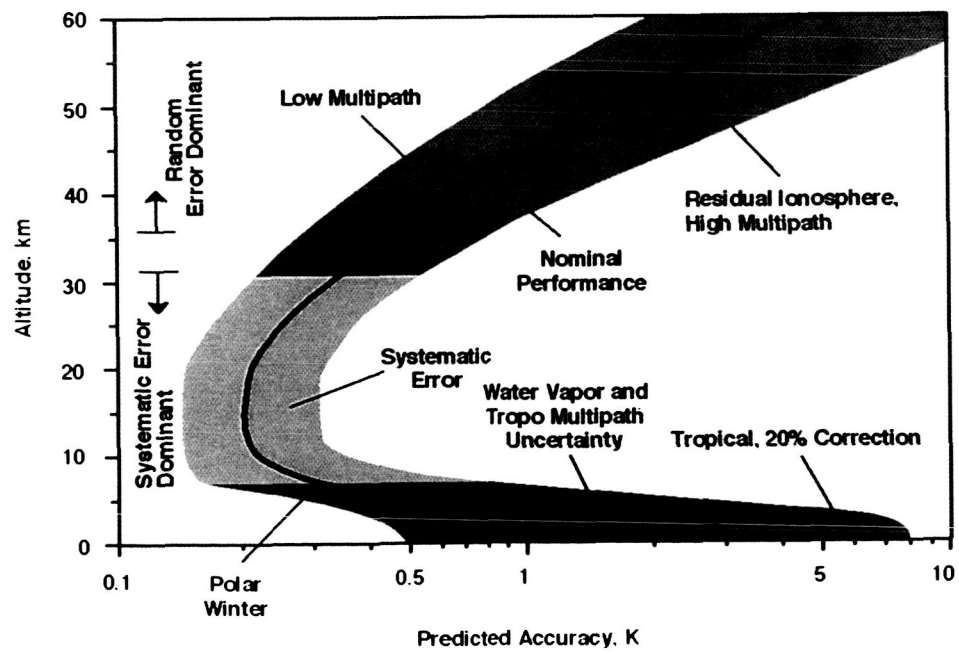


Figure 2-3. Predicted accuracy of temperature profile recovery using GPS/MET.

1 km at the Earth's surface; thus, the atmosphere creates a measurable signal with over six orders of magnitude in dynamic range (Figure 2-4).

A typical setting occultation event lasts 1–2 min, beginning with the signal path passing through the mesopause and ending with it grazing the Earth's limb, or shortly thereafter if diffraction effects are included. About 500 usable and globally distributed occultation events occur per day (Figure 2-5) and are observed by a single LEO in a polar orbit and at an altitude of roughly 800 km. Half of these are rising and half are setting.

A flight-rated GPS receiver is required; this receiver must not only achieve high accuracy in the dual-band carrier phase measurements but provide phase and signal amplitude measurements under adverse signal-tracking conditions (i.e., weak and highly dynamic signals in both phase and amplitude) as the ray passes through the lower troposphere.

Defocusing from refractivity gradients can reduce signal strength by more than one order of magnitude at the Earth's limb. Inhomogeneities in the refractivity of the lower troposphere result in multipath-induced signal interference, which can cause wide excursions in signal amplitude and phase. Multipathing can also lead to multiple tones that must be concurrently "tracked" if the total atmospheric information content in the signal is to be preserved. Dual-band L1 and L2 capability is required to remove ionospheric effects on the carrier phase observations. Dual-band tracking provides a nearly complete calibration of ionospheric effects, but the residual error due to third- and higher order ionospheric delays

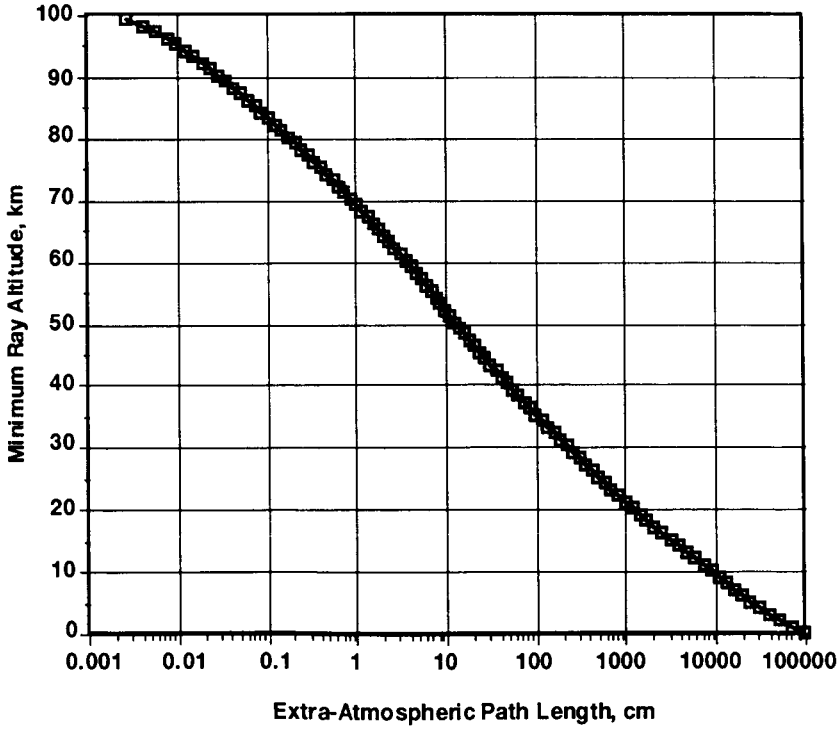


Figure 2-4. Extra-path delay resulting from the GPS signal passing through the atmosphere. U.S. Standard Atmosphere with 50% relative humidity. In addition, during the last seconds before extinction (~10 s), the GPS signal received by the LEO will be accelerated, attenuated by over 10 dB as a result of defocusing, and subject to multipathing.

will limit the accuracy of recovered temperature profiles in the upper stratosphere. Dual-band tracking also yields valuable information about ionospheric electron density distributions and dynamics.

Although GPS receivers used for occultation measurements can and probably will fly on multisensor platforms such as EOS, the great virtue of microsats like GPS/MET or GLIMPSE is that they can be developed and deployed quickly and at a relatively low cost. We advocate an initial mission that serves as a proof-of-concept demonstration. It would provide not only valuable scientific data but also serve as a basis for assessing the merits of deploying a constellation of operational satellites for long-term monitoring and for calibration and verification of other remote sensing systems. The initial microsat with a GPS receiver payload could be built and launched within three years of project start using current technology. The total proof-of-concept mission could be completed within four years for less than $10M, including mission operations, ground data processing, and archival operations. Costs are kept low because of

(1) the low power and mass requirements of the flight receiver,

(2) the high degree of heritage and synergy provided by related GPS-based geodynamics programs and radio occultation programs in planetary science,

(3) the existing infrastructures for ground operations,

(4) the enormous progress in GPS-specific very large-scale integration (VLSI) circuitry and the consequent simplification of GPS flight receiver architectures,

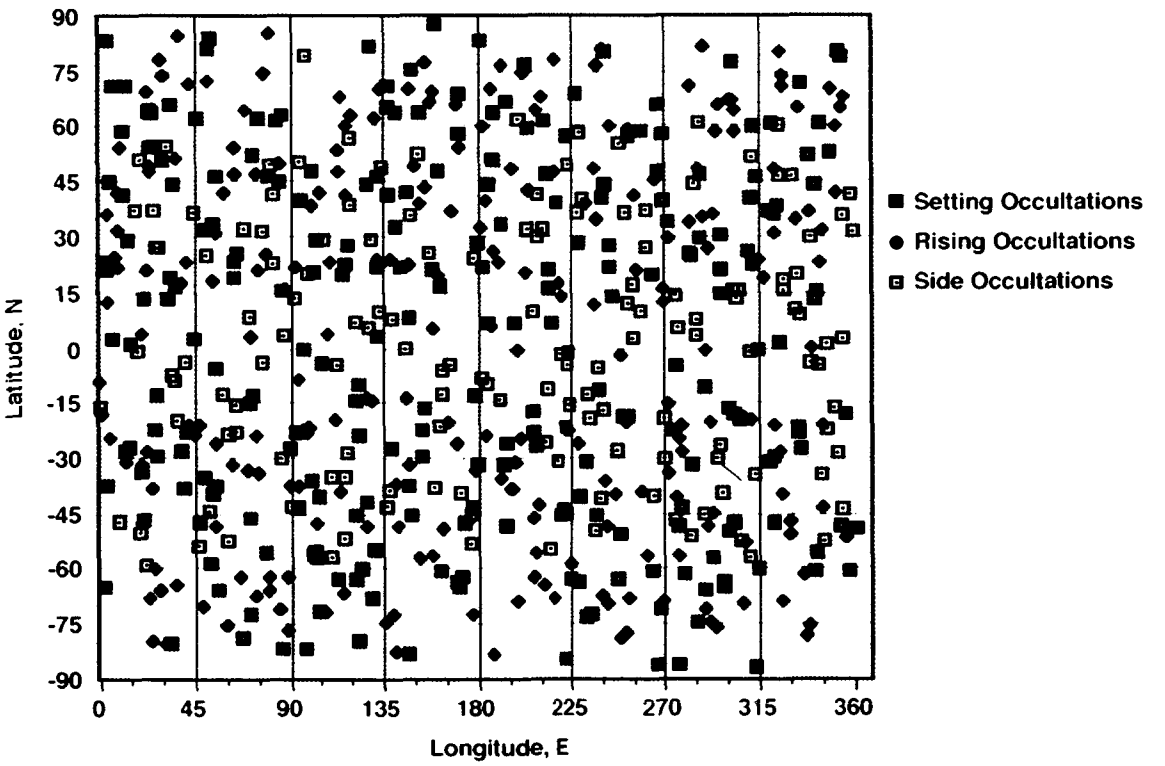


Figure 2-5. Global distribution of about 700 occultations in a 24-hour period for one LEO with a full GPS constellation. Orbital altitude of the LEO is 800 km at 97 deg inclination. About 500 occultations rise or set within ±45 deg of the local along-track direction.

(5) the maturity and low cost of commercial microsat technology, and

(6) the proliferation of low-cost commercial launch systems.

Ultimately, low-cost operational payloads (with marginal costs of less than $1M) could be carried piggyback on a variety of proposed LEO commercial satellite fleets. Thus, operational data might become available from private satellite operators. Alternatively, a fleet of 20–30 dedicated, second-generation GPS microsats could provide 5,000–10,000 soundings per day. Either way, missions such as GPS/MET or GLIMPSE will benefit from substantial economic and technical leverage available by exploiting existing GPS signals.

The GPS radio occultation concept requires a globally distributed ground network of about a dozen GPS receivers that can concurrently track the dual-band signals from the GPS constellation (Figure 2-6). These concurrent data streams and those from the LEO are centrally processed to maintain highly accurate ephemerides of the GPS satellites and the LEO, which are essential to maintaining high-accuracy reference radial velocities. The concurrent observations from the ground also help eliminate errors arising from short-term oscillator instabilities by processing the flight and ground data streams synchronously. During each occultation event, at least one ground station must concurrently observe two GPS satellites in common view by the LEO: (1) the occulted satellite and (2) the "clock" satellite. Differencing the phase measurements of these two GPS satellites eliminates errors arising from oscillator instabilities in the flight receiver. Using these

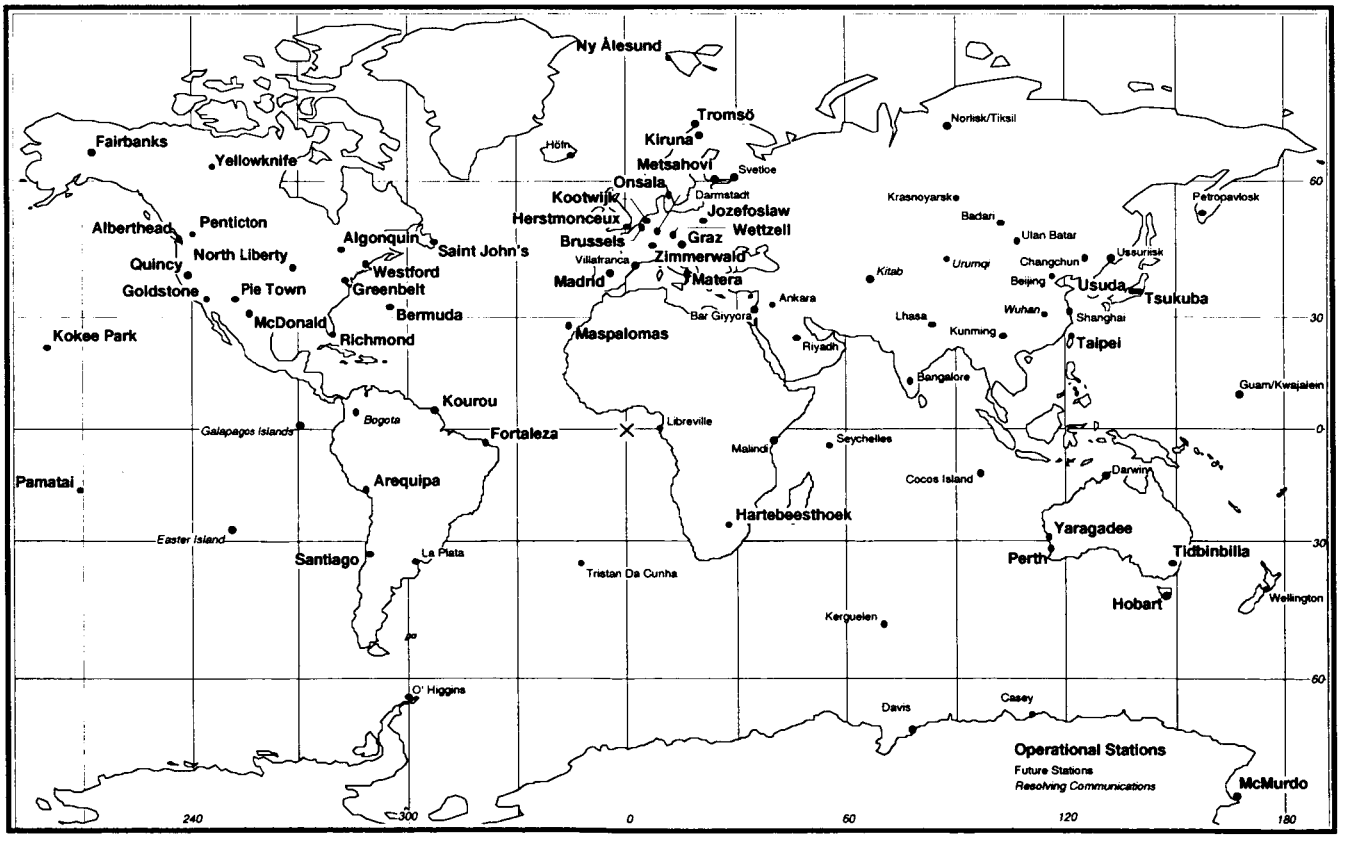


Figure 2-6. GPS global network for geodynamics studies and space operations. The IGS international GPS tracking network in 1994.

differenced phase measurements with similarly differenced phase measurements taken from the ground station eliminates errors arising from oscillator instabilities in the GPS satellites. The ground stations must sample at a rate consistent with the short-term oscillator stability of the GPS satellite clocks (cesium) to ensure accurate interpolation to achieve simultaneity in the observations. A 1-Hz or less sample rate is required. The nominal sample rate by the LEO would be 50 Hz; this would ensure adequate Nyquist sampling. Thus, both the LEO and certain stations in the ground network must be provided with almanacs and schedules for observations of the occulted and clock satellites at these higher rates instead of at nominal rates, which typically are 0.1–0.01 Hz. To obtain nearly complete global coverage (85% for GPS/MET), it is sufficient to have the high-rate sampling capability at six ground stations, which are uniformly distributed ~±30 deg about the equator. Thus, the six-station subnetwork supporting TOPEX/POSEIDON would suffice, as would several variants. Future plans call for expanding the high-rate network to 12 stations.

Figure 2-7 shows a diagram of the key flight/ground system elements of the GPS-based tracking system. The GPS tracking system consists of the GPS constellation, the flight GPS receiver on board the LEO, a global network of GPS ground receivers, and a central monitor, control, and processing facility. The precision orbit determination (POD) strategy requires continuous tracking of the visible GPS satellites by the ground and flight receivers. Data from all receivers are brought together and processed in a grand solution in which the LEO and all GPS orbits, receiver and transmitter clock offsets, ionospheric corrections, and other parameters are estimated. Simultaneous sampling at all receivers (which may be achieved by later interpolation) eliminates the effect of common errors, such as clock errors.

Heritage From Planetary Radio Science

Radio occultation is a well-established technique in planetary science; it has been part of the experiment mix in NASA's planetary program since the early 1960's. The occultation technique was introduced in astronomy in the eighteenth century when it was used for timing celestial events. More recently, it has been used to study the physical properties of planetary atmospheres by observing scintillations and variations in stellar brightness and spectra as a star is occulted by a planet (see, for example, Chandrasekhar 1952; Salpeter 1967; Hubbard et al. 1978). The occultation technique has also been used in conjunction with lunar occultations to study brightness distributions of extragalactic radio sources (Hazard 1962).

In the early 1960's, Stanford University developed a one-way phase coherent dual-band radio frequency (RF) system using the early Pioneer spacecraft. The RF signals were transmitted from a ground-based radio telescope to the spacecraft; the phase difference recovered from the dual-band carriers was returned via a telemetry link to the Earth. The change in these differential phase measurements was used to study spatial and temporal distributions of the solar wind, exploiting the dispersive refractive index of a plasma. For NASA JPL's Mariner 3 and 4 missions to Mars in 1965, the Stanford group at first proposed (1962) to use the group's dual-band uplink approach. NASA's support was requested to develop and deliver to the Mariner project the necessary on-board hardware and processing systems. The objective of the proposed experiment was primarily to probe the Martian ionosphere and the planetary surface (Fjeldbo 1964). However, probing the neutral atmosphere was not considered fruitful with this approach because the signature of the then-putative, very thin Mar-

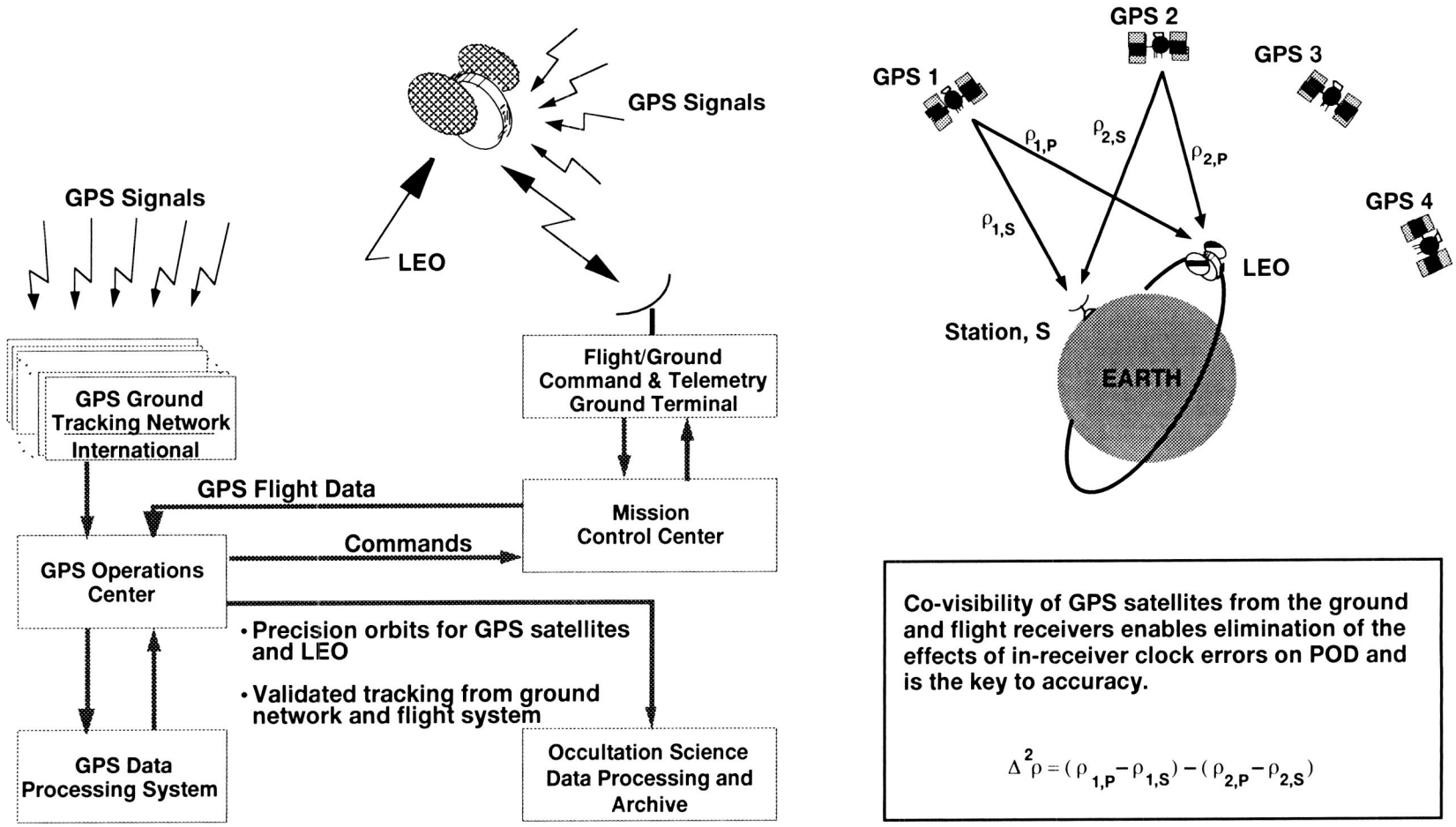


Figure 2-7. Organizational diagram of the GPS tracking system for LEO.

tian atmosphere (surface pressure in the 10–25 mbar range) on the one-way phase measurements of the Stanford system would have been masked by on-board reference oscillator frequency drifts. Moreover, defocusing effects from vertical gradients in atmospheric refractivity would have caused only about a 20% drop in signal amplitude for a ray just grazing the Martian limb, yielding relatively crude information about the neutral atmosphere. For these and other reasons pertaining to mission design and project schedule and budgetary constraints, the Stanford proposal was rejected by NASA.

Meanwhile, JPL, which had developed a two-way coherent counted Doppler system in the late 1950's for spacecraft navigation to the Moon and Venus, recognized in 1963 the potential of the two-way differential phase measurements for recovery of pressure and temperature profiles from the atmosphere of a distant planet using the occultation technique. The two-way system eliminates the problem of on-board reference oscillator drifts. This scheme involves a ground-based transmission from a radio telescope to the spacecraft, which coherently returns the RF signal at a slightly offset frequency to the transmitting radio telescope on the Earth. There it is combined with the transmitting signal to obtain a precise differential carrier phase measurement of two-way range change (see, for example, Melbourne 1976). Errors in the carrier phase measurements caused by frequency drifts in the reference oscillator of the transmitter depend only on those drifts that are incurred over the round-trip light time of the signal, or about 10 min for a Mars close encounter. The short-term stability of the rubidium frequency standards used in the ground stations at that time was state of the art. The total atmospheric delay on the RF signal for grazing incidence at Mars is about 2 m, but the instrumental accuracy of the counted Doppler measurement at that time was better than 2 cm.

A JPL group therefore proposed, independent of and without knowledge of the earlier rejected Stanford proposal, to probe the Martian neutral atmosphere using the two-way Doppler system on Mariners 3 and 4 (Kliore et al. 1964). NASA opted for JPL's proposal partly because the two-way Doppler system was already included in the mission for navigation and because it could accurately probe the neutral atmosphere. Accurate knowledge of the Martian neutral atmosphere was important in the design of a future Mars lander. The major accommodation required by the JPL proposal was changing the planned Martian approach trajectory of the Mariner spacecraft to an occulting one.

However, in 1964 NASA formed an experiment team including both JPL and Stanford scientists, with Stanford focusing on the Martian ionosphere (which could also be studied with the JPL system) and with JPL focusing on the neutral atmosphere. From this somewhat contentious beginning, a remarkable collaboration between the scientific and engineering groups at Stanford and JPL has ensued, making these groups world leaders in exploiting radio occultation for the study of planetary atmospheres (see, e.g., Kliore et al. 1965; Fjeldbo and Eshelman 1968; Fjeldbo et al. 1971; Eshelman 1973; Lindal et al. 1983; Marouf et al. 1982; Newman et al. 1984; Marouf et al. 1986; Tyler 1987; Lindal 1992). Radio occultation has been used to probe the atmospheres and ionospheres of the planets and their moons, as well as certain physical properties of planetary surfaces and planetary rings.

CHAPTER

3 SCIENTIFIC APPLICATIONS OF GPS RADIO OCCULTATION MISSIONS

The Atmosphere

GPS radio occultation provides an all-weather limb sounding that recovers vertical temperature profiles, particularly in the upper troposphere and lower stratosphere regions, with accuracies exceeding those of other remote-sensing techniques. The vertical resolution of GPS radio occultation will range from ~1 km in the stratosphere down to 100 m near the surface where the refractivity gradients are large; these resolutions are comparable to or better than those of the best limb sounders. Moreover, under conditions where the horizontal gradients are small, the vertical resolution can be pushed beyond the Fresnel diffraction limit by sampling, subsequently deconvolving the diffraction pattern.

Because aerosol, clouds, and precipitation particles present in the troposphere cause signal extinction at visible and infrared wavelengths, routine limb sounding at these altitudes must be done at relatively long wavelengths. Radio occultation wavelengths are very long (~20 cm) by atmospheric remote-sensing standards. The GPS frequencies are well below any absorption lines, and the effect of particle scattering will be generally negligible. This allows routine observations to be made down to the Earth's surface in a limb-sounding geometry without the fundamental difficulties that scattering and absorption cause at shorter wavelengths.

Radio occultation provides a self-calibrating system; it uses precise phase measurements of artificial signals in a differential mode in which oscillator-induced phase errors cancel and where POD establishes a very accurate (~±10 ppb) reference phase profile that would have been obtained in the absence of the intervening atmosphere. In contrast, the observations by radiometry of the absorption and emission spectral profiles from natural sources usually are obtained from amplitude measurements that are more vulnerable to long-term stability errors in their calibration systems. Radio occultations can provide an absolute standard and/or calibration system for other temperature-monitoring systems, thereby satisfying the requirement for long-term continuity and stability in global temperature monitoring. With respect to the monitoring and tracking of long-term climate trends, the more precise and consistent the measurement, the shorter the time period required to identify the trend. The radio occultation technique is well suited to the task. Furthermore, the temperature data obtained from these experiments can be applied to climatological studies to monitor trends, particularly in the upper troposphere and lower stratosphere where the occultation technique yields its most accurate results.

The refractivity itself should be very accurate throughout most of the troposphere and strato-

sphere. This accuracy combined with the simple calculation required to estimate it from model data may ultimately cause refractivity to become a principal weather and climate observable. Globally distributed daily profiles of atmospheric refractivity, which from a single LEO (at 500 occultations per day) would have a mean spacing of about 1000 km, would be an additional constraint on climatological and weather models and improve their accuracy. A constellation of LEO satellites (see Chapter 6), each observing all visible satellites in the GPS constellation, would not only increase the density of the profiles but could enable tomographic analysis to be performed, providing improved spatial resolution of variability in atmospheric refractivity down to the mesoscale level.

The Lower Troposphere and Water Vapor

Although temperature profiles from satellites have been available for more than two decades, their impact on forecast models has been rather small, particularly for the Northern Hemisphere. Radiosonde data are generally available only over land areas and for a few island stations, leaving large information gaps over oceanic regions. Better vertical resolution of temperature and moisture observations is also needed for more accurate forecast models.

Weather analysis and forecast models require as input four-dimensional (three spatial and time) measurements of temperature, moisture, pressure, and wind. In addition, information on surface and cloud features is needed for many weather and climate models. Meteorological satellites provide global coverage for some key weather parameters. However, current capabilities of operational weather satellites fall short in meeting the requirements for accurate and timely measurements of most weather parameters needed to study global change.

Radio occultation measurements of temperature are potentially more accurate and have finer vertical resolution than those from nadir-viewing satellite temperature sounders. Hence, these new data will be valuable as input to numerical weather models. Currently these models routinely handle information from a variety of satellite and ground-based sensors. Thus, data from GPS occultation measurements could be incorporated into the existing models with little modification.

Because water is so fundamentally important to the Earth's weather and climate system and because presently there is a lack of accurate water vapor profiles on a global scale, meteorologists and climate modelers are interested in how radio occultations might be used to determine water vapor distributions. Knowledge of water vapor density in the lower troposphere is needed to determine (1) the latent heat suspended in the atmosphere, (2) the radiative forcing due to vapor and clouds (inferred from humidity), and (3) improved inputs for weather forecasting, climate, and hydrology. Retrievals of water vapor in tropical latitudes are of special interest because this is the region where most of the incoming solar energy is deposited and where water plays such a primary role in distributing this energy to higher latitudes and, ultimately, radiating the energy back to space. Tropical latitudes are largely covered by ocean, so remote sensing is almost a necessity in order to provide the information needed to better understand evaporation and precipitation.

The large permanent dipole moment of the water molecule and the short scale height of water vapor in the troposphere combine to create a disproportionately large (up to ~30% near the surface in the tropics) contribution to the total measured phase delay from GPS occultations. To isolate water vapor profiles from the observed

refractivity profiles, ancillary information must be provided, e.g., independent measurements or models of temperature. In the tropical lower troposphere, occultation techniques have the potential to recover the vertical profile of water vapor densities to accuracies better than 10% below inversion layers (Kursinski et al. 1993a; see also Chapters 8 and 9) and total columnar water vapor content to better than 5%. Because this is a limb-viewing geometry, it will probably work best over oceans where, in general, the horizontal gradients are least. The very sharp and large contrast in refractivity at the marine boundary layer between the moist air below and relatively dry air above induces dramatic changes in the amplitude and phase of the signal as the ray path crosses this boundary. In studying weather and climate, the height of the marine boundary layer is of special interest but is very difficult to recover accurately with nadir-viewing sounders. However, knowledge of the height of the marine boundary layer obtained from occultation data will improve the interpretation of nadir-viewing measurements of water vapor. Provided the atmosphere is horizontally uniform over the sampling dimension of about 160 km, the sharp discontinuity in refractivity at an inversion layer will generate a diffraction pattern that should provide sufficient information to recover the height of the inversion with ~100-m resolution.

The high vertical resolution obtainable in the troposphere creates opportunities for tropospheric weather and climate monitoring. A key issue is how to deal with and recover horizontal structure. There is a tendency by some to view the occultation geometry and conclude that the horizontal resolution is relatively coarse and, therefore, of limited value. However, the vertical-to-horizontal resolution ratio (~1:160) is certainly appropriate for synoptic scale weather characterization. It has also been pointed out by Lindzen and Fox-Rabinovitz (1989) that horizontal and vertical resolutions are strongly coupled and, therefore, must be increased in a self-consistent manner, both in atmospheric models and in observational systems. Their results, in fact, indicate that vertical resolution is inadequate in virtually all large-scale weather models and observing systems; consequently, improved vertical resolution is sorely needed.

Temperature retrievals at high-latitude winter conditions provide an example of how the radio occultation observations can complement data from passive sounders. Accurate temperature retrievals by passive infrared (IR) remote-sensing techniques under these conditions are difficult because of the very cold temperatures, the common near-surface thermal inversions, and the high percentage of ice clouds that limit IR soundings to the cloud tops. In contrast, the radio occultation technique proves advantageous in colder temperatures because the air is dry, relatively dense, and scale heights are relatively small—all of which result in a more accurate temperature retrieval. Preliminary simulations of temperature retrievals under these conditions indicate that accuracies better than 1 K can be obtained throughout the troposphere down to the surface (see Figure 2-3). A dense sampling of the high latitude regions would be useful for routinely providing a detailed and very precise vertical reconstruction of the thermal structure of the polar vortex regions. Combining these precise vertical temperature profiles with other observations of ozone densities and dynamic models should improve our understanding of the conditions that lead to formation and dissipation of polar stratospheric clouds and how the particles on which the heterogeneous chemical reactions leading to ozone loss are believed to occur. From a climate change

perspective, the change in surface temperatures caused by an increase in greenhouse gas densities is generally predicted to be largest and, therefore, most apparent first at high latitudes.

The Tropopause

The tropopause is the transition region between the dynamically controlled thermal structure in the troposphere below and the radiative control in the middle atmosphere above (see Figure 2-1). This region is of scientific interest although it is a difficult region to characterize via remote sensing. The GPS occultations are expected to provide high-accuracy temperature and pressure profiles ranging from 7–35 km, thus providing specifications of the tropopause height and temperature.

Exchange between tropospheric and stratospheric air masses involves the tropopause structure. The minimum temperatures that characterize the tropopause act as a cold trap, wringing water out of the tropospheric air; thus any air reaching the stratosphere from below contains very little water vapor. The vertical temperature structure in the vicinity of the tropopause controls static stability and, therefore, determines where and how deeply air will penetrate into the stratosphere from below. This interaction tends to be difficult to monitor and depends largely on the local temperature, water vapor, and atmospheric stability. Therefore, very accurate, globally distributed, high-vertical-resolution temperature profiles provided by GPS occultation measurements should contribute to our understanding of how the stratosphere and troposphere interact.

Climatologically, the temperature in the upper troposphere/tropopause region is of interest because of its influence on the amount of energy radiated to space. Changes in concentrations of atmospheric radiative-forcing constituents such as aerosols, cloud particles, and greenhouse gases force the atmosphere to alter its temperature structure to maintain an energy balance with incoming solar radiation. Major volcanic eruptions inject a large amount of aerosol material into the lower stratosphere, significantly altering the radiative forcing of the atmosphere. Detailed monitoring of the evolution of the atmospheric thermal structure following these events can only be done on a global scale with remote sensing and done remotely with wavelengths sufficiently long to be unaffected by the enhanced aerosol concentrations.

Accurate high vertical-resolution temperature reconstructions in the upper troposphere will increase our understanding of the conditions under which cirrus clouds form. For example, cirrus clouds will generate a positive feedback effect if global warming displaces a given cloud layer to a higher and colder region, then the colder cloud will emit less radiation, forcing the troposphere to warm to compensate for this decrease (Houghton et al. 1990). The changes in the altitude region associated with anthropogenic increases in greenhouse gases are poorly understood, as indicated by the variability of Global Circulation Model (GCM) predictions. Accurate temperature retrievals from occultation measurements combined with high horizontal-resolution temperatures derived from nadir-viewing microwave radiometers (Spencer and Christy 1990) should provide a powerful data set for climate studies of the Earth's lower atmosphere.

The atmospheric sounders on operational weather satellites all use passive infrared and/or microwave nadir-viewing radiometers. An inherent attribute of the retrieval of temperature or moisture profiles from these radiometers is the weighting function or contribution to the observed

brightness temperature by various thicknesses within the atmosphere. The observed brightness temperature at each frequency is a result of the radiance from a weighted layer extending over about one scale height or 6–8 km within the troposphere and stratosphere. Although increasing the number of channels helps to improve the resolution, radiance emanating over a scale height limits the vertical structure that can be retrieved.

The comparatively sharp change in the lapse rate and the cold temperature make it difficult to retrieve from radiometric measurement the structure near the tropopause. The results of simulations and comparisons between radiosondes and retrievals from satellite radiometric data consistently show increased errors in the retrieved temperature near the tropopause. Comparing radiosonde data with retrievals from High-Resolution Infrared Sounder 2 (HIRS/2) plus the Microwave Sounding Unit (MSU), Reale et al. (1988) found that the root-mean-square (RMS) errors of the retrieved soundings were about 2.5 K near the surface and the tropopause and about 1.5 K within the mid-troposphere. For particularly sharp changes for a tropical sounding at the tropopause, the coldest temperature retrieved can be more than 10 K warmer than the observed temperature (Yates et al. 1989).

Temperature measurements by the radiosondes are themselves subject to systematic errors at tropopause heights. Because the air is thin at these altitudes, the physical temperature of the radiosonde thermistor is actually determined by a balance between conductive heat transfer with the air temperature, absorbed long-wave and solar radiation, and the long-wave radiation emission of the sensor. This makes the measured temperature sensitive to a variety of factors such as cloud cover, cloud top temperatures, solar zenith angle, surface temperature, vertical temperature structure, and the vertical distribution of aerosols, ozone, water vapor, and carbon dioxide (Finger and Schmidlin 1991; McMillin et al. 1988). Present estimates indicate that the maximum magnitude of these errors is on the order of 1 to 3 K (Ahnert 1991; Schmidlin 1991; McMillin et al. 1988).

From simulations using GPS occultation data, the temperature and the height of the tropopause can be retrieved to within 1 K and less than 1 km, respectively (Kursinski et al. 1993a; Kursinski et al. 1993b).

The radio occultation technique is somewhat unique in that it measures properties of the bulk atmosphere. To this point, we have tended to regard temperature and water vapor as the primary parameters of interest. However, very accurate reconstructions of pressure as a function of geometric height will also be provided wherever accurate temperature retrievals are produced. Pressure versus geometrical height is potentially an interesting diagnostic of tropospheric climate change because the height of any pressure surface is a function of the integrated temperatures below (Gary 1992). Therefore, if the troposphere warms in response to the anthropogenically increased greenhouse gas, as is generally predicted, this will force the average pressure scale height across the troposphere to increase and cause the height of a given pressure level to rise. The tropopause altitude regime should yield the largest effects because the height of pressure levels here represents the integrated effect of the warmer temperatures below. At higher altitudes in the stratosphere, predicted cooling will reduce pressure scale heights, somewhat negating the effect of the expanding troposphere below. A simple order-of-magnitude estimate indicates that a 2-K increase (a magnitude often discussed in climate change simulations) will produce about a 70-m increase

in height. This is equivalent to a change in pressure at fixed altitude of approximately 1%. Since the accuracy of each sounding will potentially be ~0.1%, detection of such a change appears promising. In fact, the sensitivity implies that changes in temperature of less than 2 K may be detected, and, thus, changes in temperature or pressure fields will be identified much earlier. The problem to be addressed then may shift to the separation of the climate change signal from natural climatic variability.

The Stratosphere

Once GPS occultation measurements become available for atmospheric research, the precise temperature data could be applied to map the structure of the stratosphere, particularly in the polar regions where temperature is believed to be an important contributor to the minimum levels of ozone observed in the spring. Recent observations of the decrease of ozone in the lower stratosphere, combined with the ozone's known radiative properties, have generated much interest in how high vertical-resolution measurements of temperature can assess climate change and model accuracy (Gille 1991).

A large-scale volcanic eruption can inject massive amounts of aerosols into the lower stratosphere and, consequently, can significantly limit IR observations of the stratosphere and lower regions. As these aerosols change the radiative forcing of the atmosphere, this is precisely the time when accurate observations are needed to determine how the atmospheric thermal structure is adjusting to achieve overall radiative energy balance. The GPS occultation technique will still be fully effective in the presence of aerosols and, thus, will provide temperature structure when other satellite techniques are subject to error.

Climate models predict that the presently increasing abundances of atmospheric CO_2 and other greenhouse gases will give rise to an increase in radiation within the stratosphere and will, therefore, produce significant cooling. The GPS observations provide a very accurate, long-term data set for monitoring both regional and global scales. The observed temperatures will be compared against model predictions to provide both a method for detecting and characterizing stratospheric climatic variations as well as a means for evaluating the performance of model behavior at stratospheric altitudes.

With sufficiently dense sampling, the high vertical resolution and very accurate density, pressure, and temperature profiles will improve the reconstruction of wind fields in the stratosphere and knowledge of stratospheric circulation. This is interesting in its own right, but these profiles will also improve our knowledge of advective terms in stratospheric chemistry and particularly in the ozone.

The Ionosphere

Radio occultation measurements could enable global mapping of the ionosphere with sufficient temporal and spatial resolution to investigate many important dynamic processes in the ionosphere/thermosphere system and their relation to processes in the atmosphere and solid Earth. We can investigate traveling ionospheric disturbances (TIDs) caused by internal gravity waves and their relation to the flow of energy and momentum through and among the Earth's layered systems. Also, mesoscale ionospheric structures such as the mid-latitude trough and equatorial bubbles may be reconstructed via tomographic techniques.

Measuring Total Electron Content

With a simple linear combination of the carrier phase measurements from the dual-frequency GPS data, the total electron content (TEC) along all GPS-to-receiver ray paths can be continuously measured. Pseudorange by itself will give absolute TEC with an accuracy of about 10^{16} e/m^2 (a 16-cm delay accuracy at L1) averaged over 1 s. Smoothing pseudorange against precise carrier phase will improve this to better than 10^{15} e/m^2 (<2 cm), or about 0.1% of the daytime zenith peak and 0.5% of the typical nighttime minimum. Continuous carrier phase measurements by themselves will give TEC change with a precision of better than 10^{14} e/m^2 (<2 mm). These estimates are based on the demonstrated performance of the Jet Propulsion Laboratory's TurboRogue receiver. Down-looking TEC measurements with the PRARE system, and possibly with other platform instruments such as DORIS and other multifrequency radio beacons, will strengthen ionospheric sampling. The global coverage offered by these space platforms and the GPS ground sites will provide an unprecedented examination of ionospheric structure and short-term variability.

*Traveling Ionospheric Disturbances**

Traveling ionosphere disturbances (TIDs) are a manifestation of internal gravity waves, also called acoustic gravity waves (AGWs), propagating in and intersecting with the ionosphere. AGWs are present at all levels of the atmosphere and have been detected by a variety of experimental techniques (Hines 1960; Yeh and Liu 1974). The sources of AGWs are many and diverse: volcanic eruptions, earthquakes, and nuclear explosions in the atmosphere. A magnitude 7 earthquake, for example, will create a surface uplift of several centimeters over several hundred kilometers, launching a gravity wave. Upon reaching the ionosphere, the AGW can temporarily increase the TEC by 5 x 10^{15} e/m^2 or 8 cm of phase advance which is easily observable on the L1 carrier. In most cases, however, observed AGWs have their origin in meteorological events such as strong wind shear, distortion in jet streams, convective storms and fronts, and auroral-related high-latitude events.

Once generated, AGWs carry energy and momentum from the source to other parts of the atmosphere. As they ascend through the rapidly thinning medium, the particle motions are amplified through conservation of energy flux and become an increasingly important component of the local atmospheric energetics. The wave-associated velocity increases with height until nonlinear effects such as resonant interactions or wave breaking take place. When this happens, energy and momentum carried by the wave may be deposited to the background atmosphere. For example, it is currently believed that AGWs play a key role in determining the general large-scale circulation in the middle atmosphere (Geller 1983; Fritts et al. 1984). In addition, a principal means by which the lower atmosphere couples with the thermosphere is through the dissipation of AGWs, resulting in momentum and heat flux divergences and enhanced turbulent mixing. At high latitudes, energy input from the magnetosphere generates AGWs that propagate around the globe in the thermosphere (Hunsucker 1982) providing a possible channel for coupling of energy from high to low latitudes. Hines (1965) estimated that in the ionospheric E-region, heating rates from TIDs are comparable to the heating from solar radiation and have levels ranging from 10 K/day near 95 km

* C. Liu, then at the University of Illinois, contributed some of the text in this and the following sections on the ionosphere.

to 100 K/day near 140 km. At F-region heights (~300 km) there is evidence that gravity waves may seed plasma instabilities that cause "spread-F" irregularities (Kelley et al. 1981).

Ionospheric Tomography

Computerized tomography (CT) is a technique whereby an image of an object is constructed from a set of projections, or integrated densities, taken along many lines through the object. CT is widely used in medicine to produce two-dimensional cross-sectional X-ray images. More recently, three-dimensional acoustic and seismic tomography have been applied to the oceans and the solid Earth. In ionospheric tomography, first described by a group at the University of Illinois (Austen et al. 1986; Austen et al. 1988) and more recently by other groups (see, for example, Chiu et al. 1991; Hajj et al. 1994b), the measurement of TEC along a ray path is the analog of the integrated density registered by an X-ray. TEC measurements taken along many ray paths can be used to construct maps of electron distribution in two or three dimensions. There are two basic methods for inverting the measurement set to construct the tomographic image: (1) transform methods and (2) finite series expansion. Only the latter is suitable with the nonuniform sampling available from practical ionospheric measurement systems. However, with only ground-based observations available, very limited work in ionospheric tomography has been possible to date, and this limitation will continue until suitable ionospheric sensors are placed in orbit.

Irregular structures of different sizes are generated in the ionosphere as a result of its electromagnetic, corpuscular, and dynamic coupling with the magnetosphere as well as its dynamic and aeronomic interactions with the thermosphere and middle atmosphere. For example, the mid-latitude ionospheric trough lies close to the plasmapause; hence, the dynamic behavior of the trough is directly related to magnetospheric processes such as substorms. The equatorial bubble, to take another example, is believed to be caused by Rayleigh–Taylor instabilities on the underside of the F-region, which develop nonlinearly into strong depletions of the ionization. The associated polarization of the electric field can drive these bubbles upward to form plumes. In recent years, high-power, RF waves have been used to modify the ionosphere through nonlinear interactions, giving rise to irregularities of different sizes in the F-region. These heat-generated irregularities are believed to result from a variety of plasma instabilities. The ionosphere, in short, is a complex, highly mutable matrix containing an assortment of transient, inhomogeneous structures. To begin to understand it, we need an accurate physical description of its different features and we must be able to follow the evolution of those features in time and space. Over the years, many observational techniques have been devised for this purpose. One of the most important of these is the measurement of TEC.

If a number of LEOs carrying all-in-view GPS flight receivers were deployed (see Chapter 6) and were supplemented by a moderately dense global network of ground receivers such as the 200+ fiducial stations soon to be operating under the International GPS Service (IGS) aegis, an immensely rich and continuous global TEC data set would become available for ionospheric tomography. Sun-synchronous orbits of the polar platforms will give a detailed picture of the variation of ionospheric structure not only from the daytime peak through the nighttime minimum, but with latitude, with season, and across the solar cycle.

CHAPTER

4 OVERALL MISSION DESCRIPTION

An Example Design for an Operational System

Two mission concepts are of interest: one that will demonstrate the occultation technique and another that will exploit this technique in the long term. First, we will consider how this technique will be used over the long term.

The Global Positioning System

The 24-satellite GPS constellation is maintained and operated by the U.S. Air Force. The full suite of satellites was completed in early 1994.

The LEO Constellation

For an operational system, we estimate that about 10,000 soundings per day would provide more than adequate spatial coverage to support weather forecasting and climate models. This requires a constellation of about 20 LEO satellites. Any orbit altitude in the five hundred to several thousand kilometer range would yield an adequate number of limb soundings and have adequate orbital lifetime. Sun-synchronous orbits offer the advantage of simpler spacecraft operations, but such orbits only provide occultations at fixed local times, a disadvantage for adequate temporal sampling. The orbital inclination of the Block II GPS satellites is 55 deg; this results in coverage "holes" about the celestial poles where no GPS satellites can be observed. Consequently, the LEO in polar orbit observes no fore or aft occultations as its groundtrack passes through the equatorial zones, resulting in uneven geographical distribution of occultations and significant gaps in the equatorial zone. An orbital inclination for the LEO of about 60–65 deg appears to provide the best compromise between achieving good global coverage and obtaining equatorial coverage. Such a constellation could be established in five to eight orbital planes with three to four satellites per plane. This configuration would cover all occultation points on the equator and at higher latitudes, even when occultations more than 45 deg from along-track are discarded. The constellation could be built one plane at a time, with three to four satellites carried by a single launch vehicle.

The Payload

The GPS flight receiver, ultrastable oscillator, and GPS antennas constitute the payload. Fore and aft antennas enable both rising and setting GPS satellites to be tracked and provide sufficient position dilution of precision (PDOP) for high accuracy POD. Some positive gain (~10 dB) in the fore and aft antennas would help the signal-to-noise ratio (SNR) limitations encountered in the lower troposphere and upper stratosphere, but their reduced beamwidth might require an addi-

tional vertical antenna for POD. The receiver must be a high-precision, dual-band unit capable of tracking low signal levels with relatively high dynamic properties. A sample rate of up to 50 samples per second may be required to capture the signal dynamic changes expected in the lower troposphere. To accommodate AS, the receiver will have to operate either in the crypto mode to track the Y-code or in the enhanced codeless mode. (See Chapter 5.) Over-the-air rekeying of the receiver will be accomplished through a decryption chip embedded in the receiver itself. The receiver will not need to correct for selective availability (SA). Not converting for SA keeps the downlink data unclassified and makes the flight data compatible with the ground data, which is also not corrected for SA. The total mass of the payload should be under 5 kg, and the required power should be under 10 W. The pointing requirements for the antennas are modest: ±5-deg pointing control. Knowledge of the yaw rate that is equivalent to a rotation about the vertical axis by the antenna phase of better than 0.1 mm/s is required to avoid corrupting the basic accuracy of the carrier-phase measurements.

The Satellite

A dedicated microsatellite in the 50- to 80-kg class is required. Options for a dedicated satellite include (a) the ORBCOMM bus being developed by Orbital Science Corp., (b) the kind of microsats being flown piggyback by ARIANE for a fee of ~$200,000, and (c) the QuickStar bus being offered by Ball Aerospace. Another practical way to collect occultation data might be to replace a single-frequency GPS receiver (being carried for navigation and time synchronization by a satellite in the Light-Sat class) with a precision dual-band receiver. An example of this configuration is the satellite currently being developed for Motorola's IRIDIUM communication system.

The Launch System

A dedicated launch on a Pegasus would allow optimization of the system design and operational management. The Pegasus XL has the capacity to deliver approximately 200 kg to a 900-km polar orbit and approximately 300 kg to a 400-km polar orbit. This capacity can support deployment and maintenance of the constellation.

To be feasible for the ultimate system, a "piggyback" option must be associated with launches to support a constellation of satellites. For example, the Delta is likely to be used to maintain the GPS constellation (55-deg inclination orbits). It may be possible to deploy as many as four QuickStar-class satellites riding piggyback on the Delta's second stage on its way to the GPS orbit.

Mission Operations, Ground Control, and Data Receiving System

Even without compression, the data rates for the flight system are low enough (less than 65 Mbytes/satellite/day for the 50 sample-per-second maximum sample-rate case) to make a direct satellite-to-ground communication system practical. All of the data collected in a day could be easily received at a single ground station. Operating at 1 Mb/s, all data could be transmitted in less than 10 min (i.e., in less than one typical tracking pass). Because of the higher frequency of contacts, a high-latitude station is desired for monitoring the health of the satellite itself.

The Ground-Based GPS Global Tracking Network

NASA Code O plans to have a globally distributed 12-station network of GPS receivers operating 24 hr per day to support frame-of-reference

calibration functions of the Deep Space Network (DSN) and to precisely determine the orbits of altimetric missions such as TOPEX/POSEIDON. Code O has already supported the implementation of a six-station network for the GPS flight experiment on board TOPEX/POSEIDON. Radiometric data and station status information from both the flight system and the ground stations flow continuously to the GPS Operations Center at JPL, where monitor and control functions for the network are performed and where tracking data are generated. Because a GPS tracking network for space operations requires a high degree of reliability, most but not all of the system elements within the network must be dedicated to space operations support. For example, NASA and its international partners in geodynamics research are also implementing a 40–60 station network (the core network of NASA's FLINN program) that came on line in 1992 with over 25 stations; a full complement of 50+ stations will be implemented by the end of 1994. It would be a relatively straightforward procedure to convert several of the sites in the FLINN network, including ground communications, central processing, and monitor and control, to complete the network of approximately 12 space-operations ground-receiver stations required for a radio occultation system. Thus, a certain degree of synergy between these two programs should be evident, not only in shared data acquisition systems, but also in other system elements such as ground communications, monitor and control, hardware and software configuration management, and so on. JPL will operate both networks for NASA.

NASA-generated ephemerides for the GPS satellites have been available for general use from NASA's FLINN network since the summer of 1992. This network operates under somewhat less stringent reliability conditions than the space network; nevertheless, it has configuration control and operations standards that are rigorously followed to ensure accuracy, stability, integrity, and timeliness in distributing GPS data products. To eliminate POD as a source of velocity error for recovery of upper stratospheric temperatures, the POD accuracy for the LEO must be better than a few parts in 10^8 to ensure that the error in the Doppler measurement is at or below 0.1 mm/s. Most of the power in the orbit error spectrum is concentrated around a frequency range once per revolution. The POD position accuracy on TOPEX/POSEIDON from the GPS flight experiment appears to be better than 10 ppb (Melbourne et al. 1994; Schutz et al. 1994; Yunck et al. 1994). Moreover, recent global network site position results from NASA's geodynamics ground programs appear to surpass 5 ppb (Blewitt et al. 1993).

Central Data Processing and Archive System

The GPS operation center at JPL retains data for a short period of time, usually 30 days. The system supports automatic delivery of certified data products to archive and processing centers that make data accessible to end users.

Concept Demonstration

For demonstrating the radio occultation concept, we conceived of several approaches to mission design: (1) as an adjunct to a larger mission that already has a GPS receiver for another purpose (e.g., GGI, GAMES/GRACE), (2) as a microsat attached to a Delta- or ARIANE-launched satellite, or (3) as a microsat on a dedicated or shared Pegasus launch.

Although the dedicated launch on a small satellite may be the most expensive way to achieve a proof-of-concept demonstration, it has the advantage of being more closely related to the

deployment of a complete system for long-term monitoring of weather and climate. This is the approach that we have chosen to examine for a proof-of-concept GPS/MET or Ørsted class of mission.

Mission Design

The microsat carrying the GPS occultation receiver would be launched into a circular and near-polar orbit by a Pegasus XL, for example. A mission that collected a few days' worth of data would be a success, but it would be preferable to collect data over several seasons. Therefore, a six-month mission would be desirable.

The exact parameters of the orbit are not critical for a successful proof-of-concept mission. However, a Sun-synchronous orbit, at an altitude of about 900 km, an inclination of ~99 deg, and with its nodal line in the terminator plane, offers attractive operational features, such as simplifying the satellite solar power system. It also provides a longer duration for the mission in full Sun, which could supply a test receiver with higher power than would be available from an operational system. A circular orbit with an altitude of 900 km can provide a one-day ground track repeat pattern after 14 orbits. Such an orbit also yields a half-day, quasi-repeat pattern of occultation locations, which disperses slowly over the ensuing days because no on-orbit station-keeping capability is assumed for the microsat. This could be important in planning experiments that compare results of in-situ measurements.

The Pegasus XL can deliver approximately 200 kg to a 900-km polar orbit. As many as four microsatellites could be launched at once. Only one of them would need to be dedicated to the GPS proof of concept.

GPS Flight Receiver

In the spring of 1994, a demonstration unit was delivered to the GPS/MET project at UCAR; the mission is planned for launch in October 1994. The flight hardware and software were delivered in approximately 12 months by upgrading a geodetic-quality commercial receiver (the TurboRogue). Some hardware changes were needed to meet the functional requirements of the mission, but performance optimization was deferred to later units. Changes to the TurboRogue included

(1) foaming the components in place to add resistance to launch vibration,

(2) a new enclosure with a mechanical interface,

(3) designing new software to support automatic operation on-orbit,

(4) adding a telemetry and command interface with the satellite,

(5) adding a power supply interface with the satellite, and

(6) modifying the receiver-to-antenna interface.

This unit has a mass of less than 3 kg and requires less than 18 W of power. For the proof of concept, GPS's antispoofing function will be turned off periodically to support the mission. Three or four one-week periods with AS off over the course of a six-month mission should be an adequate sample of the seasonal effects on the measurement.

Proof-of-Concept Operations

Mission operations, ground control, the data receiving system, the GPS global tracking network, and the central data-processing and archive system would be similar to those in an operational system. To a large extent, existing institutional facilities would be used for the proof of concept.

User Segment for Proof of Concept

Although the central data-processing facility of the GPS global network could and probably would generate temperatures, pressures, and water-vapor distributions from the data records in the data information system (DIS), we believe that a proof-of-concept demonstration would be enhanced by the participation of scientists from a number of disciplines, including meteorology, atmospheric science, and radio science. Several federal and university laboratories involved with global change problems should participate through an announcement of opportunity (AO) process; these groups would access the radio occultation DIS data products and analyze them for further assessment of the system performance in terms of accuracy and evaluation of error sources, resolution, and limitations. These groups would also be able to compare and combine radio occultation results with data gathered from other ground and flight programs studying global change and using extensive modeling and analysis programs. In general, these groups should evaluate the radio occultation technique for its potential role in long-term global change research.

CHAPTER

5 TECHNOLOGY READINESS ISSUES

Microsatellite Launch Vehicles

Microsatellite technology has matured considerably in recent years. Semi-standard satellite buses with masses in the 50-kg class and below are now available from several commercial suppliers. Similarly, commercial space transportation companies have been focusing more on the small satellite market.

GPS Flight Receiver

While the required flight receiver is not yet completely developed, its component technology is well understood and readily available. The NASA–JPL-sponsored Monarch receiver, launched on TOPEX/POSEIDON in August 1992, is the first civilian flight receiver for high-accuracy positioning applications. However, possible adverse signal conditions in the lower troposphere have led to performance specifications that drive GPS receiver design toward lower SNR operations. This includes the use of the P-code (or Y-code) for carrier recovery, improved sky acquisition techniques, and the recording and reporting of carrier phase measurements at Nyquist sampling rates for the probable bandwidth of the signal spectrum in adverse signal conditions. Several design changes in the Monarch would be needed to meet these requirements. For example, the Monarch uses a signal squaring-detection scheme during its GPS signal acquisition phase. Although this scheme works well when the GPS signal is strong, with rising GPS satellites early acquisition will be difficult to achieve because of probable adverse signal conditions experienced in the lower troposphere. In addition, the Monarch also decimates the carrier phase measurements; it reports 1 sample per second of a 17-Hz B_L tracking loop, corresponding to about 30 ms of data. This strategy was forced upon the Monarch design by throughput limitations in the 1750A microprocessor. Unlike many scientific applications that are not SNR limited, radio occultation is SNR limited in the upper stratosphere and lower troposphere.

The JPL-developed TurboRogue (Meehan et al. 1992), which has recently been commercialized by Allen Osborne Associates, Inc. (AOA) for ground operations, is currently being modified for a low-Earth orbit experiment (GPS/MET). It appears to best meet the performance requirements for radio occultation if it can be economically flight rated. The TurboRogue hardware design is well suited for low-Earth orbit science applications. It features a compact, low-power (18 W) design that uses a pair of high-density gate array chips, an applications-specific integrated circuit (ASIC), and the LSI Logic 100,000 series (which is driven by a powerful 32-bit microprocessor (AMD 29050)

with enough excess throughput to adapt to the rigorous tracking environment of atmospheric occultations). Heavy emphasis has been placed on using digital processing techniques to produce a data set where error sources are precisely understood if not eliminated. The basic design can track up to eight GPS satellites simultaneously and extract all readily available navigation signals.

Testing of the ASIC set and the AMD 29050 microprocessor for total dose radiation and susceptibility to single-event upsets from high-energy particles was completed in 1992. Use of alternative space-rated 32-b microprocessors and commercially available flight-rated gate array architectures is also being evaluated. Initial data on the GPS-specific VLSI circuitry (designed to Mil. Spec.) indicate sufficient radiation hardness (~30 krad) for medium duration in low-Earth orbit applications. Coupled with a static memory architecture and radiation-tolerant programmable logic, this will provide a good margin of safety against the hazards of the space environment. Although changes to the specific hardware layout are anticipated in order to further improve heat transfer and reduce vibration effects, no major redesign of the TurboRogue should be necessary. Thus, it is reasonable to project a relatively short, low-risk development period, consistent with the strawman schedule herein.

Flight Receiver Performance Under AS Operations

An operational GPS occultation system will need receivers capable of operating with AS on. Antispoofing gives U.S. military users anti-jamming capability. In the future, one can expect to see AS on for the duration of specific military actions (i.e., for at least several months). With AS on, the P-code on both the L1 and L2 carriers is coherently encrypted by superimposing a classified code known as the W-code, which has a single-sided bandwidth of approximately 500 kHz. The resultant P⊗W code is known as the Y-code. Although recovery of the L1 carrier phase by the receiver can still be effected through use of the unencrypted clear access (C/A) code, recovery of the L2 carrier phase is compromised through denial of the P-code unless the receiver operates with knowledge of the W-code and, therefore, is in a classified mode.

With AS on, the TurboRogue has a codeless operating mode that cross-correlates the Y-code on L1 with the Y-code on L2 to produce a measure of the ionospheric delay on the signal from which the measured L1 carrier phase can be corrected and the L2 carrier phase can be inferred. The thermal error in the phase measurement derived from cross-correlating goes as 1/SNR for low SNR values and approaches $1/\sqrt{\text{SNR}}$ for large SNR values. When the P-code is used, the thermal error goes as $1/\sqrt{\text{SNR}}$. Figure 5-1 shows the thermal noise on the L2 carrier phase recovery for both P-codeless and P-code techniques. While the cross-correlation technique (the "TR XCORR" curve shown in Figure 5-1) has adequate SNR for ground applications, it does not have adequate SNR for the limb-sounding applications in the upper stratosphere (where the signature from the atmosphere is weak) and in the lower troposphere (where scintillations in phase and signal strength are likely and where multipath effects are encountered). Moreover, rapidly changing ionospheric delays encountered by the LEO exacerbate the L2 tracking problem. To support a near-term technology demonstration, a temporary waiver to DoD's AS on policy is needed. Theoretically, such waivers can be obtained between cooperating federal agencies and the DoD by using established protocols. In practice, these procedures tend to be time-consuming and are not 100% reliable. The DoD

has consistently urged partner agencies requiring P-code performance to implement a Y-code capability in their systems.

For an operational system supporting weather and climate modeling, two options are possible: (1) the enhanced codeless option (the "code enhanced" curve in Figure 5-1), which exploits the relatively narrow RF spectrum of the W-code to greatly reduce thermal noise on the L1–L2 carrier phase measurement derived from the correlation operation, and (2) the Y-code option. Enhanced codeless techniques that improve the SNR to support orbital operation have been developed by Magnavox, Co., Ashtech, Inc., and JPL. The ASIC needed to convert any P-code tracking GPS receiver to Y-code operation is available from a

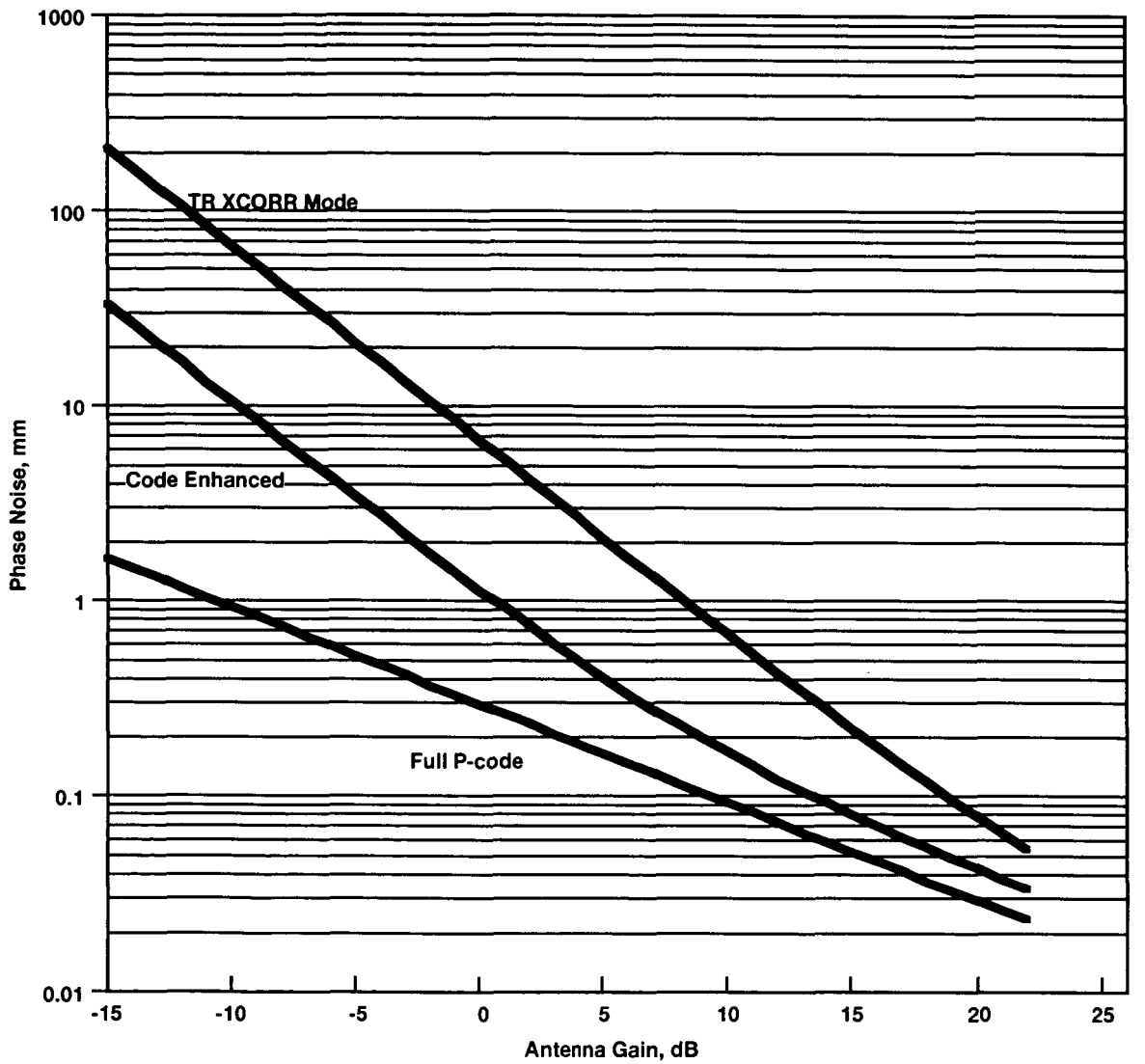


Figure 5-1. TurboRogue L2 phase noise versus antenna gain for 1-s averaging and a system temperature of approximately 250 K at 6-dB antenna gain. The "TR XCORR" curve recovers L1–L2 carrier phase by cross-correlating the L1 and L2 signals and subtracting the recovered phase from the C/A-based L1 carrier phase; thus, the noise penalty for this mode involves the (signal + noise power)^2 across the entire P-code RF spectral range. The "code enhanced" curve involves similar correlation operations, but utilizes the narrower W-code RF spectrum. The "full P-code" curve recovers the L2 carrier phase directly by correlating the received signal with either the receiver-generated P-code or the Y-code; it involves the noise power across the P-code RF spectrum linearly.

number of sources. For example, Motorola, Inc. is currently implementing its Y-code ASIC in a radiation-hard process developed by UTMC. In addition to the increase in complexity and power consumption, the main disadvantage of the Y-code option is the prelaunch security cost, particularly when a non-U.S. partner is involved and the launch is from a foreign site. After launch, AS-related flight and ground command and communications would be conducted in the open.

Figure 5-1 shows that over the 0–10 dB range, which corresponds to the signal strengths likely to be encountered by the LEO from the middle troposphere upward, the Y-code offers only a factor of 2 to 3 advantage in phase noise over the enhanced codeless technique. However, for the weak (–10 dB and less) and highly dynamic signal conditions found in the lower troposphere, one incurs a significant penalty using the codeless techniques. Future versions of the Turbo-Rogue flight receiver for radio occultation applications will include both the enhanced codeless and Y-code options.

Flight/Ground Communications

The requirements on the communications system are modest. Assuming that occultations data are collected at 50 samples per second, each satellite produces less than 65 Mbytes/day. This can be downlinked in ~10 min at 1 Mbps. A 1-kbps uplink would make it possible to perform 25-kbyte modifications to flight software in a single tracking pass from a low-latitude station.

Satellites can operate in a store-and-forward mode. Only one or two ground terminals would be needed to support the whole constellation. One high-latitude site, which would have six to eight contact periods with each satellite per day, is desirable for monitoring satellite operation and troubleshooting. Low-latitude sites typically have contact with the satellite two or three periods per day.

The ground control and data receiving system will consist of readily available products. The U.S. Air Force has developed ground equipment to support their Space–Ground Link System (SGLS) that has all the necessary elements needed to synthesize the ground terminals and to support the proposed constellation. If needed, SGLS offers an option for over-the-air rekeying of GPS flight equipment. Furthermore, the National Oceanic and Atmospheric Administration (NOAA) and the small satellite industry both have well-developed systems for command, control, and data acquisition from satellites.

CHAPTER

6 COST ESTIMATES AND SCHEDULE

Cost of Implementing the System

In this chapter we discuss both the cost of a system that would demonstrate the occultation concept and the cost of the ultimate system. We chose for illustration a single-purpose satellite and dedicated launch vehicle, either the Pegasus XL or a competitive launch vehicle. The non-recurring cost of developing the first single-purpose satellite flight system with the ground system support for the constellation is estimated between $10M and $15M. The cost to deploy, operate, and maintain a constellation of 20 to 30 microsatellites is estimated to be between $40M and $50M per year; this includes the cost of producing refractivity profiles for the user community. The breakdown of the annual cost estimate is illustrated in Figure 6-1 (see Appendix A for further details on the baseline cost estimate).

Figure 6-1 provides a range of cost estimates for the main system elements. It serves as a baseline for comparing alternatives for reducing cost. There are a number of ways to reduce the

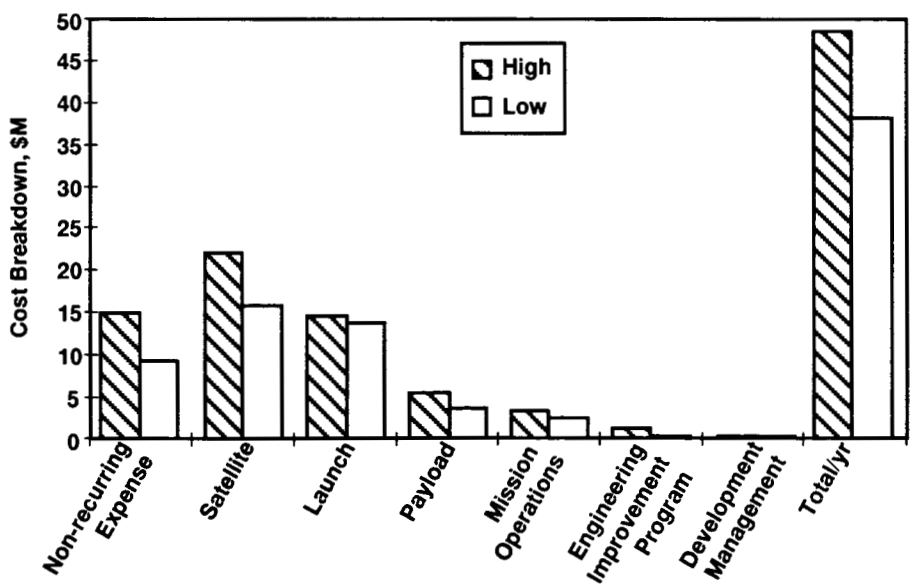


Figure 6-1. Baseline budget breakdown for a single-purpose satellite system using Pegasus XL launch vehicle (four satellites per launch).

cost of deploying a constellation of GPS limb-sounding instruments. More satellites per launch and cheaper launches clearly are advantageous. Some of the alternatives are discussed below.

Depending on the level of cost sharing for the first launch, a demonstration mission with the first article of the single-purpose flight system could be conducted for $15M to $25M; this includes the cost of development (NRE) plus the launch.

Launch Costs for Microsatellites

There are several possibilities for launching the constellation of microsatellites needed for the system. Several options are discussed below. Choosing the Pegasus XL as the dedicated launch option is the conservative approach and the most costly option.

A piggyback option seems viable for the ultimate system if the primary purpose of the launches also involves supplying satellites to a compatible constellation. Satellite constellations currently planned include the 24-satellite GPS constellation maintained by the U.S. Air Force, the 66-satellite IRIDIUM constellation proposed by Motorola, the 24-satellite ORBCOMM constellation proposed by Orbital Science Corporation (OSC), the 810-satellite project by Teledesic, and the 48-satellite constellation by GlobalStar.

Dedicated Launch With Pegasus XL

The price of a Pegasus XL launch has varied in recent years, but the reported costs range from $8M to $15M. We suspect that the price may depend on many factors. To stay competitive, the base price must be maintained at the $9M to $10M level, and development costs for unique mission interfaces must be added to this price. We estimate that development costs could be between $1.5M and $2M. Because the Pegasus XL is expensive on a cost-per-pound-to-orbit basis, using this launch vehicle may not be the best approach in the long run. However, the expense of a dedicated launch of this kind is balanced against the control and flexibility that it offers project management. With the Pegasus XL, management can optimize the system with a minimum number of constraints.

Piggyback Microsat on the ARIANE

At $0.2M per satellite, the ARIANE option is inexpensive, but the service it offers is limited. Locations for up to six 50-kg microsatellites are provided in the adaptor ring. But because of ARIANE's direct-ascent operation, the orbit must be similar to the orbit of the primary payload. In addition, space is offered on the basis of payload mass availablity. The primary payload customer must agree to ARIANESPACE launching the microsats. Moreover, ARIANESPACE must be assured that the microsats will not interfere with or present risk to the primary payload. We know of no plans for the ARIANE to maintain a constellation of satellites of any kind.

Piggyback DELTA II on the Ball Aerospace QuickStar

The U.S. Air Force and NASA are both encouraging piggyback launches. The piggyback option is offered for the cost of developing the interface and integrating the microsatellite with the Delta second stage. The cost of developing a new interface ranges from $2M to $4M. The cost of integrating the microsatellites with the launch vehicle ranges from $1M to $2M. Ball Aerospace suggests attaching four 170-kg satellites to the second stage of the Delta II. At least one mission has been conducted in this mode: LOSAT-X in 1989, a single satellite. More piggyback launches are planned for the Delta II. The Delta II is used to

launch the GPS Block II satellites and is a candidate for launching the GPS Block IIR satellites.

To make piggyback launches feasible, the problem of constellation compatibility must be resolved. Because of the 55-deg orbit inclination, the number and quality of high-latitude limb soundings are reduced. Furthermore, the deployment scenario adds risk to the GPS mission profile. The dual-constellation deployment scenario that maximizes payload capability requires that the Delta be initially deployed into an orbit with an apogee similar to the microsat orbit altitude (e.g., 900 km). At the apogee of the initial orbit, the microsatellites are deployed and at this point provisions must be made to circularize the orbits of the microsatellites. At the perigee of the initial orbit, the second stage of the Delta II fires again (into the transfer orbit to reach the GPS orbit altitude). It is not at all clear whether the risk of deploying three or four microsatellites before the GPS satellite would be acceptable. After all, *GPS is paying for the launch.*

Constellation Deployment Model

If the atmospheric refractivity is worth monitoring, it is worth monitoring over an extended period of time, perhaps decades. The strategy for implementing the system needs to account for this. Specifically, a system must be established to produce, launch, operate, and maintain a constellation of satellites for an indefinite period.

An efficient system must balance the satellite production rate (SPR) with the satellite consumption rate (SCR). This means that the satellite design life (SDL) should equal the time it takes to deploy the constellation plus one launch interval. The first satellites fail when the satellite production system has produced N+1 satellites, where N equals the number of satellites in the constellation. It follows that

$$SDL = N/SPR \tag{6.1}$$

In practice, it is not possible to accurately engineer the life of satellites; they either last longer than expected or fail prematurely. The strategy suggested here is to let the size of the constellation N be the independent variable ($N = SDL \times SPR$), and then empirically, to tune the satellite's life expectancy to achieve the desired size of the constellation.

To illustrate this idea, we have chosen a model budget for producing six satellites per year. Initially, we targeted for a satellite design life between one and two years. Assuming this, the constellation will stabilize at between 6 and 12 satellites. As time goes on, our deployment plan holds the production rate constant and expands the size of the constellation by increasing the satellite design life through a planned design improvement program. Because experience with satellite failures will influence the design improvement process, this strategy minimizes the cost of initial design of the satellite and allows the engineering effort to be concentrated on the most important design deficiencies. When the actual satellite life reaches four years, the constellation will move toward a stable size of 24 satellites. The model budget includes the costs for this design improvement process at 10% of the initial development cost each year.

This deployment strategy is consistent with achieving a 12-satellite constellation in 5 to 6 years and a 24-satellite constellation in 16 to 18 years. The growth scenario for the constellation is illustrated in Figure 6-2 for two different mean-time-between-failure (MTBF) scenarios.

A schedule for developing a 12-satellite constellation by the end of 2002 is presented in Figure 6-3. To achieve this objective, the development program must be started in 1995.

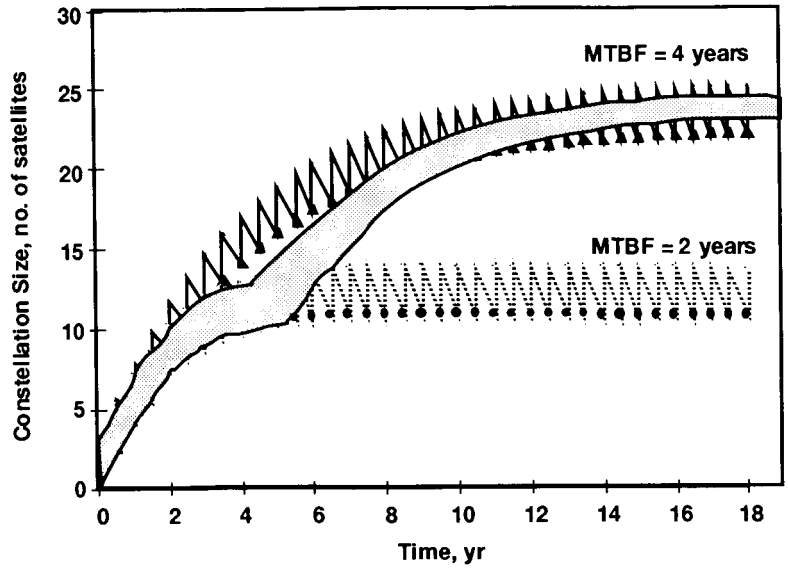


Figure 6-2. Management standard for building the constellation.

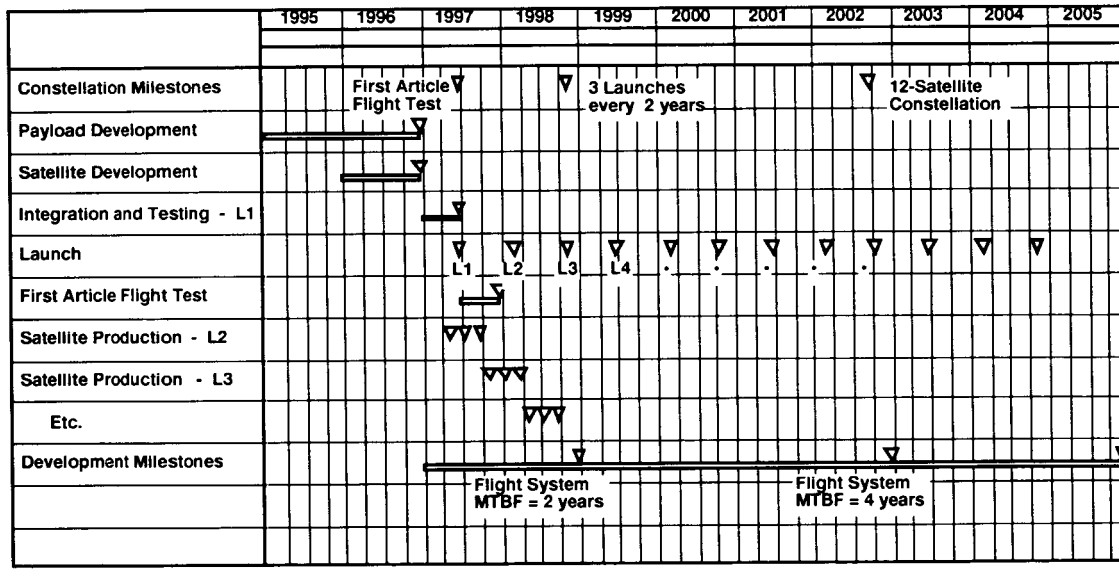


Figure 6-3. Top-level development schedule for 24-satellite constellation.

Satellite Cost Model

The satellite subsystems are essentially off-the-shelf systems developed by DoD and commercial programs to reduce the cost of operating in Earth orbit. By using off-the-shelf systems, we assume that the second article will cost 80% of the initial unit. This estimate is on the high side according to some manufacturers.

Payload Cost Model

The development cost for the flight TurboRogue with Y-code capability is based on class-C reliability standards. Our estimate includes the cost of the commercial P-code TurboRogue that will be integrated with an off-the-shelf W-code generator (e.g., Motorola's Sharpshooter™ chip) and Red-Prom technology for over-the-air rekeying of the

W-code generator. We assume a fair amount of attention will be devoted to the antenna engineering required to control the multipath from the satellite and to calibrate the antenna. An ultrastable oscillator (USO) similar to the Mars Observer USO will be purchased from The Johns Hopkins University Applied Physics Laboratory.

Based on an estimate of the amount of engineering that will be involved in developing the first article, we assume that the second article can be produced for 20% of the initial unit cost. For example, the second GPS flight receiver on TOPEX/POSEIDON cost only 15% of the cost of the first receiver.

Learning Curve Allowance

Experience in the production environment shows that cost goes down as cumulative output increases. A simple logarithmic model has been used to account for this effect on the cost of the satellite, payload, and launch vehicle: an empirical exponent for this effect is 0.9 (i.e., commonly called a 90% learning curve).

The Mission Operation System Cost Model

A dedicated satellite communication terminal should be able to easily handle the data flow between the ground and the satellites in the constellation. Allowance has been made for an upgrade of six GPS receivers in the GPS Global Network and an upgrade in the GPS operations center that will handle the ground data. To manage the operation, we estimate that a staff of approximately 25 people working 1.4 shifts per week will be sufficient. (The U.S. Air Force Consolidated Space Operations Command at the Falcon Air Force Base employs five shifts per week to control the GPS constellation. This is handled by a staff of 125.) Staffing of the GPS operation center is based on the assumption that the existing operation at JPL is supplemented.

We also assume that the size of the mission operation staff is invariant with the size of the constellation, from 1 to 30 satellites. We believe that this is a reasonable estimate given the time needed to train people, the likelihood of more frequent problems early in the deployment phase, and the expectation of shorter life for the flight equipment when the system has not yet matured. The assumed constant launch rate is consistent with this staffing level.

No learning curve allowance is made for the operations crew needed for the deployment phase. Because of the volume of data they must handle and the number of decisions they must make, the operators' workload will probably increase for several years into the deployment phase.

Mission Annual Cost and System Performance Trade-offs

Although still in the conceptual stage of development, the model budget does illuminate some important trade-offs that need more study. The annual budget is clearly dominated by the cost of producing and launching satellites. Launch cost can be reduced if more than three satellites can be carried on each launch. For the constant satellite production-rate scenario, the slower launch rate would presumably save money. However, an even distribution of satellites in each of the assumed eight planes will be more difficult to maintain. Further investigation needs to be conducted in how the number of planes impacts the utility of the GPS occultation data product. Fewer orbit planes offer more opportunities to reduce overall system cost.

CHAPTER

7 RADIO OCCULTATION METHODOLOGY

Introduction

As the LEO orbits the Earth, the received navigation signal from a distant and setting (rising) GPS satellite passes through successively deeper (higher) layers of the Earth's atmosphere. The GPS signal is both bent and retarded, causing a delay in arrival at the LEO (Figure 7-1). At the required sampling rate, the LEO records the dualband carrier phase measurements, the C/A and P-code group delay measurements, and the signal strength measurements made by the flight receiver; these are transmitted from time to time to the ground at designated telemetry sites. To eliminate errors arising from short-term oscillator instabilities in the satellites and receivers, at least one of the stations in the GPS ground network must concurrently track the same occulted satellite as well as at least one other satellite observed by the LEO. At a central network facility, the data streams from the flight receiver and from the ground network are processed in parallel at synchronous epochs to eliminate the effect of the oscillator errors. Highly accurate ephemerides of the LEO and GPS satellites referenced to the ground-based terrestrial frame (which is defined by the geocentric locations of the ground

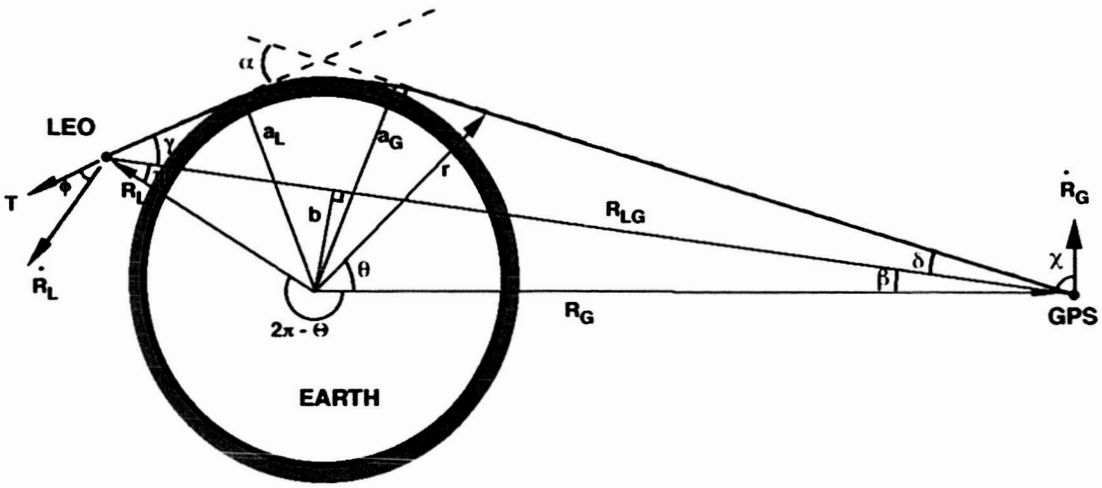


Figure 7-1. The ray path geometry for the LEO/GPS satellite combination is exaggerated. The maximum bending angle α, at grazing incidence with the Earth's surface, is about 1 deg for dry air. The maximum difference between b and r_o, the grazing distance of the actual ray, is about 60 km.

receivers) are produced along with the carrier phase and amplitude profiles arising from the occulted GPS satellites.

The radio occultation technique takes advantage of the extremely precise phase and amplitude measurements of the GPS navigation signals that pass through the Earth's atmosphere to provide accurate retrievals of the vertical refractivity profile (Figure 2-1). The nominal instrumental accuracy (i.e., excluding non-receiver error sources) of the L1 carrier phase measurement for the TurboRogue, which uses both the C/A and P-codes to obtain in-phase and quadrature carrier phase estimates every 20 ms, is around 0.1 mm at a 1-Hz sampling rate. Therefore, a detectable effect will be observed in the carrier phase residuals from a GPS signal passing through the mesopause as high as 90 km. The TurboRogue also measures the amplitude of the signal to an accuracy of about 0.3% at a 1-Hz sampling rate at nominal SNR conditions. The information in the amplitude measurement is complementary to that in the phase measurement. The amplitude data are used along with the phase data primarily for spectral analyses of signal structure as a result of multipathing in the lower troposphere and for sharpening the resolution using Fresnel transform techniques.

Radio Occultation Inversion Techniques

Although the open literature is replete with discussions of various aspects of radio occultation theory applied in probing planetary atmospheres in the Solar System (see also Lusigan et al. 1969, for an early discussion of the use of Earth satellites), there has been only limited discussion of the specific application with GPS (Melbourne et al. 1988; Yunck et al. 1988; Hardy et al. 1992; Bauer et al. 1992; Ware 1992; Hardy et al. 1993; Yuan et al. 1993; Kursinski et al. 1993a; Kursinski et al. 1993b; Gorbunov and Sokolovskiy 1993; Hajj et al. 1994a; Hajj et al. 1994b). Also, discussion of the basic methodology is somewhat scattered over a nearly 30-year period. Consequently, a description of the methodology is provided here for completeness.

Strictly speaking, the propagation of the GPS signal through the atmosphere obeys Maxwell's equations in which the propagation medium is characterized by a three-dimensional spatial (and possibly temporal) distribution of a complex and dispersive refractive index. Here, it is convenient to assume that the refractive index is real (i.e., zero absorption) and to assume that the signals are monochromatic; both of these assumptions are largely valid for GPS signals. Because the wavelengths of the GPS radio signals are small compared to the characteristic scale of the problem, a geometric optics approach will also be used. (Fresnel diffraction topics are covered in Chapter 11.)

For simplicity, we assume coplanarity for the ray path of the GPS signal, which is a good approximation for the Earth's atmosphere. A sufficient condition required for the ray path to be coplanar is one in which the component of the refractivity gradient normal to the plane is zero everywhere along the path. This plane osculates with time, depending on the out-of-plane components of the orbital velocities. The orbital radii of a GPS satellite and the LEO are about 4.1 and 1.1 Earth radii, respectively. The geocentric angular rate for the LEO is about 4 rad/hr and for the GPS satellite, about 0.5 rad/hr.

The resulting ray path from the GPS satellite to the LEO, in a geometric optics context, is *by definition* a path of stationary phase. Specifically, the phase delay incurred on this path is stationary with respect to the phase delay that would have resulted from following any other neighboring

and physically realizable path. This simply means that the ray satisfies Fermat's principle globally and Snell's law locally. Let $\Delta\rho$ be the additional delay as a result of the GPS signal traveling through the atmosphere. From the stationary phase concept (Appendix B), it follows that this ray must satisfy the requirement

$$\Delta\rho = \int n(\mathbf{r})\, ds - R_{LG} = \text{a stationary value} \quad (7.1)$$

where $n(\mathbf{r})$ is the real part of the refractive index, $\mathbf{r}$ is the geocentric position vector of any point on the ray, s is the arc length along the path, and R_{LG} is the geometric straight line distance between the LEO and the emitting GPS satellite, that is,

$$R_{LG} = |\mathbf{R}_L - \mathbf{R}_G| \quad (7.2)$$

where $\mathbf{R}_L$ and $\mathbf{R}_G$ are the geocentric position vectors to the LEO and GPS satellite, respectively (see Figure 7-1). In the following, we use a geocentric inertial frame to define the relevant quantities.

From the calculus of variations (or from Snell's law), it can be shown (see Appendix B) that a necessary local condition on the curvature of the ray, which must hold at all points along the path to achieve stationary phase (corners on the path involving discontinuities in $n(\mathbf{r})$ are discussed below), is given by

$$n\mathscr{R}^{-1} = |\mathbf{T} \times \nabla n| \quad (7.3a)$$

where $\mathscr{R}(\mathbf{r})$ is the radius of curvature of the ray and $\mathbf{T}(\mathbf{r})$ is its unit tangent vector, that is, $\mathbf{T} = \mathbf{k}/k$, where $\mathbf{k}(\mathbf{r})$ is the vector wave number of the ray at point $\mathbf{r}$ and ∇n is the gradient vector of the refraction index. The radius of curvature condition in eq. (7.3a) is an alternate form of the Euler equation, which is a necessary condition for the ray path if it is to be a path of stationary phase. The vector form for $\mathscr{R}(\mathbf{r})$ (see Appendix B) is given by

$$n\frac{\boldsymbol{\mathscr{R}}}{\mathscr{R}^2} = \mathbf{T} \times (\nabla n \times \mathbf{T}) = \nabla n - (\nabla n \cdot \mathbf{T})\mathbf{T} \quad (7.3b)$$

It follows that $\boldsymbol{\mathscr{R}}$ lies in the osculating plane defined by the normal vector parallel to $(\nabla n \times \mathbf{T})$, but $\boldsymbol{\mathscr{R}}$ is perpendicular to $\mathbf{T}$ and directed toward increasing n. Thus, the ray path is concave toward increasing n. At corner points on the ray path involving discontinuities in $n(\mathbf{r})$, the usual care must be taken in defining the right- and left-hand limiting values of ∇n.

The corner conditions for the boundary intercept problem from the calculus of variations (see Appendix B) also yields Snell's law, namely,

$$n(\mathbf{r})[\mathbf{T}(\mathbf{r}) \times \mathbf{N}(\mathbf{r})] \text{ is continuous along the ray} \quad (7.4)$$

where $\mathbf{N}$ is the unit vector normal to any boundary surface transected by the ray and across which $n(\mathbf{r})$ may be discontinuous. The condition in eq. (7.4) holds in three dimensions, providing the vector version of Snell's law, $n_1 \sin\varphi_1 = n_2 \sin\varphi_2$, where φ is the angle between $\mathbf{T}$ and $\mathbf{N}$.

The differential bending angle $d\boldsymbol{\alpha}$ accruing on the ray path over a differential arc length ds, and at any point possessing a continuous slope, is defined by $(\boldsymbol{\mathscr{R}} \times d\boldsymbol{\alpha}) = \mathbf{T}ds$; it follows from eq. (7.3b) that $d\boldsymbol{\alpha}$ is given by

$$d\boldsymbol{\alpha} = n^{-1}(\mathbf{T} \times \nabla n)\, ds \quad (7.5)$$

This can also be readily obtained by applying Snell's law eq. (7.4) over an infinitesimal arc length ds of the ray path if $n(\mathbf{r})$ is continuous. For a ray path without corners, that is, for a ray path that encounters no discontinuities in $n(\mathbf{r})$, the total bending angle $\boldsymbol{\alpha}$ is given by

$$\alpha = \int_{\mathscr{E}} n^{-1}(\mathbf{T} \times \nabla n)\, ds \quad (7.6a)$$

where $\mathscr{E}$ denotes a path integral along the ray. For ray paths with corners, eq. (7.6a) becomes

$$\alpha = \int_{\mathscr{L}} n^{-1}(\mathbf{T} \times \nabla n)\, ds + \sum \Delta\alpha_k \qquad (7.6b)$$

where $\Delta\alpha_k$ follows from applying Snell's law, that is, $(\mathbf{T}_{k-1} \times \mathbf{N}_k)\, n_{k-1} = (\mathbf{T}_k \times \mathbf{N}_k) n_k$, at the kth corner ($\sin(\Delta\alpha_k) = |\mathbf{T}_{k-1} \times \mathbf{T}_k|$).

Because the ray path as defined by eq. (7.1) is a path of stationary phase, the measured phase residual is sensitive only to second-order variations in the position and slope profile of the ray path. As a consequence, using the residual phase measurements directly to recover the atmospheric refraction profile leads to instability in the ray-tracing and refractivity recovery algorithms because of their insensitivity to ray path variations. On the other hand, the property of stationary phase also implies that the time derivative of the residual delay, that is, the Doppler residual, has no explicit first-order dependence on the velocity of the ray path itself as its shape and location evolve with time through the Earth's atmosphere (see Appendix B). Consequently, the Doppler residual depends only on *local* conditions at the LEO and the emitting GPS satellite, a much simpler relationship. Specifically, for both satellites the Doppler residual only depends upon the component of the satellite's velocity vector in the direction of the local ray path. Because $\dot{\mathbf{R}}_L$ and $\dot{\mathbf{R}}_G$ are known from POD information obtained before, during, and after the occultation event, the Doppler residual and the POD information yield ϕ, the angle between $\dot{\mathbf{R}}_L$ and $\mathbf{T}(\mathbf{R}_L)$ (Figure 7-1), and χ, the angle between $\dot{\mathbf{R}}_G$ and $\mathbf{T}(\mathbf{R}_G)$. The deflection angles of the ray at the LEO, γ, and at the GPS satellite, δ, readily follow. Thus, it is the Doppler residual, recorded by the LEO, not the phase residual, that is used in conjunction with the bending angles for ray tracing to recover $n(\mathbf{r})$. Moreover, it is eq. (7.6), not eq. (7.1), that provides the basis for the inversion algorithms. We present below two inversion approaches; the first is a general and discrete approach that assumes ∇n varies as a piecewise constant along the ray path; the second is a continuous approach that uses eq. (7.6a) but assumes spherical symmetry.

Differentiating eq. (7.1) with respect to time and invoking the endpoint transversality conditions from the calculus of variations (Appendix B), we obtain an expression for the Doppler residual that is given by

$$\Delta\dot{\rho} = \left[n\dot{\mathbf{R}} \cdot \mathbf{T}\right]\Big|_{\mathbf{R}_L} - \left[n\dot{\mathbf{R}} \cdot \mathbf{T}\right]\Big|_{\mathbf{R}_G} - \dot{R}_{LG} \qquad (7.7)$$

This expression, which depends only upon local coordinates and velocities of the satellites at the reception and transmit epochs, provides a constraining condition on the deflection angle γ of the ray at the LEO and the deflection angle δ at the emitting GPS satellite in terms of the observed Doppler residual and the satellite orbital velocity vectors provided by POD.

Ray-Tracing Techniques

The condition in eq. (7.7) and the supporting POD information are sufficient to begin a recursive ray-tracing program to reconstruct the ray path from the LEO back to the emitting GPS satellite for each observational epoch within the occultation event. Figure 7-2 shows a now-classic example of this approach using the so-called onion skin model. Here, the gradient of the index of refraction within each layer is assumed to be constant, but the layers themselves need not be spherical nor the resulting ray path coplanar; they can be defined to best match the local stratigraphy of the refracting medium. We can either require continuity in the refractivity across the boundaries or, if not, invoke Snell's law at the boundaries. Also, the layers need not be of equal thickness, nor in practice are they, because the

observational epochs are not necessarily spaced to yield equal thickness. Usually *local* spherical symmetry (i.e., corrected for flattening) has been used for the terrestrial planets and ellipsoidal symmetry for the Jovian planets (Lindal 1992).

A recursive approach enables us to reconstruct the path of the ray through the atmosphere for each Doppler observation epoch and, therefore, to obtain the refractivity at the point of closest approach by that ray and the position of the point of closest approach. From the ensemble of Doppler residual observations, we sequentially build a profile of the atmospheric refractivity. This can be represented schematically by a triangular matrix for the linear system that results from the differential correction process. Here, the coefficient X_{ij} in Figure 7-2, which depends on the particular stratigraphy adopted for the onion skin model, represents the partial derivative of the total bending angle, $\alpha = \gamma + \delta$, with respect to the gradient of the refractivity in the jth layer for the ray whose lowest atmospheric penetration is through the ith layer. The quantities $\delta\alpha_i$ and $\delta(\nabla n)_i$ in Figure 7-2 are the differential corrections to their respective nominal values. For example, X_{11} would account for the ray path curvature and, hence, the bending that results from the vertical refractivity gradient encountered by the ray as it traverses across the central cap of the top layer.

In effect (but not strictly in practice), the matrix in Figure 7-2 is sequentially inverted from

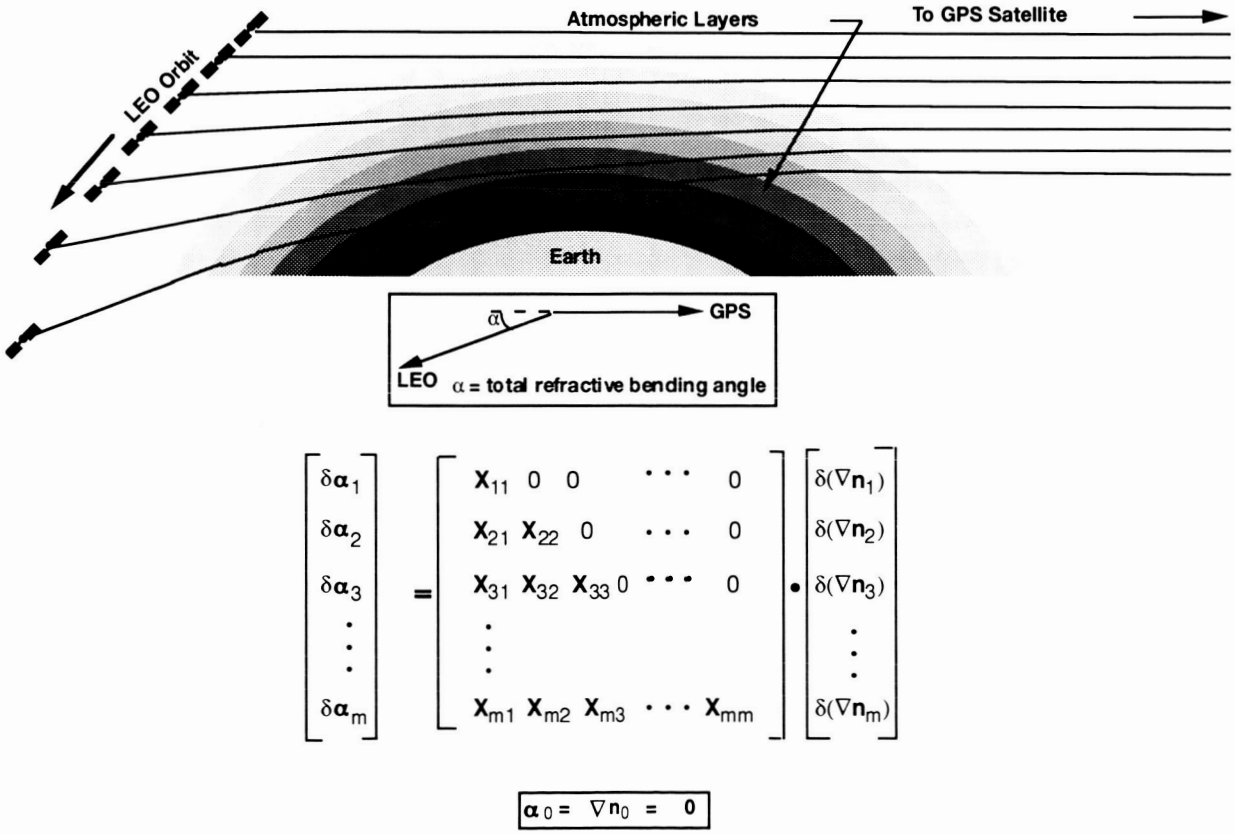


Figure 7-2. Onion skin model of atmospheric refraction. The gradient of the index of refraction is assumed constant within each layer. Continuity in the refractivity is required across the boundaries; otherwise, Snell's law is applied across each layer boundary. Starting at a 3-Hz sampling rate in the upper stratosphere and reducing to about 1 Hz in the lower troposphere, the total number of layers (m) would be about 100. The layer thickness would be about 1 km.

the top down using the previously recovered values of the refractivity gradients associated with the overlying layers. This sequential approach enables one to recover the entire vertical profile n(r) from top to bottom by integrating the recovered gradient series along the space curve normal to the boundary surfaces.

This recovery process can be readily described for the special case where the LEO and GPS satellites travel in circular orbits, which would be essentially true. For this case it follows from Figure 7-1 and eq. (7.7) that the deflection angles γ and δ are constrained by the Doppler condition

$$\Delta\dot{\rho} = \dot{\theta}_L R_L \sin(\psi+\gamma) - \dot{\theta}_G R_G \sin(\beta+\delta) - b\dot{\Theta} \quad (7.8a)$$

or

$$\Delta\dot{\rho} = V_{\perp L}\gamma - V_{\perp G}\delta + O[(\gamma,\delta)^2] \quad (7.8b)$$

where all quantities in eq. (7.8) except γ and δ are known from POD orbital position and velocity vector information for the LEO and the GPS emitting satellite.[1] Here, the angles ψ and β are defined in Figure 7-1. The quantities $\dot{\theta}_L$ and $\dot{\theta}_G$ are the components of the geocentric angular rates of the two satellites in the plane of propagation relative to an inertial reference; therefore, the angular rate of the central angle Θ is given by $\dot{\Theta} = \dot{\theta}_L - \dot{\theta}_G$. The quantity $V_\perp$ is the component of the orbital velocity perpendicular to the LEO–GPS line. The quantity b is the "impact parameter" or distance of closest approach of the undeflected ray to the geocenter; that is, the path that the radio signal would have followed in the absence of the Earth and its refracting atmosphere—it is also known from POD. It follows from Figure 7-1 that

$$b = R_L \sin\psi = R_G \sin\beta \quad (7.9)$$

Also from Figure 7-1, we obtain

$$a_L = R_L \sin(\psi+\gamma) \quad (7.10a)$$

and

$$a_G = R_G \sin(\beta+\delta) \quad (7.10b)$$

where a_L and a_G are, respectively, the impact parameters of the outgoing and incoming asymptotes of the deflected ray. Therefore, eqs. (7.8) and (7.10) provide information through the Doppler residual and the POD on the directions (γ and δ) and locations (a_L and a_G) of the asymptotes at each observational epoch. If a value of γ is assumed then values for the remaining quantities follow from these constraints.

Now consider the case where the refractivities of all of the layers overlying the ith layer have already been determined sequentially from the first $i-1$ Doppler residual observations. We describe the recursive ray-tracing process required to recover the refractivity for the ray that reaches a closest approach in the ith layer. Under the spherical symmetry assumption the problem will simplify considerably as discussed later. But for the general case where spherical symmetry does not necessarily apply, the convergence to the correct values of γ and n is accomplished recursively. At the beginning of the recursive process, the provisional value of γ at the ith observational epoch γ_i establishes the provisional direction of the outgoing asymptote of the ray path for that epoch (Figure 7-3). Also, the provisional value of

[1] The index of refraction has been set to unity in eq. (7.8). In fact, the nominal altitude of the LEO is within the boundaries of the ionosphere (see, for example, Figure 8-16). However, in solving for the refractivity in the neutral atmosphere, it is the ionosphere-free linear combination of the L1 and L2 carrier phase observations that must be used in eq. (7.8); this linear combination is known to include small higher order effects from the ionosphere, which should be included. (See Chapter 8.)

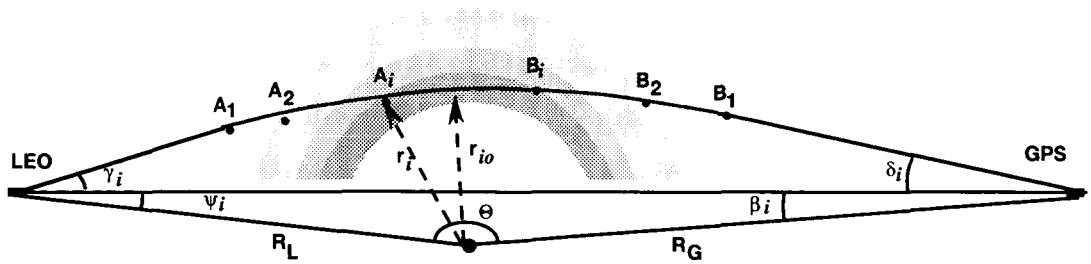


Figure 7-3. Ray tracing through a layered atmosphere. Ray path traversing the ith layer at closest approach and n_i are determined sequentially by the ith Doppler observation and the boundary conditions using previously determined values of n for the overlying layers.

a_L from eq. (7.10a) (or equivalently the boundary condition that the outgoing asymptote must intercept the LEO at the observational epoch) places the location of the asymptote. Tracing backward from the LEO along the asymptote, the ray path descends through progressively deeper layers of the overlying atmosphere. The ray's location and direction (at the points A_1, A_2, ..., A_{i-1}, in Figure 7-3) are controlled by the successive application of the ray path curvature within each layer, which is based on the values of the refractivity gradient associated with each overlying layer obtained from observations made at the $i-1$ earlier epochs within the occultation event (later epochs for emerging occultations). Near the mid-point, the ray path traverses the ith layer along a curved ray path[2] A_iB_i, which is the lowest layer for that observational epoch passing through a previously untraversed layer and, therefore, through a region of currently undetermined refractivity.[3] After reaching a minimum altitude at $\mathbf{r}_{io}$, the ray path then begins its return ascent through the overlying layers of previously determined refractivities. Finally, it exits the atmosphere along the incoming asymptote and follows it to interception with the LEO–GPS line defined by the vector $[\mathbf{R}_G(t_r) - \mathbf{R}_L(t)]$. There, it must satisfy the boundary conditions that require the wavefront normal to the ray path to intercept the emitting GPS satellite at the appropriate retarded time t_r and with the provisional deflection angle δ_i given by eq. (7.8). Specifically, these interception conditions require that the position and propagation vectors, $\mathbf{r}(t_r)$ and $\mathbf{T}(\mathbf{r}(t_r))$, of the wavefront on the incoming asymptote must satisfy

$$\begin{cases} \dfrac{\left[\mathbf{R}_G(t_r) - \mathbf{R}_L(t)\right] \bullet \mathbf{T}(\mathbf{r}(t_r))}{\left|\mathbf{R}_G(t_r) - \mathbf{R}_L(t)\right|} = \cos\delta_i \\ \mathbf{r}(t_r) = \mathbf{R}_G(t_r) \end{cases} \tag{7.11}$$

where t_r is the retarded time associated with the observation epoch t and corresponds to the epoch

---

[2] A curved path should be assumed within the layers to fully account for ray path bending. Hence, a local gradient must be included that is based on the recovered refractivity profile. If intralayer bending were not included by assuming that the refractive index varied in a piecewise constant fashion (remaining constant within each layer) and only Snell's law were applied to the straight-line ray path segments at the boundaries, then a prohibitive number of layers ($m > 10^5$) would be needed to achieve the required accuracy ($\leq 0.1\%$). (See Appendix C.)

[3] Here we have assumed that the spacing between observational epochs in fact places the lowest penetration of the ith ray in the ith layer. Clearly, the adopted vertical spacing between layers must be consistent with the sample rate and the local value of $\dot{a}$. An iteration on the sample rate and/or interpolation of the Doppler may be required to maintain commensurability with the spacing of the layers.

when the wavefront intercepts the emitting GPS satellite; i.e., $t_r = t - (|\mathbf{R}_L(t) - \mathbf{R}_G(t_r)| + Dr)/c$. It corrects for the LEO–GPS light time, about 100 ms. (Here, small relativistic effects are ignored.) In general these two (for the coplanar case) independent intercept conditions specified by eq. (7.11) will not be satisfied by arbitrary values of the two degrees of freedom available, namely, the values of γ_i and the gradient of n_i.[4]

In practice, this ray-tracing procedure must be done iteratively, adjusting the free parameters to obtain a match in the boundary conditions at the emitting GPS satellite. There are a number of methods for achieving convergence. One approach uses a dual backward/forward ray-tracing scheme as follows: Trace backward from the position of the LEO at epoch t and trace forward from the position of the GPS satellite at epoch t_r using the provisional values of γ_i and δ_i, which are mutually constrained by the Doppler observation through eq. (7.8). The segments of these two ray paths that are located in the $i - 1$ layer are provisionally determined because the refractivities and their gradients are known in the overlying layers. These provisional ray segments intersect the ith boundary at A_i and B_i, respectively, and, therefore, they provisionally define the positions on the boundary through which the ray path in the ith layer must pass. But the resulting provisional bending angle of the ray through the ith layer will not, in general, satisfy the continuity conditions in position and/or slope at A_i and B_i for any choice of the gradient of n_i. It then becomes necessary using some differential correction procedure to alter the provisional values of γ_i and δ_i, subject to the constraint imposed by eq. (7.8), and the corresponding value of the gradient of n_i to achieve convergence. When convergence is achieved, the sequential process continues to the $i + 1$ observational epoch.

For the special case of a spherically symmetric atmosphere, the recursive process described above is eliminated because both γ_i and δ_i follow directly from the ith residual Doppler observation. It will be shown later in this chapter that this symmetric case yields

$$a = a_L = a_G = R_L \sin(\psi + \gamma) = R_G \sin(\beta + \delta) \quad (7.12)$$

It follows from this and from eq. (7.8) and Figure 7-1 (with circular orbits) that

$$\Delta\dot{\rho} = \left\{ \begin{array}{c} R_L\dot{\Theta}\left[\sin(\psi+\gamma) - \sin\psi\right] \\ \text{or} \\ R_G\dot{\Theta}\left[\sin(\beta+\delta) - \sin\beta\right] \end{array} \right\} \rightarrow \gamma \text{ and } \delta \quad (7.13)$$

Thus, the location and direction of both the incoming and outgoing asymptotes are set from the Doppler residual observation and the POD information when spherical symmetry is assumed. Using the backward/forward ray-tracing scheme described above for the ith Doppler observation, we can solve directly for the positions of the points A_i and B_i on the circular boundary of radius r_i between the $i - 1$ and i layers in terms of the boundary conditions on the positions of the two satellites and the known values of γ_i and δ_i from eq. (7.13). Thus, a recursive technique is not required (except perhaps as an iteration to effect a small adjustment in t_r to account for the extra

[4] From Figure 7-3, the boundary between the i–1 and i layers is set by the choice of spatial symmetry and boundary scale (r_i) adopted for the onion skin model. The position of A_i on the ith boundary is determined by its intersection with the ray in the i–1 layer whose position and slope are set by the provisional value of γ_i and the known values of n and its gradient in the overlying layers. Choosing a value for the gradient of n_i fixes the path A_iB_i. This sets the position of B_i on the ith boundary and the exit angle of the ray into the i–1 layer at B_i, which in turn sets the remaining ray path back to the GPS satellite.

time delay $\Delta\rho/c$). From the curved path A_iB_i, we obtain r_{io}, the radius of closest approach of the ith ray; from the converged value of the gradient of n_i, we obtain the value of $n(r_{io})$ by integrating its gradient downward along the normal to the equipotential surface.

For the spherically symmetric case, a more efficient scheme using an Abel transform technique will be discussed below.

In polar coordinates eq. (7.8) becomes

$$\Delta\dot{\rho} = \left(\frac{a_L + a_G}{2} - b\right)\dot{\Theta} + \tfrac{1}{2}\left(\dot{\theta}_L + \dot{\theta}_G\right)\left(a_L - a_G\right) \tag{7.14a}$$

which for the spherically symmetric case becomes

$$\Delta\dot{\rho} = (a - b)\dot{\Theta} \tag{7.14b}$$

The bending angle α at the LEO for a thin atmosphere such as the Earth's is about 1 deg at grazing incidence for dry air and is given with good approximation by the ratio of the residual Doppler to the transverse velocity of the GPS–LEO line in the plane of propagation, that is,

$$\alpha = -\Delta\dot{\rho}/\dot{b} + O[(\Delta\dot{\rho}/\dot{b})^2] \tag{7.15}$$

This follows from eqs. (7.9), (7.12), and (7.14). To first order, γ and δ are given by

$$\begin{cases} \gamma = -\dfrac{\Delta\dot{\rho}}{\dot{b}}\dfrac{1}{1 + R_L\cos\psi / R_G\cos\beta} + O\left[\left(\Delta\dot{\rho}/\dot{b}\right)^2\right] \\ \delta = -\dfrac{\Delta\dot{\rho}}{\dot{b}}\dfrac{1}{1 + R_G\cos\beta / R_L\cos\gamma} + O\left[\left(\Delta\dot{\rho}/\dot{b}\right)^2\right] \end{cases} \tag{7.16}$$

Inasmuch as the maximum value (for dry air) of α for the Earth is 0.02, the error in the first-order expressions in eqs. (7.15) and (7.16) is less than 2%. The second term in the denominator of eq. (7.16) for γ is a parallax correction that vanishes for very distant emitting spacecraft. Although the carrier phase measurements from the TurboRogue form a statistically independent series, which is the fundamental observation set, the ionosphere-corrected Doppler observations formed from these phase measurements can be considered to be statistically independent (with some appropriate sampling interval) and to have a receiver accuracy (i.e., excluding mismodeled environmental error sources) of about 1 mm/s under nominal signal conditions. The magnitude of $\dot{b}$ is about 3 km/s for vertically descending (ascending) occultation profiles; that is, profiles for which the LEO, the geocenter, and the emitting GPS satellite remain in the same plane during the occultation event. Thus, the series of bending angle observations has an instrumental accuracy of about 300 nrad.

The Case of Spherical Symmetry: The Abel Transform

When local spherical symmetry is assumed, a remarkable simplification in the ray-tracing algorithms is available, which has its origins in seismology (Phinney and Anderson 1968; Fjeldbo et al. 1971). The simplification results in integral expressions for both the triangular matrix and its inverse that are associated with the linear system of ray paths and deflection angles described above and shown in Figure 7-2. For a spherically symmetric atmosphere, n is assumed to be a continuous function of radial position only. Upon applying the calculus of variations to eq. (7.1) in polar coordinates, we obtain for the residual phase delay[5]

$$\Delta\rho = \sqrt{R_L^2 - a^2} + \sqrt{R_G^2 - a^2} + a\alpha - 2\int_{r_o}^{\infty} \frac{n'}{n}\sqrt{\left[n^2r^2 - a^2\right]}\,dr - R_{LG} \tag{7.17}$$

[5] See footnote 1 regarding appropriate treatment of the ionosphere, which would modify eq. (7.17) slightly.

where n' = dn/dr. The angle α is the total bending angle between the incoming and outgoing ray asymptotes. The impact parameter a is given by

$$a = r_0 n(r_0) \tag{7.18}$$

where r_0 is the geocentric radial distance of the ray at closest approach.

When the assumption of radial symmetry applies, conservation of "angular momentum" holds and

$$n(r)|\mathbf{r}\times\mathbf{T}| = a = \text{constant} \tag{7.19}$$

along the ray path. This conservation result was used to obtain eq. (7.13); it is also known as Bouger's law and is a constant of integration of the Euler equation for the ray path.

The first two terms on the RHS of eq. (7.17) are straight geometric distances along the deflected asymptotes of the ray path (see Figure 7-1). The third term is approximately the extra arc length due to the bending angle α at a radius a. The fourth term is a small correction resulting from the reduced speed of light in the atmosphere.

The bending angle α at each observational epoch is given in terms of the residual Doppler observation and the POD information from eq. (7.13) or from a power series in $(\Delta\rho/\dot{b})$, the first term of which is given by eq. (7.15). Likewise, the impact parameter a is given from eqs. (7.12) and (7.13). Thus, we obtain (α,a) for each observational epoch. From eq. (7.6a) for the spherically symmetric and, hence, coplanar case it follows that

$$\alpha(r_0) = -2a\int_{r_0}^{\infty} \frac{n'\,dr}{n\sqrt{n^2r^2 - a^2}} \tag{7.20}$$

to which the change of variable

$$\mu = nr \tag{7.21}$$

may be applied to obtain

$$\alpha(a) = -2a\int_a^{\infty} \frac{1}{n\sqrt{\mu^2 - a^2}}\left(\frac{dn}{d\mu}\right)d\mu \tag{7.22}$$

The integral in eq. (7.22) admits an Abel transform that yields n in terms of a and α. Multiplying $\alpha(\xi)$ in eq. (7.22) by $1/\sqrt{(\xi^2 - a^2)}$ and integrating with respect to ξ, obtains

$$\ln[n(a)] = \frac{1}{\pi}\int_a^{\infty} \frac{\alpha(\xi)d\xi}{\sqrt{\xi^2 - a^2}} \doteq 10^{-6}N(a) \tag{7.23}$$

where N(a) is the refractivity. From the Doppler-based $\alpha(\xi)$ sequence, eq. (7.23) yields (n,a) for each observational epoch and hence $n(r_0)$. The integral equations in eqs. (7.22) and (7.23) form an Abel transform pair and have been extensively used for analyses of the atmospheres and ionospheres of the terrestrial planets. Even for the oblate Jovian planetary atmospheres, which in general require a three-dimensional approach, a perturbation technique using the Abel transform pair as a reference has been fruitful. The Abel transform pair is also highly useful for accuracy analyses, including the assessment of the effects of unmodeled error sources on the recovered refractivity profile.

A Numerical Example

We assume an atmospheric refractivity that is exponentially decreasing with height. This model enables us to obtain "back-of-the-envelope" quantitative estimates of different physical properties of the atmosphere and the signal response. The refractivity model is given by

$$N(h) = N_0 \exp\left[-\frac{h}{H}\right] \tag{7.24}$$

where $h = r - R_E$ is the altitude, R_E is the *local* radius of the Earth, H is the scale height, and N_0 is the surface refractivity adjusted to yield the appropriate integrated vertical delay for L-band microwaves. Inserting eq. (7.24) into eq. (7.22), and using a *locally* straight-line approximation for the ray, we obtain a first-order expression for α

$$10^6 \alpha_o(h_a) = N(h_a)\sqrt{\frac{2\pi a}{H}} \qquad (7.25a)$$

where h_a is the altitude corresponding to a, i.e., the lowest altitude achieved by the ray ($h_a = r_0 - R_E$). This is a valid approximation for thin atmospheres as in eq. (7.3a) (i.e., $a / \mathscr{R} \doteq Na / H << 1$) in the absence of inversion layers. We can obtain a more accurate expression for $\alpha(h_a)$ by using eqs. (7.24) and (7.25a) in eq. (7.3) to obtain a first-order expression for the vertical deflection z of the ray path from its locally straight-line approximation.[6] The deflection is given to first order in $N(h_a)$ by

$$\begin{aligned} z &\doteq \frac{s\alpha(h_a)}{2}\operatorname{erf}\left[s/\sqrt{(2aH)}\right] \\ &\quad - aN(h_a)\left[1 - \exp\left(-s^2/2aH\right)\right]10^{-6} \qquad (7.25b) \\ &= \frac{10^{-6}}{2H}N(h_a)s^2 + \cdots \end{aligned}$$

where s is the arc length measured along the straight-line approximation; s = 0 corresponds to the point of closest approach of the ray to the Earth's surface where the actual ray is tangent to its straight-line approximation. If eq. (7.25b) is then used in eq. (7.22) to account for the departures of the ray path from the straight line, we obtain an expression for $\alpha(h_a)$ to second order in $N(h_a)$ that is given by

$$\alpha(h_a) = \alpha_o(h_a)\left[1 + \left(\sqrt{2} - 1\right)\frac{a}{H}N(h_a)10^{-6} + \cdots\right] \qquad (7.26)$$

The ray path curvature term in eq. (7.26) increases the value for $\alpha(h_a)$ (up to ~10%) because the actual downward bending ray travels at lower altitudes than its straight-line approximation. Equation (7.26) is accurate to better than 98% for the Earth's exponential refractivity model.

From eq. (7.17), $\Delta\rho(a)$ becomes

$$\begin{aligned} \Delta\rho(h_a) &= -R_L\left[\cos(\Theta - \alpha_T) - \cos\Theta\right] \\ &\quad + a\alpha_T + \sum H_\kappa \alpha_\kappa \end{aligned} \qquad (7.27a)$$

or upon expanding in powers of α

$$\Delta\rho(h_a) = \tfrac{1}{2}D\alpha_T^2 + \Sigma H_\kappa \alpha_\kappa + O(\alpha^3) \qquad (7.27b)$$

where $D = -R_L\cos\Theta$, which is the distance of the LEO from the point of closest approach to the Earth's limb or to the point of tangency of the ray, about 3000 km for the LEO at an altitude of $0.1R_E$. Here, $\alpha_T = \Sigma\alpha_\kappa$ = total bending, and the subscript κ denotes the contributions from different components of the atmosphere, which may have different scale heights. The first term on the RHS of eq. (7.27b) may be interpreted as the geometric increase in the length of the ray path asymptotes due to atmospheric bending; the second term represents the delay in traveling through the atmosphere due to the reduced speed of light. For Earth orbiters the bending term is dominant.

For the dry component of the atmosphere, a scale height of 8 km is taken and a total zenith delay of 208 cm is assumed. This scale height is close to the mean value for the U.S. Standard

[6] The expression in eq. (7.25b) predicts a very small departure of the ray path from a straight line for the dry air component of the Earth's thin atmosphere. (A locally large refractivity gradient or a multipath situation obviously requires special discussion.) For a ray just grazing the Earth's surface, when $s = \sqrt{(aH)} \doteq 225$ km, $z \cong 830$ m; when $s = 3\sqrt{(aH)} \doteq 675$ km, which corresponds to a point on the ray located in altitude at 4.5 scale heights above h_a and well above the bending region, $z \cong 2.7$ km.

Atmosphere up to the middle stratosphere. For the water vapor component, a scale height of 2.7 km is taken and a total zenith delay of 33 cm is assumed—a moderately tropical atmosphere. These parameters yield surface refractivities of 260 for dry air and 120 for the water vapor. Evaluating eqs. (7.26) and (7.27) for a ray grazing the Earth's surface, we obtain

$$\alpha = \begin{cases} 20 \text{ mrad (dry only)} \\ 36 \text{ mrad (dry plus wet)} \end{cases} \qquad \Delta\rho = \begin{cases} \sim 750 \text{ m (dry only)} \\ \sim 2.2 \text{ km (dry plus wet)} \end{cases} \tag{7.28}$$

For dry air and for $h_a = 0$ ($R_E = 6378$ km), the exact value of α based on a numerical integration of eq. (7.20) using eq. (7.24) is 20.23 mrad.

It should be stressed that these exponential models are only valid in the mean, particularly in the case of water vapor. Inversion layers cause near discontinuities in the refractivity gradient (see Figure 9-5). Moreover, water vapor, which is not in hydrostatic equilibrium, can vary in density by a factor of two over vertical distances that are small compared to its mean scale height.

Signal Dynamics and Defocusing

The exponential model for atmospheric refractivity provides useful insights into the dynamics of the signal. For example, although $\dot{b}$ is nearly constant during an occultation event, about 3 km/s, $\dot{a}$ varies significantly depending on the bending of the ray. It can be shown that

$$\dot{a} = \frac{\dot{b}}{1 + \frac{\alpha D}{H}} \tag{7.29}$$

It follows in the absence of inversion layers and other local inhomogeneities that $\dot{a}$ at limb grazing, even for dry air only, will be reduced to about 13% of its value at the top of the atmosphere. When water vapor is introduced in the lower troposphere, $\dot{a}$ can drop to a few percent of its upper atmospheric value, significantly lengthening the occultation event. In addition, using eq. (7.15) for ($R_G = \infty$), we obtain

$$\Delta\ddot{\rho} = -\dot{\alpha}\dot{b} - \alpha\ddot{b} \doteq -\dot{\alpha}\dot{b} \doteq \frac{\alpha\dot{b}^2}{H + \alpha D} \to \frac{\dot{b}^2}{D}, \text{ for large } \alpha \tag{7.30}$$

Also, defocusing effects on the signal can be evaluated. The dilution factor $\zeta(a)$ of the signal power at the LEO resulting from defocusing by the atmosphere of a coplanar ray path traversing a thin atmosphere such as the Earth's is given to better than 1% accuracy by the (thin screen) approximation

$$\zeta^{-1} = 1 - \frac{d\alpha}{dr} D = 1 + \frac{\alpha D}{H} = \dot{b}/\dot{a} \tag{7.31}$$

The quantity $\sqrt{\zeta}$ is also proportional to the diameter of the first Fresnel diffraction zone in a refracting medium (see Chapter 11).

Clearly the potentially large vertical gradients, particularly those associated with water vapor concentrations, will lead to significant accelerations and jerks in the carrier phase during the last 10 or 20 seconds before extinction and to well over 10 dB of signal diminution from defocusing. Inversion layers can cause a complete loss of signal for short periods (Hajj et al. 1994a). These are difficult conditions for the flight receiver—conditions that will be exacerbated by atmospheric multipath effects. An example of the effect of large vertical gradients (on $\sqrt{\zeta}$) for an actual atmospheric profile (taken by radiosonde measurements from Hilo, Hawaii, on July 11, 1991) is displayed in Figure 9-5.

Recovering Temperature and Pressure Profiles

Although refractivity may prove to be a useful parameter for monitoring global change and for

constraining meteorological models used in weather forecasting, the principal value of the radio occultation technique is likely to be the recovery of vertical profiles of density, pressure, and temperature. Local refractivity can be accurately converted into local mass density of the dry air if no water vapor is present. Then, assuming hydrostatic equilibrium, the local densities can be integrated downward from the top to obtain a local pressure and from the gas law, a local temperature. A complication arises below the tropopause with the presence of water vapor. The index of refraction per mole of water vapor for L-band signals at nominal surface temperatures is about 17 times larger than it is for dry air. Water vapor is not in hydrostatic equilibrium and, therefore, its distribution cannot be accurately modeled if present. Thus, water vapor in the lower troposphere will tend to mask the dry error signature and will significantly compromise most temperature results unless adequate water vapor modeling or other meteorological constraints are available. However, results for altitudes above ~6 km and for the entire troposphere in winter polar regions, which are virtually devoid of water vapor, would remain valid. On the other hand, for tropical oceanic regions where temperature distributions are better understood, an alternative strategy is to use *a priori* temperature models and pressure information to recover water vapor profiles. We have already noted the importance of accurate water vapor distributions in thermal energy transport and weather forecasting problems.

The refractivity of dry air at microwave frequencies is given by the linear relation (Smith and Weintraub 1953; Thayer 1974)

$$N_a = c_a n_a \qquad (7.32)$$

where n_a is the number density of dry air and $c_a = 1.056 \times 10^{-23}$ m^3. Here, the ideal gas law eq. (7.35) is assumed to apply. On the other hand, the spectrum of the water molecule exhibits broad resonances across virtually the entire microwave spectral range. As a consequence, the refractivity N_w of water vapor is more complicated, but it is given to sufficient accuracy by the Debye relationship

$$N_w = (c_w + d_w/T)n_w \qquad (7.33)$$

where T is the ambient temperature; here, $c_w = 0.840 \times 10^{-23}$ m^3 and $d_w = 5.14 \times 10^{-20}$ m^3 K. The total refractivity of dry air and water vapor admixture is obtained by combining eqs. (7.32) and (7.33). For the physical conditions prevailing in the Earth's atmosphere, this is given with an accuracy of better than 0.5% by

$$N = cn + dn_w/T \qquad (7.34)$$

where n is the total number density ($n = n_a + n_w$) and where $c = c_a = 1.056 \times 10^{-23}$ m^3 and $d = 5.08 \times 10^{-20}$ m^3 K.

For an ideal gas in thermodynamic equilibrium, its partial pressure P_x is given by

$$P_x = n_x kT \qquad (7.35)$$

where k is Boltzmann's constant. From eqs. (7.34) and (7.35) it follows that the refractivity for the neutral atmosphere is given in terms of local pressure and temperature by[7]

$$N = 77.6 \frac{P}{T} + 3.73 \times 10^5 \frac{P_w}{T^2} \qquad (7.36)$$

[7] Atmospheric refractivity at microwave frequencies also has a slight dependence upon the liquid water content in the medium resulting from a Rayleigh scattering mechanism (Kursinski et al. 1993). An extra term of the form +1.45 W should be appended to eq. (7.36), where W is in g/m^3. Under most conditions this effect is less than 0.5% of the combined contribution of dry air and water vapor; it is neglected here.

where P is the total pressure in millibars, P_w is the partial pressure of water vapor in millibars, and T is in kelvins.

For an atmosphere in hydrostatic equilibrium, the pressure change over an incremental vertical distance dh is given by

$$dP = -\mu g n\, dh \qquad (7.37)$$

where μ is the mean molecular mass of the atmospheric constituents and g is the local acceleration of gravity. If the value of $n(h)$ is known in the overlying atmosphere from the recovered refractivity profile, eq. (7.37) may be integrated downward to obtain the pressure profile; the temperature profile follows from eq. (7.35). Alternatively, the pressure term from eqs. (7.35) and (7.37) can be eliminated to obtain

$$n(h)T(h) = n(h_M)T(h_M) + k^{-1}\int_h^{h_M} \mu g n\, dh \qquad (7.38)$$

where h_M is the altitude at which a boundary value for the temperature is taken, nominally at 60 km. For the case of dry air only, its gaseous constituents are well mixed and μ may be taken outside of the integral. It follows from eq. (7.34) when $n_w = 0$ that the temperature profile is given by

$$T(h) = \frac{N(h_M)T(h_M)}{N(h)} + \frac{\mu \bar{g}}{kN(h)}\int_h^{h_M} N(h)\, dh \qquad (7.39)$$

where $\bar{g}$ is a mean value of gravity. The integral term in eq. (7.39) clearly shows the "small divisor" effect for high stratospheric altitudes where N(h) is nearly zero. Hence, the recovered value of T(h) will be more sensitive to error sources at these altitudes. Although some starting value for T must be adopted at h_M, eq. (7.39) also shows that the effect of an error in $T(h_M)$ on T(h) at lower altitudes is rapidly attenuated with depth and is virtually eliminated within a couple of scale heights. These equations will be used in Chapter 8 on error sources where water vapor will be treated as both an error source and a quantity to be recovered.

Recovering Water Vapor

If T(h) and P(h) are provided from ancillary information sources, for example, from models and synoptic meteorological data over tropical oceanic regions, then the vertical profile of water vapor density may be recovered from the Doppler occultation observations. From eqs. (7.34) and (7.36) it follows that the water vapor density at any observational epoch is given by

$$\hat{n}_w = \frac{1}{d}\left[\hat{N}(h)T(h) - \frac{cP(h)}{k}\right] \qquad (7.40)$$

where $\hat{N}(h)$ is the observed refractivity profile inferred from the Doppler observations. Errors in the recovered value of n_w as a result of errors in the assumed values for P and T, and also the reverse situation, will be discussed in Chapter 8.

Monitoring the Ionosphere

Although the ionosphere can cause significant errors through higher order effects in the recovered temperature profile, particularly in the upper stratosphere (see Chapter 8), valuable scientific information about ionospheric electron density distributions and currents can be obtained from the radio occultation observations as a by-product. In addition to the occulted satellite, the LEO should observe for POD all visible GPS satellites over the entire upper hemisphere, as well as those lying below the hemisphere and within the beam of the fore and/or aft high-gain antenna(s). This provides a rich data set made

from a platform traveling at about 7 km/s that can be combined with similar observations made from the ground by the GPS receivers in the global network. This flight and ground data set should enable tomographic studies of the ionosphere.

The columnar electron content, TEC, along the line-of-sight in terms of the phase delay measurements ρ_1 and ρ_2 (see eqs. (8.67) and (8.69) for the basic phase delay observable and its dependence on the ionosphere) is expressed by

$$\text{TEC} = \frac{1}{40.3}\left[\frac{f_1^2 f_2^2}{f_1^2 - f_2^2}\right][\rho_1 - \rho_2 + \text{cycle ambiguity term}] + \text{bias term} \qquad (7.41)$$

where f is the carrier frequency and TEC is in units of electrons/m^2. The same expression is obtained from the P1 and P2 group delay measurements except with a reversal in sign and without the cycle ambiguity term. The phase measurements enable us to obtain with very high accuracy the change in TEC with time (to about $\pm 3 \times 10^{13}$ electrons/m^2 on 1 s samples). The group delay measurements smoothed against the phase-connected carrier phase measurements provides an accurate determination of the biased TEC (to about $\pm 10^{15}$ electrons/m^2 on 1-s samples). The bias term in the P-code group delay measurements arises from departures from synchronization of the L1 and L2 P-code modulations in the exciter chains in the GPS satellite transmitter and antenna. This bias term, which differs with each satellite, can be calibrated somewhat ($\pm 3 \times 10^{15}$ electrons/m^2, see Gaposchkin and Coster 1993) or it can be avoided with a differential strategy that uses concurrent observations of the same GPS satellite made from different receivers. The cycle ambiguity term in eq. (7.41), which is the difference in the individual integer cycle ambiguities in the L1 and L2 carrier phase measurements, is not easily evaluated absolutely because of clock instabilities in the GPS satellites and in the receivers. However, it is possible under certain conditions to resolve double-differenced cycle ambiguities, even on intercontinental baselines on the ground (Blewitt et al. 1992). For LEO–ground baselines the cycle ambiguity problem is more challenging because of the rapid motion of the LEO.

CHAPTER

8 SYSTEM ACCURACY AND ERROR SOURCES

This chapter addresses the accuracies with which various atmospheric quantities may be recovered. We will consider SNR limitations along with a number of significant error sources that arise from limitations in modeling and from environmental sources, e.g., the ionosphere, reference oscillators, multipath, and POD. We also will address the problem of water vapor and its potential recovery. Our approach is to first use the Abel transform pair to develop analytic expressions for the SNR errors and then to quantify the effect of certain error sources on the recovered refractivities and temperatures. Numerical simulations and error sensitivity analyses are presented for certain parameters in Chapter 9.

SNR Error for Refractivity

Here we use an atmospheric model that is spherically symmetric and thin, which allows us to use the Abel transform approach combined with a ray path that is approximated by a locally straight line. Also, the exponential refractivity model (either eq. (7.24) or the refined version in eq. (8.38)) is used as a reference. To simplify the computations further, we assume that the emitting spacecraft is at infinity. Even with these approximations and simplifications, the error analyses will usually yield results that are correct to within a factor of two.

From eq. (7.23) it follows that the error in the index of refraction resulting from errors in the bending angle series inferred from the Doppler observations is given by

$$\frac{\delta n(a)}{n(a)} = \frac{1}{\pi}\int_a^\infty \frac{\delta\alpha(\xi)d\xi}{\sqrt{\xi^2 - a^2}} \qquad (8.1)$$

Also, from eq. (7.15) and Figure 7-1, it follows that the bending angle error is given by

$$\delta\alpha \doteq \delta(\Delta\dot{\rho}) / \dot{b} \qquad (8.2)$$

Here $\dot{b}$ is modified to account for the noncoplanarity of the geocentric velocity vectors of the LEO and the emitting GPS satellite. Noncoplanarity effects cause the point of geocentric closest approach of the ray to descend or ascend off-vertical, which on average reduces $\dot{b}$ by about 10% for a 45-deg slant angle limit. Here, we use a nominal value for $\dot{b}$ of 3 km/s. It follows that the error in $\hat{n}(a)$, the recovered value of the index of refraction, is given by

$$\frac{\delta\hat{n}(h_a)}{n(h_a)} = \frac{1}{\pi\dot{b}}\int_0^{\tau_M} \frac{\delta(\Delta\dot{\rho}(t))\dot{\xi}}{\sqrt{\xi^2 - a^2}}dt \doteq 10^{-6}\delta\hat{N}(h_a) \qquad (8.3)$$

where $h_a = a - R_E$ is the altitude; τ_M is the time interval required for the impact parameter of the deflected ray to evolve from the value a to its

maximum value a_M, which nominally corresponds to an altitude of 60 km where, essentially, $a_M = b_M$. Thus, τ_M is given by

$$\tau_M = \int_a^{a_M} d\xi / \dot{\xi} \doteq [a_M - b(a)] / \dot{b} \qquad (8.4)$$

where $\dot{b}$ is assumed to be a constant over the ~1 min duration of the occultation event. Here, b(a) is the undeflected impact parameter and is given by

$$b(a) \doteq a - D\alpha(a) \qquad (8.5)$$

where D is the distance of the LEO from the point of closest approach of the ray to the Earth's surface.

The expression in eq. (8.3) provides a fundamental error equation for assessing the effect of various error sources through their signatures on $\Delta\dot{\rho}$. This includes errors from receiver tracking, POD, ionospheric correction, multipath, and frequency drift in the reference oscillators.

For SNR-based errors, it is convenient to consider $\delta(\Delta\dot{\rho})$ as a discrete stochastic process. Over each correlation interval the GPS receiver, if it has a Rogue or TurboRogue digital signal processing architecture, measures a residual phase, which is the difference between the observed phase and a predicted phase based upon a current model embedded in the receiver software. For L1 the model is based upon previous phase measurements and includes a Doppler term; hence, it provides the digital realization of a second-order tracking loop. The residual phase measurement is obtained from the real and the imaginary correlator outputs of the digital signal processor after averaging over a 20-ms time interval, which is the chip period of the GPS data message. The measured residual phase may be further averaged over the reporting interval $\Delta\tau$, which is an integer multiple of the averaging period, currently set to 20 ms. The measured residual phase is added to the predicted phase to yield a current observed phase. The observed phase is then reported at the observational epoch corresponding to the center of the reporting interval. For the Rogue/TurboRogue architecture, $\Delta\tau$ can be as short as 20 ms. Therefore, the observed phase is a piecewise constant process, taking on a fixed value over the reporting interval $\Delta\tau$ associated with an observational epoch; it has an expected error that is also fixed for that interval and is statistically independent with respect to the errors in the reported phases associated with all other observational epochs.

There is no reason why the receiver could not also measure the residual Doppler over the same correlation interval since the outputs of the correlators also depend on the residual Doppler. The implementation would be somewhat more complicated than the simple arctan routine used to extract residual phase from the phasor output of the real and imaginary correlators because the residual Doppler appears as a sin(x)/x function in the amplitude of the correlator outputs. We can construct alternative measurement schemes like parallel "fast/slow/synch" multiple lags for Doppler and phase extraction that are analogous to the "early/late/prompt" gate technique used for code-based group delay and phase tracking. Here, the fast and slow correlators would use model values for the Doppler that are equally spaced about the value used in the synch correlator. We would adjust the prompt model Doppler to drive the difference between the outputs for the fast and slow correlators to zero. This provides a very sensitive measurement of the residual Doppler.

Most GPS applications (the occultation application being a notable exception) do not require the GPS receiver to report the Doppler measurement. Consequently, receiver manufacturers have paid little attention to Doppler. Doppler is re-

ported in the TurboRogue, but the current operating software in the receiver obtains the Doppler from a linear combination of phase measurements, nominally a quadratic fit to the phases obtained from three or more correlation intervals. To preserve statistical independence in the time series, the reporting interval can be no shorter than the fitting interval, which for the TurboRogue is at least 60 ms for a quadratic fit. This reporting interval could be shortened to 40 ms for a linear fit. The "fast/slow/synch" scheme offers finer temporal resolution (but also results in an increased penalty in precision) by allowing the reporting of Doppler measurements that are statistically independent at every correlation interval. The finer resolution may be important to achieve a Nyquist sampling rate for lower troposphere soundings where atmospheric-induced multipath may lead to multiple tones.

We now invoke the discrete character of $\delta(\Delta\dot{\rho})$; hence, eq. (8.3) can be written as a sum

$$\delta\hat{N}(h_a) = \frac{10^6}{\pi\dot{b}} \sum_{k=1}^{M(h_a)} \delta(\Delta\dot{\rho}_k) \ln\left[\frac{\xi_k + \sqrt{\xi_k^2 - a^2}}{\xi_{k-1} + \sqrt{\xi_{k-1}^2 - a^2}}\right] \tag{8.6}$$

where k denotes the *k*th reporting interval of length $\Delta\tau$, where $\xi_0 = a$ and $\xi_M = a_M$, and where M, which is the maximum integer value of k and a function of h_a, is given by $M(h_a) = \mathrm{INT}[\tau_M/\Delta\tau]$.

The quantity ξ_k is obtained from eq. (8.5) for vertically descending or ascending profiles, which may be rewritten as

$$h_k - h_a - D[\alpha(h_k) - \alpha(h_a)] \doteq b_k - b_a \doteq |\dot{b}| t_k = |\dot{b}| k\Delta\tau, \quad k = 1,2,\cdots,M(h_a) \tag{8.7}$$

where $h = \xi - R_E$ is the altitude of the point ξ, and $b_a = b(a)$.

The error in $\delta(\Delta\dot{\rho}_k)$ is assumed to obey white Gaussian statistics with mean zero. Hence,

$$\left\langle \delta(\Delta\dot{\rho}_k)\delta(\Delta\dot{\rho}_j) \right\rangle = \sigma_{\dot{\rho}}^2 \delta_{kj} \tag{8.8}$$

For either the quadratic fit or the "fast/slow/synch" approaches used for extracting Doppler, the effective standard deviation of the Doppler measurement in terms of the standard deviation of the L1 phase measurement is given by

$$\sigma_{\dot{\rho}} = 2 \times 3 \times 2\sqrt{3} \frac{1}{\Delta\tau} \sqrt{\frac{1\,\mathrm{sec}}{\Delta\tau}} \frac{\sigma_{\rho_o}}{\sqrt{\zeta}} \tag{8.9}$$

where σ_{ρ_o} is the nominal SNR-based standard deviation in the L1 phase measurement over a 1-s averaging interval. The factor $1/\sqrt{\zeta}$ accounts for the loss of SNR resulting from atmospheric defocusing (see eq. (7.31)), which is particularly significant in the lower troposphere. The factor $2\sqrt{3}$ is the ratio of the Doppler to phase extraction from the correlator outputs either using a quadratic fit to the phase observations or the "fast/slow/prompt" strategy.[1] The factor 3 is the magnification factor resulting from linearly combining L1 and L2 to obtain a correction for the ionosphere (3.3 is a better value to account for the 3 dB less of power in the L2 signal than in the L1 signal). Depending on the level of short-term ionospheric variability, this factor of 3 possibly can be improved down to a lower bound of $1/\sqrt{2}$ (but probably more typically $\sqrt{3}$) (Wu et al. 1993) using short-term smoothing techniques on the ionospheric term; for example, by treating it as a first-order Markov process. The factor 2 arises from the elimination of reference oscillator errors in the LEO and GPS satellite by treating the clock

[1] Strictly speaking, for a quadratic fit the factor $2\sqrt{3}$ should be replaced by $2m\sqrt{[3/(m^2-1)]}$ where m is the number of 20-ms correlation intervals spanned by the quadratic fit. For the "fast/slow/synch" approach, the factor for optimal spacing of the fast and slow tones is $2\sqrt{3}$. Hence, the latter technique yields a somewhat smaller factor than the quadratic fit approach if the correlation interval is extended to $\Delta\tau$.

epochs as white noise processes (analogous but not equivalent to "double differencing"). If the LEO carries an ultrastable oscillator (USO) and the ground stations are comparably equipped, only a "single difference" strategy is required to eliminate the GPS satellite clock error; therefore, the factor could be reduced to $\sqrt{2}$. Using the TurboRogue with P-code tracking under nominal signal conditions for the Block II GPS satellites and with 0 dB antenna gain, $\sigma_{\rho_o} = 0.1$ mm at 1-s averaging for the L1 carrier phase. To keep in mind the relative coarseness of these numbers, note that a 3-dB increase in antenna gain, a relatively easy factor to achieve, improves σ_{ρ_o} by $1/\sqrt{2}$.

Squaring eq. (8.6) and taking its expected value, we therefore obtain

$$\sigma_{N(h_a)} = 10^6 \frac{12\sqrt{6}}{\pi\sqrt{a\dot{b}\Delta\tau}} \sigma_{\rho_o} F(\Delta\tau, h_a) \quad (8.10)$$

where F (Figure 8-1) is a dimensionless weighting function resulting from the spherical geometry and from a vertical atmospheric sampling interval that varies as a result of the variation in $\dot{a}$ with altitude. The weighting function F is of the order of unity and is given by

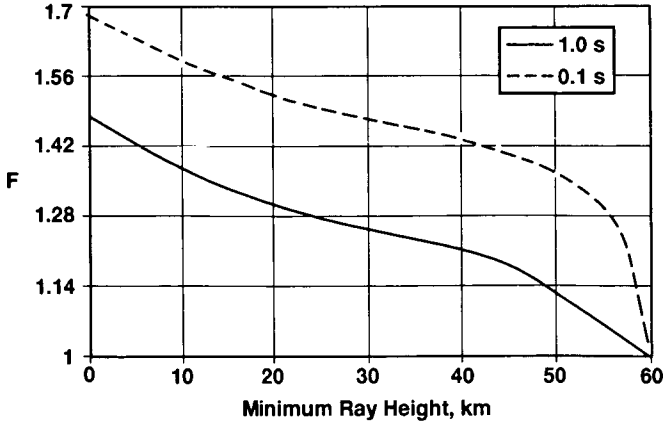


Figure 8-1. Refractivity weighting function $F(\Delta\tau, h_a)$ for a constant temporal sampling period of 1 s or 0.1 s with $h_M = 60$ km.

$$F^2(\Delta\tau, h_a) = \frac{a}{2\dot{b}\Delta\tau} \sum_{k=1}^{M(h_a)} \ln^2\left[\frac{\xi_k + \sqrt{\xi_k^2 - a^2}}{\xi_{k-1} + \sqrt{\xi_{k-1}^2 - a^2}}\right] \Big/ \zeta_k \quad (8.11)$$

where ζ_k is the mean value of the defocusing factor on the kth interval. Let the quantity η be defined by $\eta_k = (h_k - h_a)$; then $\eta_k/a << 1$ and the limiting form of F becomes

$$F^2(\Delta\tau, h_a) \doteq \frac{1}{\dot{b}\Delta\tau} \sum_{k=1}^{M(h_a)} \left[\sqrt{\eta_k} - \sqrt{\eta_{k-1}}\right]^2 / \zeta_k \quad (8.12)$$

For small $\Delta\tau$

$$F^2(\Delta\tau, h_a) \to 1 + \tfrac{1}{4} \ln\left[\frac{h_M - h_a}{\dot{a}\Delta\tau}\right] \quad (8.13)$$

where $\dot{a}$ is given by eq. (7.29) and evaluated at h_a.

The relative accuracy from SNR phase errors predicted by eq. (8.10) (at a sampling rate of 1 s and 0 dB of antenna gain) for the refractivity recovery at sea level is about 40 ppm, degrading to about 1% at a 50-km altitude. Note that σ_N is proportional to $F/\Delta\tau$, which contains a $1/\sqrt{\Delta\tau}$ term to account for correlator averaging, and another $1/\sqrt{\Delta\tau}$ "random walk" term arising from the number of vertical slices of the atmosphere that are sampled. In other words, for a fixed altitude, the SNR precision of the recovered refractivity times the vertical sampling interval is essentially constant (ignoring the weak logarithmic dependence of F on $\Delta\tau$).

SNR-Based Accuracy of the Temperature Profile

From eq. (7.39) (ignoring the initialization error at h_M), it follows that the error in $\hat{T}(h)$, the recovered temperature profile due to Doppler measurement errors, is given by

$$\delta\hat{T}(h) = \frac{\mu\bar{g}}{kN(h)}\left[\int_h^{h_M}\delta N(h')dh' - \frac{\delta N(h)}{N(h)}\int_h^{h_M} N(h')dh'\right] \tag{8.14}$$

The first integral in eq. (8.14) can be evaluated (applying the same approximations used earlier)

$$\int_h^{h_M}\delta N(h')dh' = \frac{10^6}{\pi\dot{b}}\int_h^{h_M}\delta(\Delta\dot{\rho})\cos^{-1}\left(\frac{R_E+h}{R_E+h'}\right)dh' \tag{8.15}$$

and the second integral is simply

$$\int_h^{h_M} N(h')dh' \doteq HN(h) \tag{8.16}$$

when the exponential model in eq. (7.24) is used for N(h). Squaring eq. (8.14), taking the expected value, and using eqs. (8.5) through (8.9) obtains

$$\sigma_{T(h_a)} = 10^6 T^*(h_a)\frac{12\sqrt{6}}{\pi}\frac{\sigma_{\rho_o}}{\Delta\tau}\frac{G(\Delta\tau,h_a)}{N(h_a)}\sqrt{\frac{1\,\text{sec}}{\dot{a}\dot{b}}} \tag{8.17}$$

where $T^*(h_a)$ is obtained from eq. (7.39), that is,

$$T^*(h_a) = T(h_a) - T(h_M)N(h_M)/N(h_a) \tag{8.18}$$

is essentially $T(h_a)$ for h_a more than two scale heights below h_M. The quantity G (Figure 8-2) is a dimensionless weighting function similar to F. For $\eta_k/a = (h_k - h_a)/a << 1$, G is given in its limiting form by

$$G^2(\Delta\tau,h_a) = \frac{1}{\dot{b}\Delta\tau}\sum_{k=1}^{M(h_a)}\left[\left[\sqrt{\eta_k}-\sqrt{\eta_{k-1}}\right]^2 - \frac{4}{3H}\left[\sqrt{\eta_k}-\sqrt{\eta_{k-1}}\right]\left[\sqrt{\eta_k^3}-\sqrt{\eta_{k-1}^3}\right] + \frac{4}{9H^2}\left[\sqrt{\eta_k^3}-\sqrt{\eta_{k-1}^3}\right]^2\right]\Big/\zeta_k \tag{8.19}$$

For small $\Delta\tau$

$$G^2 \to \left[F^2(\Delta\tau,h_a) - \frac{1}{H}(h_M - h_a) + \frac{1}{2H^2}(h_M - h_a)^2\right] \tag{8.20}$$

Figure 8-3 is a plot of σ_T versus h_a (using $T(h_a)$ from the U.S. Standard Atmosphere) and clearly shows the sub-kelvin precision of the temperature recovery below an altitude of 40 km with $\Delta\tau$ = 1-s sampling and 0-dB antenna gain. A cutoff altitude h_M of 60 km has been assumed. Figure 8-3 displays σ_T for both an austere SNR scenario and a more sanguine one. Clearly, for a given vertical resolution, the use of (a) higher gain fore and/or aft antenna arrays, (b) a USO in the LEO to avoid double differencing, and (c) some degree of ionospheric smoothing would result in greatly improved precision of the temperature recovery in the SNR-limited upper stratosphere. For a

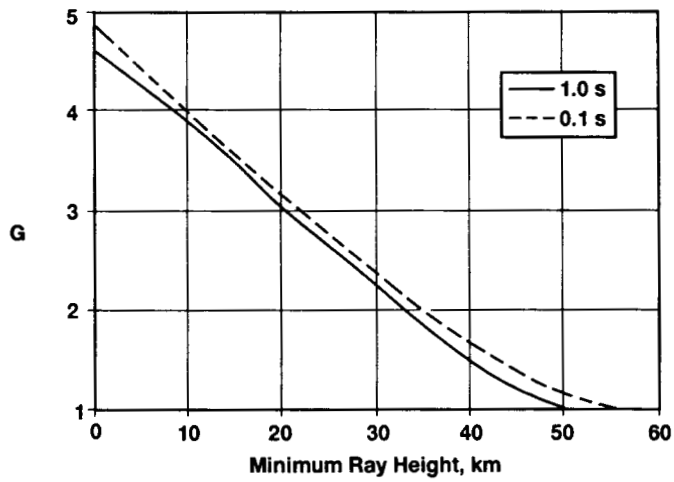


Figure 8-2. Weighting function $G(\Delta\tau,h_a)$ for SNR-limited accuracy of temperature recovery. Constant temporal sampling period of 1 s or 0.1 s with h_M = 60 km.

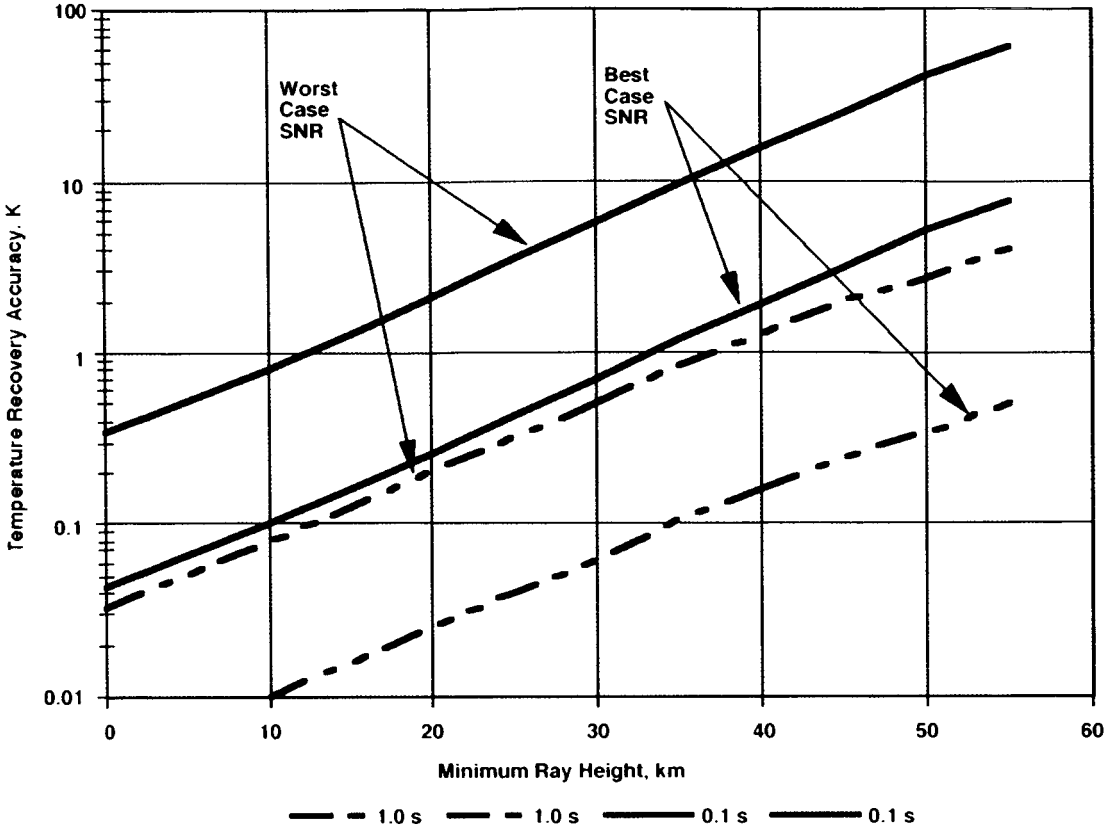


Figure 8-3. SNR-based accuracy of the recovered temperature profile using a sampling period of 1 s or 0.1 s assuming code-based in-phase and quadrature tracking and nominal signal conditions. Worst-case SNR assumes 0 dB antenna gain, no ionospheric smoothing, and a double differencing scheme for clock error elimination. Best-case assumes 10 dB gain, ionospheric smoothing, and single differencing.

constant temporal sampling interval $\Delta\tau$, the vertical atmospheric sampling interval varies with altitude as ζ, as given by eq. (7.31). Figure 8-4 shows the variable spatial sampling versus the minimum altitude of the ray path for dry air, which results from the variable rate of descent or ascent of the ray path through the atmosphere. Hence, the vertical sampling interval varies by nearly one order of magnitude when a constant temporal sampling interval is used; also, the diameter of the first Fresnel zone varies by a factor of three. A modified strategy would use a variable temporal sampling interval $\Delta\tau_k$. In this case, the weighting function F would be generalized to

$$F'^2(h_a) = \frac{a\Delta\tau_o^2}{2b} \sum_{k=1}^{M(h_a)} \ln^2\left[\frac{\xi_k + \sqrt{\xi_k^2 - a^2}}{\xi_{k-1} + \sqrt{\xi_{k-1}^2 - a^2}}\right] \Big/ (\Delta\tau_k)^3 \zeta_k \tag{8.21}$$

To obtain σ_N, we would replace $\Delta\tau$ in eq. (8.10) with $\Delta\tau_o$, where $\Delta\tau_o$ is a fixed-reference sampling interval (e.g., 1 s), and use eq. (8.21) for F. Similar expressions would result for G and σ_T.

The near-linear trade-off between sampling interval and precision suggests a strategy that uses coarser temporal sampling in the upper stratosphere to improve SNR-based precision but finer

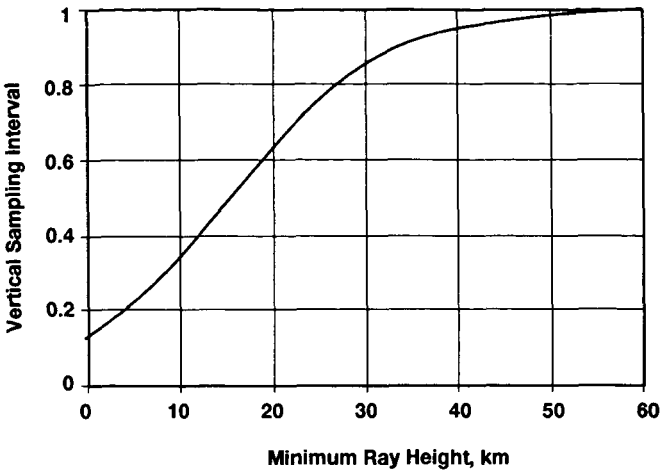


Figure 8-4. Relative vertical atmospheric sampling interval for a constant temporal sampling interval with an exponential atmosphere and a scale height of 8 km. For a sampling interval of 1 s, the vertical interval is 3 km at an altitude of 60 km for an LEO orbital radius of 7000 km.

sampling in the lower atmosphere, where SNR is not limiting, to achieve the highest resolution.

The Effect of POD Errors

Over the 1-min duration of the occultation event, POD errors will manifest themselves essentially as a linearly varying error in the Doppler of the form

$$\delta(\Delta\dot{\rho}) = \delta\dot{\rho}_o + \delta\ddot{\rho}_o t \tag{8.22}$$

where $\delta\dot{\rho}_o$ and $\delta\ddot{\rho}_o$ are constants and t denotes the time interval from the onset of the occultation event ($a = a_M$). Renormalization of the Doppler residual profile to zero at $a = a_M$ should mostly eliminate the effect of the $\delta\dot{\rho}_o$ term, but it is included here for completeness. It follows from eq. (8.3) that the error in the recovered refractivity for this case is given by

$$10^{-6}\delta\hat{N}(h_a) = \frac{\delta\dot{\rho}_o}{\pi\dot{b}}\ln\left[\frac{a_M + \sqrt{a_M^2 - a^2}}{a}\right] + \frac{\delta\ddot{\rho}_o}{\pi\dot{b}^2}\left[\sqrt{a_M^2 - a^2} + DN_o \exp\left[-\frac{h_a}{H}\right]\right] \tag{8.23}$$

Similarly, it follows from eq. (8.14) that the temperature error in a power-series is given by

$$\delta\hat{T}(h_a) = \frac{T^*(h_a)}{\pi\dot{b}}\frac{10^6}{N(h_a)} \times \left[\delta\dot{\rho}_o\left\{-1+\frac{2}{3}\left(\frac{h_M - h_a}{H}\right)\right\}\left(\frac{2(h_M - h_a)}{R_E}\right)^{1/2} + \frac{\delta\ddot{\rho}_o R_E}{3\dot{b}}\left\{-1+\frac{2}{5}\left(\frac{h_M - h_a}{H}\right)\right\}\left(\frac{2(h_M - h_a)}{R_E}\right)^{3/2}\right] + \cdots \tag{8.24}$$

Figure 8-5 displays the sensitivity of the recovered temperature profile to POD errors in Doppler (Ks/mm) and acceleration (Ks^2/mm) for a maximum cutoff altitude of h_M = 60 km. An exponential model for refractivity has been assumed. Alternative models for the refractivity profile and cutoff altitude will change the shapes of the curves somewhat, but their scale should remain about the same. A once-per-revolution orbit error of 3 cm, approximately the current state of the art, maps into amplitudes for the velocity and acceleration errors of about 0.03 mm/s and 0.03 $\mu m/s^2$, respectively. Thus, eq. (8.24) shows that the POD errors can be made negligible for the temperature recovery except near the highest altitudes, if a high-accuracy orbit determination methodology is followed. A similar conclusion follows from examination of POD errors on the derived orbital positions and angles such as Θ, ψ, and β in Figure 7-1.

Multipath Errors

Figure 8-5 is also useful for quantifying the error in temperature recovery that results from multipath effects primarily at the LEO but also at the ground stations. During an occultation, the change in angular direction of the occulted GPS satellite as seen from the LEO is about 4 deg. Far-field—and, hence, short-period—multipath effects should be quite low for the LEO. However, near-field multipath, that is, multipath from nearby

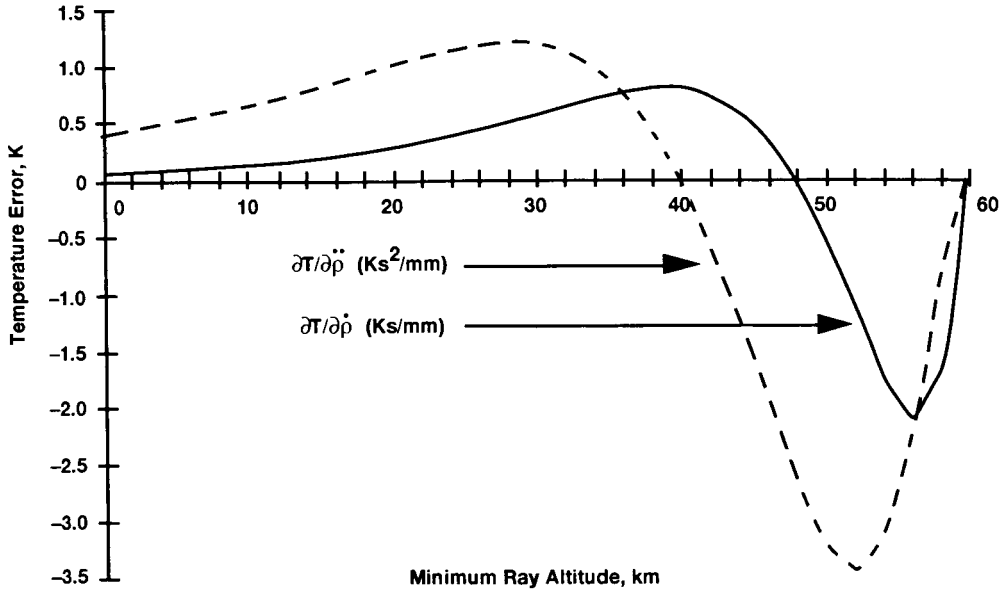


Figure 8-5. Sensitivity of recovered temperature to errors in Doppler and acceleration based on exponential model for atmospheric refractivity (8-km scale height) and a cutoff altitude of 60 km.

objects with spectral power primarily at long periods (e.g., min), can be significant relative to SNR errors.

The Effect of Along-Track Variations of Refractivity

One of the major questions concerning the radio occultation technique is its horizontal or along-track resolution. Resolution topics are discussed in Chapters 10 and 11. In this section, we evaluate the sensitivity of the recovered refractivity and temperature to along-track variations in the true refractivity that departs from the assumed spherically symmetric distribution.

Let the true refractivity N_t be given by

$$N_t = N_m(h)[1 + \varepsilon(h,\theta)] \qquad (8.25)$$

where N_m is the assumed spherically symmetric refractivity distribution, nominally the exponential model given in eq. (7.24). Here $N_m(h)\varepsilon(h,\theta)$ represents the spatial variability of the refractivity, that is, the departures from the spherically symmetric model.

From eq. (7.6a) the "error" in the bending angle—that is, the difference between the bending angle resulting from the spherically symmetric distribution and the true bending angle—is given by

$$\delta\alpha = \int_{\mathscr{E}_m} n_m^{-1}(\mathbf{T}\times\nabla n_m)ds - \int_{\mathscr{E}_t} n_t^{-1}(\mathbf{T}\times\nabla n_t)ds \qquad (8.26)$$

where the subscripts m and t denote the model and the true version, respectively; $\mathscr{E}_m$ and $\mathscr{E}_t$ denote the corresponding ray paths that are followed. For the case of a thin atmosphere with low refractivity gradients, i.e., no inversion layers, this distinction in ray paths is negligible for evaluating $\delta\alpha$.

It is convenient to use a separation of variables approach by expressing $\varepsilon(h,\theta)$ as the product of two functions, $\varepsilon(h,\theta) = z(h)f(\theta)$. We will consider two extreme cases for the functional form of z(h): First, z(h) is non-zero only within a single anomalous refracting layer of the atmosphere of thick-

ness Δh and located at an altitude of h_o. We assume that $\Delta h << H$, the local scale height of the refractivity. Thus, the variability in refractivity is highly localized in altitude. Second, $z(h) = 1$, in which case the horizontal variability of the refractivity is uniform in altitude. It is also convenient to represent $f(\theta)$ in a polynomial form such as

$$f(\theta) = \beta_0 + \beta_1\theta + \frac{1}{2!}\beta_2\theta^2 + \cdots \qquad (8.27)$$

From eq. (7.5) it follows that the differential bending angle at any point on the ray path is sensitive only to the component of the refractivity gradient perpendicular to the ray path. Consequently, although $f(\theta)$ does introduce an along-track or horizontal gradient through the β_1 term, its principal contribution to the bending angle is through the variation it causes in the radial component of the gradient. In addition, because the weighting function or kernel in eq. (7.6a) of the form $N_m(r)/[n^2r^2 - n_o^2r_o^2]^{1/2}$ is an even function about the center line or mid-point of the ray path through the atmosphere, only even powered terms of $f(\theta)$ corrupt the recovered refractivity.

First consider Case 1 where the variability in refractivity is confined within a single layer at a mean altitude h_o and of thickness Δh (Figure 8-6). The thickness Δh equals $h_2 - h_1$. For $h_a < h_1$, the ray passes through the layer twice; therefore, the resulting bending angle $\delta\alpha(h_a)$ for a given altitude of closest approach h_a is given by

$$\delta\alpha(h_a) = \begin{cases} 2N_m(h_o)\left[f(\theta_2)\cot\theta_2 - f(\theta_1)\cot\theta_1\right] + O\left(N_m^2\right) & \text{for } h_a < h_1 \\ 2N_m(h_o)f(\theta_2)\cot\theta_2 + O\left(N_m^2\right) & \text{for } h_1 \le h_a < h_2 \end{cases} \qquad (8.28)$$

where θ_2 and θ_1 are, respectively, the central angles of the intersection points of the ray having a closest-approach altitude of h_a with the outer boundary at an altitude h_2 and the inner boundary at h_1.[2] These are given by

$$\begin{cases} \cos\theta_2 = (R_E + h_a)/(R_E + h_2) \\ \cos\theta_1 = (R_E + h_a)/(R_E + h_1) \end{cases} \qquad (8.29)$$

Inserting eq. (8.28) into eq. (8.1), we obtain the change in $\hat{N}(h_a)$ due to the horizontal variability $N_m(h_o)f(\theta)$ within the single layer. For β_0, which is not strictly a horizontal variation but rather a spherically symmetric departure from N_m, we obtain

$$\frac{\partial\hat{N}}{\partial\beta_0} = \begin{cases} 0 & \text{for } h_a < h_1 \\ N(h_o) & \text{for } h_1 \le h_a < h_2 \\ 0 & \text{for } h_2 \le h_a \end{cases} \qquad (8.30)$$

For the corresponding error in the recovered temperature and using eqs. (8.30) and (8.14), we obtain

$$\frac{1}{T}\frac{\partial\hat{T}}{\partial\beta_0} = \begin{cases} \dfrac{N(h_o)}{N(h_a)}\dfrac{\Delta h}{H} \doteq 0 & \text{for } h_a < h_1 \\ -1 + \dfrac{h_2 - h_a}{H} \doteq -1 & \text{for } h_1 \le h_a < h_2 \end{cases} \qquad (8.31)$$

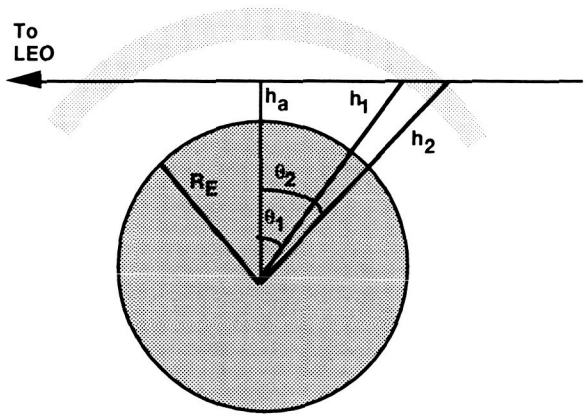


Figure 8-6. Geometry for anomalous layer.

[2] For either θ_1 or θ_2 near zero, eq. (8.28) breaks down and the second-order terms should be included. Fortunately, the integration in the Abel transform does not require the second-order terms to recover $\delta\hat{N}$.

The physical interpretation of the temperature result in eq. (8.31) follows from the condition of hydrostatic equilibrium, which requires that a variation in local density $\delta N/N$ be compensated by a corresponding temperature variation $-\delta T/T$ within the anomalous layer because there can be no appreciable change in pressure there.

For a quadratic horizontal variation in refractivity within the anomalous spherical layer, the error in recovered refractivity is given by

$$2a^2 \frac{\partial \hat{N}}{\partial \beta_2} = \begin{cases} N(h_o) a \Delta h & \text{for } h_a < h_1 \\ N(h_o) a (h_2 - h_a) & \text{for } h_1 \le h_a < h_2 \end{cases} \tag{8.32}$$

The quantity $[2a(h_2 - h_a)]^{1/2}$ is just the one-sided along-path distance within the anomalous layer; thus, the error in $\hat{N}$ is half of the total variation in refractivity along the ray. For the error in the recovered temperature, we obtain

$$\frac{2a^2}{T} \frac{\partial \hat{T}}{\partial \beta_2} = \tag{8.33}$$
$$\begin{cases} -\frac{N(h_o)}{N(h_a)} a \Delta h \left[1 - \frac{h_2 - h_a}{H}\right] & \text{for } h_a < h_1 \\ -a(h_2 - h_a)\left[1 - \frac{h_2 - h_a}{H}\right] \doteq -a(h_2 - h_a) & \text{for } h_1 \le h_a < h_2 \end{cases}$$

The upper relation in eq. (8.33) shows that the error in $\hat{T}(h_a)$ decays exponentially with altitude below the anomalous layer.

For Case 2 where $z(h) = 1$ throughout the atmosphere, the situation is substantially different. Evaluating eq. (8.26) using the exponential refractivity model (eq. (7.24)), and a locally straight-line approximation for the ray path, we obtain an approximation for the departure of α from its spherically symmetric value that is given by

$$\delta\alpha(h_a) = \alpha_m(h_a) \sum_{k=0} \beta_k \Pi_k \left[\frac{H}{a}\right]^{k/2} \tag{8.34}$$

where

$$\Pi_k = \begin{cases} 1 & k = 0 \\ \dfrac{2}{\left(\frac{k-2}{2}\right)! k 2^{k/2}} & k = \text{even} \\ 0 & k = \text{odd} \end{cases} \tag{8.35}$$

and where $\alpha_m(h_a)$ is the spherically symmetric model-based value for the bending angle.

When the exponential refractivity model is used as a reference, the variation with height of this particular bending angle "error" is proportional to the nominal bending angle itself. Because α and N form an Abel transform pair, it follows that the error in the recovered refractivity $\delta\hat{N}$ is given by

$$\delta\hat{N}(h_a) = N_m(h_a) \sum_{k=1} \beta_k \Pi_k \left[\frac{H}{a}\right]^{k/2} \tag{8.36}$$

For example, a quadratic variation in refractivity that amounts to X% over an along-track distance L from the center line results in a change in the recovered refractivity $\delta\hat{N}_2$, given by

$$\frac{\delta\hat{N}_2}{N_m} = \frac{XaH}{100L^2} \tag{8.37}$$

Thus, a 1% quadratic variation in true refractivity over a distance from the center line of 300 km results in a corresponding 0.5% change in the Doppler measurements and a 0.5% change in the bending angle and recovered refractivity.

For the error in the recovered temperature, the exponential refractivity model, which implies an isothermal atmosphere, yields a null result. This unrealistic result is a consequence of the peculiarities of the exponential refractivity model, which causes a complete cancellation of the two principal terms in eq. (8.14) because the error in $\hat{N}$ as a function of altitude varies linearly with

$\hat{N}$ itself. Therefore, a more realistic refractivity profile is required as a reference.

For example, suppose the scale height of the refractivity model is a slowly varying and essentially linear function of height. In this case $N_m(h)$ may be given by

$$N_m(h) = N_m(0)\ \exp\left[-\left(\frac{h}{H_o} + gh^2\right)\right] \quad (8.38)$$

With H_o, the scale height at sea level, fixed at around 9 km and g at around 0.001 km^{-2}, this model corresponds more closely to the U.S. Standard Atmosphere, at least through the troposphere. Unlike the isothermal atmosphere implied by the purely exponential model, this model yields to first order a linearly decreasing temperature with height. From eq. (7.39) it can be shown that the temperature profile is given by

$$T_t(h) = T_t(0)\left[1 - 2\left(gH_o^2\right)\left(\frac{h}{H_o}\right)\right] + O\left[\left(gH_o^2\right)\left(\frac{h}{H_o}\right)\right]^2 \quad (8.39)$$

which is close (for $gH_o^2 = 0.1$) to the average lapse rate for the U.S. Standard Atmosphere up to the tropopause.

From eqs. (7.20) and (8.38), we obtain

$$\alpha_m(h_a) = 10^{-6} N_m(h_a)\sqrt{\frac{2\pi a}{H_o}} \times \left[\sqrt{1 + 2gH_o^2\left(\frac{h_a}{H_o}\right)} + \frac{gH_o^2}{\left(1 + 2gH_o^2\left(\frac{h_a}{H_o}\right)\right)^{3/2}}\right] \quad (8.40)$$

where $N_m(h_a)$ is given by eq. (8.38). Similarly, using the Abel transform, we obtain for the error in the recovered refractivity as a result of a quadratic along-track variation in the true refractivity

$$\delta\hat{N}(h_a) = N_m(h_a)\left\{\beta_0 + \beta_2\left[\frac{H_o}{a}\right]\left[1 - \left(gH_o^2\right)\left(\frac{h_a}{H_o} - 2\right)\right]\right\} + O\left[\left(gH_o^2\right)\left(\frac{h}{H_o}\right)\right]^2 \quad (8.41)$$

and from eq. (8.14) we obtain for the temperature error

$$\delta\hat{T}(h_a) = -\left(gH_o^2\right)T^*(h_a)\left\{\beta_0 + \beta_2\left[\frac{H_o}{a}\right]\right\} + O\left[\left(gH_o^2\right)\left(\frac{h}{H_o}\right)\right]^2 \quad (8.42)$$

Because the value of the (gH_o^2) term is about 0.1, eqs. (8.41) and (8.42) predict about a 90% reduction in the magnitude of the effect of an along-track variation in the true refractivity on the recovered temperature versus its effect on the recovered refractivity. The degree of mutual cancellation of the two principal terms in eq. (8.14) in calculating the temperature error is highly sensitive to the assumptions made in the atmospheric model and in the ray path approximations. Consequently, a prudent choice is to set $\delta\hat{T}$ equal to one of the two principal terms in eq. (8.14) times some dimensionless factor Γ that is probably at most a few tenths in magnitude. In this case $\delta\hat{T}$ for a quadratic variation in along-track refractivity would be given by

$$\delta\hat{T}(h_a) = \Gamma T^*(h_a)\left[\beta_0 + \frac{XaH_o}{100L^2}\right] \quad (8.43)$$

Thus, eq. (8.43) predicts that a 1% quadratic along-track variation in true refractivity over a horizontal distance from the center line of 300 km results in an error of 0.2 K in the recovered temperature for $\Gamma = \left(gH_o^2\right) = 0.1$. This topic is discussed further in the following chapter on simu-

lations. For a quadratic along-track variation in refractivity, these simulations should verify the form of eq. (8.43) and refine the value of Γ for a realistic atmospheric model and a more accurate ray-tracing algorithm.

The Effect of Water Vapor on Temperature Recovery

For a given profile for the recovered refractivity, which is controlled by the occultation data set, it follows from eq. (7.34) that the variations in number density and temperature of the atmospheric constituents are constrained by the condition

$$0 = c\delta n + d\delta n_w / T - d\delta T n_w / T^2 \quad (8.44)$$

which must hold at each observational epoch. Here, n is the total number density of the atmospheric constituents, and n_w is the water vapor number density. Let n_t be the true value of the total number density, and let n_f be the false or forced value of the total number density assuming no water vapor is present. Then, for a given recovered refractivity profile, $\hat{N}$, it follows that

$$cn_t + \frac{d}{T_t} n_w = \hat{N} = cn_f \quad (8.45)$$

where $T_t(h)$ is the true temperature profile. Water vapor creates a more efficient refracting medium ($1 + d/cT_t \approx 17$). The recovered total number density n_f in compensation is forced to a much larger value than the true value to maintain the same observed refractivity. For example, for $n_w/n_t = 0.01$, n_f would be 16% larger than the true value, which would lead to a denser atmosphere whose constituents have a slightly higher mean molecular mass; this would result in significant errors in the recovered temperature and pressure.

From eq. (7.38) it can be shown that the presence of unmodeled water vapor in the amount n_w results in an error in the recovered temperature profile that is given by

$$\frac{\delta\hat{T}}{T} \triangleq \left.\frac{\hat{T}_f - T_t}{T_t}\right|_h = \left(\frac{\int_h^\infty \left(\frac{d}{cT_t} + \nu\right) n_w dh'}{\int_h^\infty (n_t - \nu n_w) dh'} - \left.\frac{d}{cT_t}\left(\frac{n_w}{n_t}\right)\right|_h \right) \div \left[1 + \left.\frac{d}{cT_t}\left(\frac{n_w}{n_t}\right)\right|_h \right] \quad (8.46)$$

where ν is the ratio of the molecular mass difference between dry air and water, and is given by

$$\nu = \frac{\mu_a - \mu_w}{\mu_a} = 0.378 \quad (8.47)$$

The ν terms in eq. (8.46) reflect the slight increase of the mean molecular mass of the forced atmosphere, which has a small effect, compared to the d/cT_t term. Likewise, the error in the recovered pressure is given by

$$\frac{\delta\hat{P}}{P} \triangleq \left.\frac{\hat{P}_f - P_t}{P_t}\right|_h = \frac{\int_h^\infty \left(\frac{d}{cT_t} + \nu\right) n_w dh'}{\int_h^\infty (n_t - \nu n_w) dh'} \quad (8.48)$$

The errors in temperature and pressure recovery given by eqs. (8.46) and (8.48) clearly depend on the vertical profiles of n_w and T_t in the lower troposphere. For illustrative purposes, we use eq. (8.39) for the temperature profile and a profile for n_w that is exponentially distributed with a scale height of H_w. For this case, the error in the recovered temperature at an altitude h using eq. (8.38) as the reference model for $n_t(h)$ is given by

$$\frac{\delta\hat{T}}{T} \doteq -\left[1-\left(\frac{H_w}{\overline{H}}\right)\left[1+2\left(gH_o^2\right)\left(\frac{H_w}{H_o}\right)\right]\right] \times\left[\frac{n_w d/cT_t}{n_t+n_w d/cT_t}\right]+O\left[\left(gH_o^2\right)^2\right] \quad (8.49)$$

and for the recovered pressure

$$\frac{\delta\hat{P}}{P} \doteq \left(\frac{H_w}{\overline{H}}\right)\left[1+2\left(gH_o^2\right)\left(\frac{H_w}{H_o}\right)\right] \times\left[\frac{d}{cT_t}\frac{n_w}{n_t}\right]+O\left[\left(gH_o^2\right)^2\right] \quad (8.50)$$

where n_t, n_w, T_t, and $\overline{H}$ in eqs. (8.49) and (8.50) are evaluated at the altitude h. The ν terms have been neglected in these expressions because of their small value when compared to the d/cT_t term. Here, $\overline{H}(h)$ is the weighted scale height for n_t from eq. (8.38). It is a slowly varying function of height and given by

$$\overline{H}(h)=\frac{\int_h^\infty n_t(h')dh'}{n_t(h)} \quad (8.51)$$

For $H_o = 9$ km and $g = 0.001$, $\overline{H}(h) = 7.2$ km at $h = 0$; $\overline{H}(h) = 6.2$ km at $h = 5$ km. Note that for an unmodeled exponential water vapor distribution amounting to only 10 cm of zenith delay, eq. (8.49) and eq. (8.50) predict surface temperature and pressure errors of about 23 K and 60 mbar, respectively (for $H_w = 2.7$ km and $H_o = 9$ km). Thus, mismodeled water vapor is a very serious problem for temperature recovery in the lower troposphere. However, meterological models that partially constrain the relationship between temperature and the relative abundance of dry air and water vapor can be used to lessen the susceptibility of the recovered temperature to water vapor-induced errors.

The physical interpretation of the results in eqs. (8.49) and (8.50) is as follows. From eq. (8.45), the increasing relative abundance of water vapor as the altitude lowers forces n_t to be increasingly larger than n_t, which would force P_f to be correspondingly higher. But the system is in local hydrostatic equilibrium and, therefore, the pressure cannot increase above that sustainable by the weight of the overlying atmosphere. Consequently, the false temperature in compensation is forced to drop below the true temperature.

The pressure error predicted by eq. (8.50) is so large that synoptic meteorological data, which should be accurate to about 10 mbar in surface pressure, combined with model-based estimates of H_w, should provide constraints that limit water vapor-induced errors in the recovered temperature. For example, it follows from eqs. (8.49) and (8.50) that

$$\frac{P}{T}\left(\frac{\partial T}{\partial P}\right)_w=-\frac{\overline{H}-H_w\left[1+2\left(gH_o^2\right)\left(H_w/H_o\right)\right]}{H_w\left[1+2\left(gH_o^2\right)\left(H_w/H_o\right)\right]} \times\frac{1}{1+\frac{d}{cT_t}\frac{n_w}{n_t}}+O\left[\left(gH_o^2\right)^2\right] \quad (8.52)$$

If ΔP is the difference in pressure inferred from the occultation technique minus that obtained from synoptic, or from *in situ* or other information, then roughly the correction $\Delta\hat{T}$ that should be applied to the recovered temperature to compensate for the error due to mismodeled water vapor is given by

$$\Delta\hat{T} = -\left(\frac{\partial T}{\partial P}\right)_w \Delta P \approx \frac{\overline{H}-H_w}{H_w}\frac{T}{P}\Delta P \quad (8.53)$$

Note that $\left(\frac{\partial T}{\partial P}\right)_w$ is moderately sensitive to the assumptions about the vertical distribution of water vapor as characterized by the value of H_w. Nevertheless, a correction based upon ancillary information that is accurate to 1% in pressure and

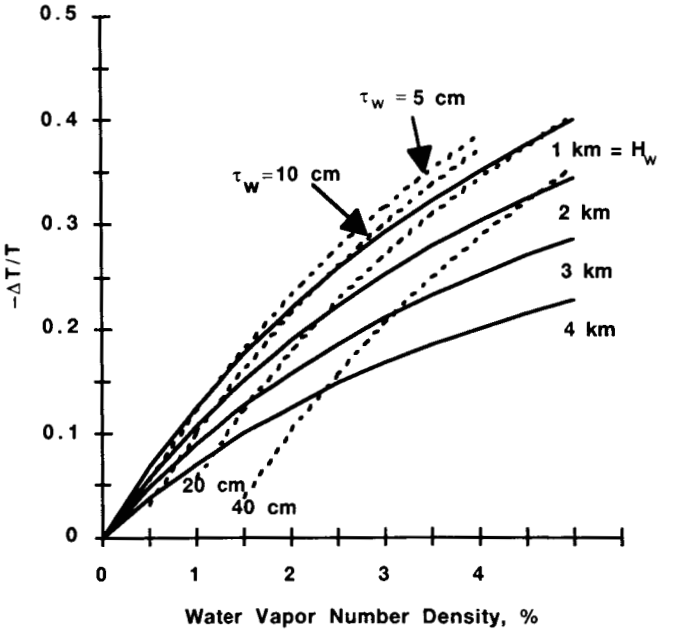


Figure 8-7. Water-vapor-induced error in temperature recovery for different assumed values of the scale height of the water vapor distribution. Contours of constant zenith delay are also shown.

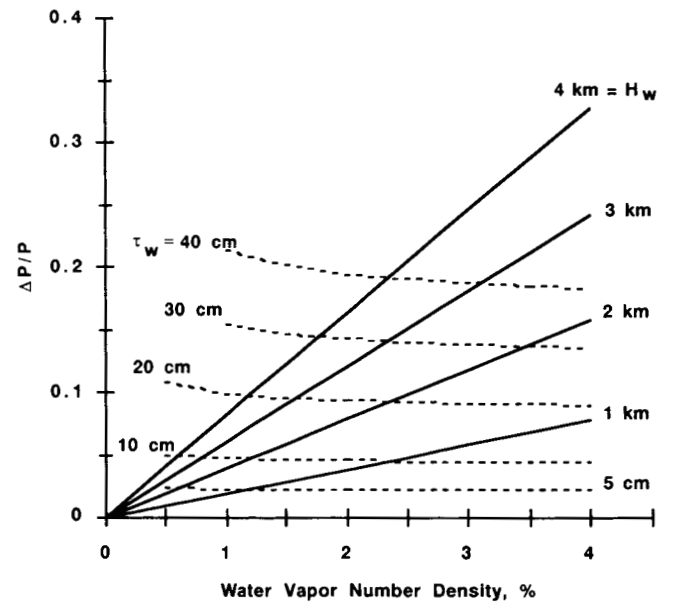


Figure 8-8. Water-vapor-induced error in pressure recovery.

to about 10% in $\left(\frac{\partial T}{\partial P}\right)_w$, might bound the surface temperature error to below 5 K.

Figures 8-7 and 8-8 are plots of the recovery errors predicted by eqs. (8.49) and (8.50) for a range of surface values for n_w/n_t, water vapor

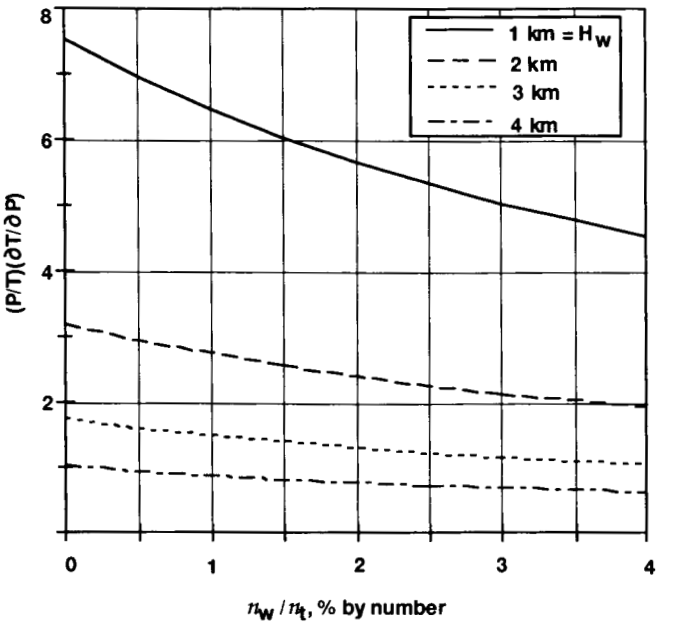


Figure 8-9. Correction curves for adjusting recovered temperature.

scale heights and zenith delays ranging from 40 cm down to 5 cm. It should be noted that this exponential model for the water vapor distribution becomes less physically realizable for larger values of H_w because of saturation limits imposed by the ambient temperature.

Figure 8-9 displays $\left(\frac{\partial T}{\partial P}\right)_w$ based upon eq. (8.52) versus n_w/n_t for a range of scale heights for the vertical distribution of water vapor.

An interesting case is the fully saturated lower troposphere, a condition that is often approached in the tropics. The partial pressure of water vapor essentially varies exponentially with ambient temperature. An empirical relation for n_w for water vapor in fully saturated air in thermodynamic equilibrium with liquid water (solid water yields different equilibrium levels) is given by

$$n_w = \frac{4.436 \times 10^{25}}{T} \exp[17.26(T-273)/(T-35.7)] \text{ m}^{-3} \tag{8.54}$$

and shown in Figure 8-10 (based on information from *Handbook of Chemistry and Physics*, 70th

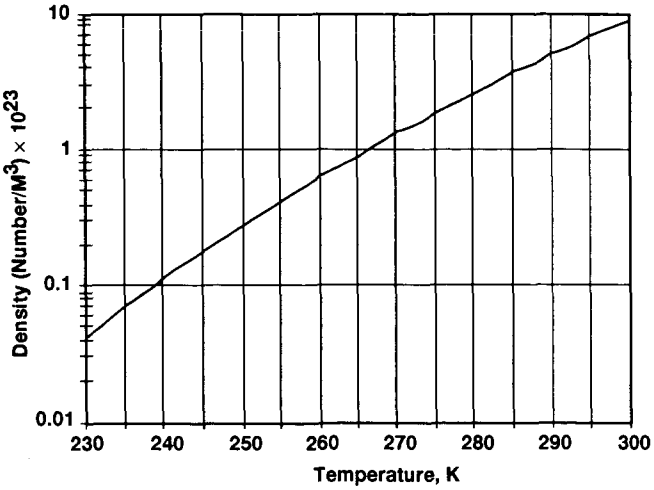


Figure 8-10. Number density of water vapor in fully saturated air that is in thermodynamic equilibrium with liquid water.

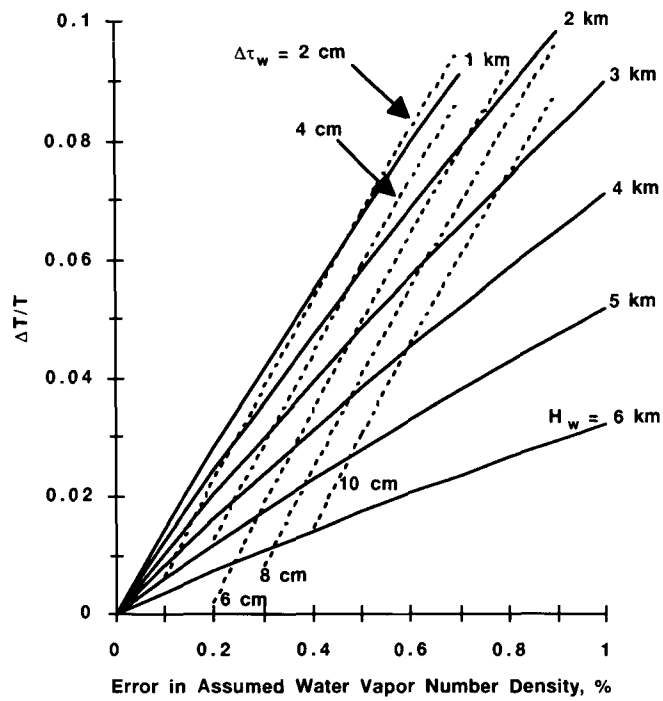


Figure 8-11. Error in recovered temperature as a result of an error in the assumed surface density for water vapor and/or zenith delay is shown for different assumed scale heights for its vertical distribution.

Edition; see also Tetens 1930). For a given vertical profile of temperature, this relationship can be integrated to obtain the total zenith delay due to saturated water vapor. For a lapse rate of 6 K/km, we obtain from eq. (8.54) a mean scale height for the water vapor density of 2.3 km. For a temperature at sea level of 300 K, we obtain a total zenith delay of about 40 cm, which is about the maximum observed in the tropics. For this total water vapor zenith delay and scale height, the resulting errors in the recovered temperature and pressure are 85 K and 180 mbar, respectively!

The error relations in eqs. (8.46) and (8.48) use a zero reference value for n_w. Because the ratio n_w/n_t is at most a few hundredths, these expressions are also valid for determining the error in the recovered temperature and pressure as a result of an error δn_w relative to a non-zero reference value for n_w. We simply replace n_w with $-\delta n_w$ in eqs. (8.46) and (8.48).

Figure 8-11 shows the error in the recovered temperature as a result of mismodeled water vapor; it provides the same information given in Figure 8-7 but in more detail. Figure 8-11 also shows the high sensitivity of the error in recovered temperature—for a given error in the total zenith delay due to water vapor $\Delta\tau_w$—to the vertical distribution of that water vapor, as characterized by the scale height H_w. On the other hand, the error in recovered pressure, shown in Figure 8-12, for an assumed zenith delay due to water vapor is largely insensitive to its vertical distribution; pressure is more constrained by the bulk properties of the atmosphere at large.

Errors in Water Vapor Recovery

Conversely, we can also evaluate the error in the recovered water vapor as a result of errors in the assumed pressure and temperature profiles. Using eq. (7.38) it can be shown that the error $\delta\hat{n}_w$ in the recovered water vapor number density is given by

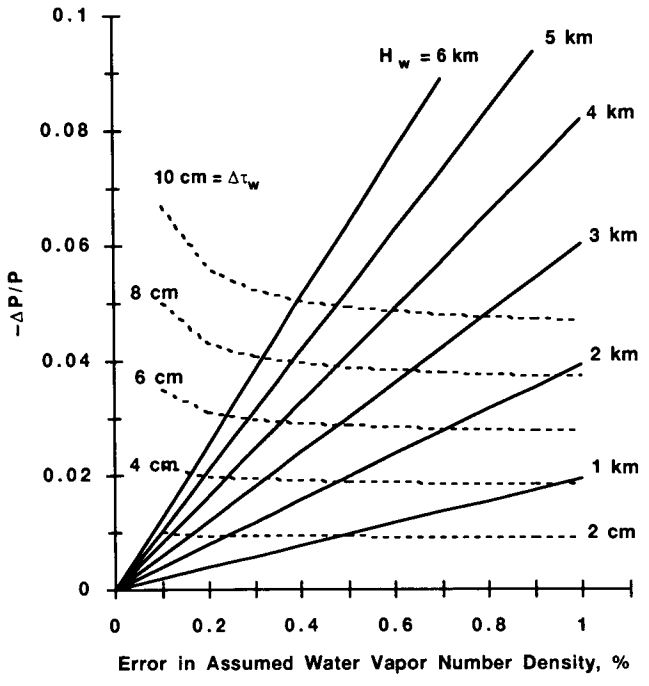


Figure 8-12. Error in recovered pressure as a result of an error in the assumed surface density of water vapor and/or zenith delay is shown for different assumed scale heights for the vertical distribution of water vapor.

$$\delta\hat{n}_w = \frac{\hat{N}}{d}\delta T - \frac{c}{kd}\delta P \tag{8.55}$$

where δT and δP are the errors in the *assumed* temperature and pressure profiles and where $\hat{N}$ is the recovered refractivity. It follows from eqs. (7.34) and (8.55) that

$$\frac{T}{n_w}\left(\frac{\partial\hat{n}_w}{\partial T}\right)_P = \frac{n_t + n_w d/cT_t}{n_w d/cT_t} \tag{8.56}$$

and

$$\frac{P}{n_w}\left(\frac{\partial\hat{n}_w}{\partial P}\right)_T = -\frac{n_t}{n_w}\frac{cT_t}{d} \tag{8.57}$$

To evaluate these expressions, we would assume reference profiles for n_w, n_t, and T_t. These sensitivity partials may also be integrated in height to yield the sensitivity of the estimate of the total or columnar content of water vapor to errors in the temperature and pressure profiles—an important quantity in meteorology for the study of atmospheric thermal energy transport and for weather prediction. Let the columnar content of the water vapor be defined by $\mathcal{W}$, which is given by

$$\mathcal{W} = \int_0^\infty n_w dh \tag{8.58}$$

Suppose that the error in the temperature profile $\delta T(h)$ is a constant. Then, the error in the estimate $\hat{\mathcal{W}}$ due to a temperature error, while holding P fixed, is given by

$$\left(\frac{\delta\hat{\mathcal{W}}}{\mathcal{W}}\right)_P \stackrel{\Delta}{=} \frac{1}{\mathcal{W}}\left(\frac{\partial\hat{\mathcal{W}}}{\partial T}\right)_P \delta T \doteq \frac{\delta T}{T_t}\left[\frac{\tau_a(1+d/cT_t)}{\tau_w\, d/cT_t} + 1\right] \approx \frac{\delta T}{T_t}\left[\frac{221}{\tau_w} + 1\right] \text{ (at sea level)} \tag{8.59}$$

where τ_a and τ_w are the zenith delays (in centimeters) of dry air and water vapor, respectively.

Similarly, for an error in surface pressure, the error in $\hat{\mathcal{W}}$ holding T fixed is given by

$$\left(\frac{\delta\hat{\mathcal{W}}}{\mathcal{W}}\right)_T \stackrel{\Delta}{=} \frac{1}{\mathcal{W}}\left(\frac{\partial\hat{\mathcal{W}}}{\partial P}\right)_T \delta P(0) \doteq -\frac{\delta P(0)}{P(0)}\left(\frac{\tau_a}{\tau_w}\right)\left[1 + cT_t/d\right] \approx -\frac{\delta P(0)}{P(0)}\left[\frac{221}{\tau_w}\right] \tag{8.60}$$

Figure 8-13 displays eqs. (8.59) and (8.60) for a 1-K constant temperature error and –10-mbar surface pressure error. Over tropical oceanic regions that approach fully saturated conditions, it follows that a 1-K temperature error results in a 2% error in the total water vapor content and that a 10-mbar surface pressure error results in a 5% error; roughly speaking, a 1% error in the as-

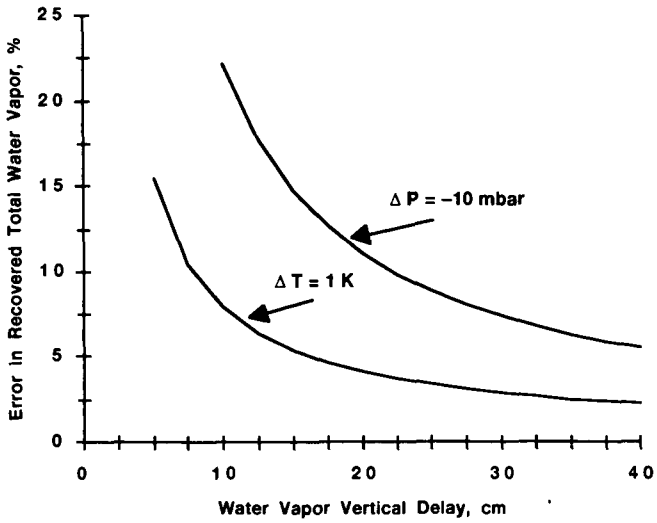


Figure 8-13. Relative error in total water vapor recovery due to errors in the assumed temperature profile of 1 K and -10 mbar in the surface pressure.

sumed quantities results in about a 5% error in the recovered total water vapor content, or about 2 cm of equivalent zenith delay. Over temperate regions, a 1% error in the temperature or pressure profiles would still result in about a 2-cm error in equivalent zenith delay, but the total delay in this case is typically only 10–20 cm.

The Effect of Ionospheric Mismodeling on Temperature Recovery

The Earth's ionosphere, which extends from about an 80-km altitude upward, acts as a lens that overlays the neutral atmosphere. In an occultation geometry, the signal must pass through the ionosphere on its way into and out of the neutral atmosphere below a 100-km altitude. This implies that a removal of this ionospheric effect on the signal is necessary before we can accurately deduce the neutral atmospheric profile. The dispersive nature of the ionosphere causes the two GPS signals to travel at different speeds. A simple linear combination of the L1 and L2 signals can be formed to subtract out most of the ionospheric delay. This commonly used simple linear combination, however, assumes that the two GPS signals are traveling along exactly the same paths. Moreover, it also ignores higher order terms in the expansion of the ionospheric index of refraction. These residual ionospheric effects, if left uncalibrated, act as an error source that maps into neutral atmospheric profile errors. In this section we will examine these ionospheric residuals in some detail.

Starting with the Appleton–Hartree formula for the ionospheric index of refraction (Papas 1965) and expanding in powers of the inverse carrier frequency, the total atmospheric index of refraction can be written as (see e.g., Bassiri and Hajj 1993)

$$n = 1 - \frac{1}{2}X \pm \frac{1}{2}XY\cos\theta_B - \frac{1}{4}X\left[\frac{1}{2}X + Y^2\left(1+\cos^2\theta_B\right)\right] + 10^{-6}N_{neut} \quad (8.61)$$

where

$$X = \left(\frac{f_p}{f}\right)^2 = \frac{\left(n_e e^2/4\pi^2\varepsilon_0 m\right)}{f^2} \quad (8.62)$$

$$Y_T = Y \sin\theta_B; \quad Y_L = Y\cos\theta_B \quad (8.63)$$

$$Y = \left(\frac{f_g}{f}\right) = \frac{\left(|e|B_0/2\pi m\right)}{f} \quad (8.64)$$

where n_e is the number density of electrons; e and m are the electron charge and mass, respectively; ε_0 is the permittivity of the free space; f_p, f_g, and f are the plasma-, gyro-, and carrier frequencies, respectively; θ_B is the angle between the Earth's magnetic field $\mathbf{B_0}$ and the direction of propagation of the wavefront $\mathbf{k}$. B_0 is the magnitude of $\mathbf{B_0}$; N_{neut} is the refractivity of the neutral atmosphere. By definition $\mathbf{Y} = e\mathbf{B_0}/2\pi fm$, and because the sign of e is negative for the electron, $\mathbf{Y}$ is antiparallel to $\mathbf{B_0}$. For the Earth's ionosphere,

with $n_e = 10^{12}$ electrons/m^3 the plasma-frequency $f_p \approx 8.9$ MHz. The gyro-frequency for an electron in the Earth's magnetic field (2×10^{-5} Tesla) is $f_g \approx 0.59$ MHz.

The second, third, and fourth terms on the right-hand side (RHS) of eq. (8.61) are proportional to the inverse-square, inverse-cube, and inverse-quartic powers of frequency, respectively. The plus and minus signs of eq. (8.61) correspond to the ordinary and extraordinary wave modes of propagation, respectively. When the carrier frequency is large compared to plasma- and gyro-frequencies, as is the case with the GPS frequencies, the principal modes of propagation are dominantly circularly polarized. Ignoring the left circular polarized (LCP) component of the GPS signal, which has less than 0.35% and 2.5% of the total power for L1 and L2, respectively, only the (–) sign will be of relevance to us in the subsequent analysis.

The phase delay of a signal passing through the atmosphere is given by

$$\rho_i = \int_{\text{curve } i} n_i \, ds_i \qquad (8.65)$$

where i = 1, 2 for L1 and L2, respectively; ds_i is an element of length along the curved path that the ith signal travels (see Figure 8-14). For the case where (n–1) << 1 and where the gradient of the index of refraction is slowly varying along the path, we can use the stationary phase property of the actual ray path (see Appendix B, eq. (B-19); also Moritz 1961; Williams 1975; and Gu and Brunner 1990) to rewrite eq. (8.65) in a more convenient form. For this case it can be written with sufficient accuracy as

$$\rho_i = \int_0^{R_{LG}} n_i dx - \frac{1}{2}\int_0^{R_{LG}} \frac{1}{x^2}\left[\left(\int_0^x \frac{\partial n_i}{\partial z} x' dx'\right)^2 + \left(\int_0^x \frac{\partial n_i}{\partial y} x' dx'\right)^2\right] dx \qquad (8.66)$$

where dx and dx' are elements along the straight line connecting the transmitter and the receiver (Figure 8-14). All integrations in eq. (8.66) are carried out along this line. The coordinates y and z form a mutually orthogonal coordinate set with x. The partial derivatives $\partial n_i/\partial y$ and $\partial n_i/\partial z$ are evaluated at (y,z) = 0. (For more accuracy, they should be evaluated at (y,z) of the curved path at a given x.) The first integral on the RHS of eq. (8.66) corresponds to the effect of the refracting medium on the signal had it traveled along a straight line. The second integral corresponds to a curvature correction or bending term. It should be noted from eq. (8.66) that the bending term is a second-order term because of the stationarity property of the actual path, which is proportional to the square of the transverse gradient of the index of refraction. From eq. (8.61), it follows that the f^{-2} term in the refractivity brings an f^{-4} bending term to the phase observables.

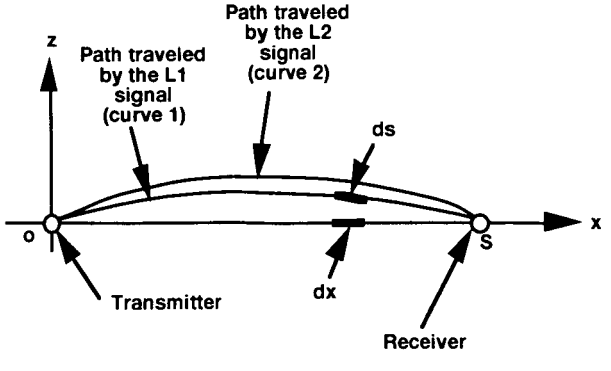


Figure 8-14. Representation of the path traveled by the GPS L1 and L2 signals.

Substituting eq. (8.61) into eq. (8.66), we can write the phase delays, ρ_1 and ρ_2, up to order f^{-4} as

$$\rho_1 = R_{LG} + \rho_{neut} - \frac{q-\varepsilon}{f_1^2} - \frac{1}{2}\frac{v}{f_1^3} - \frac{1}{3}\frac{w}{f_1^4} - \frac{\beta}{f_1^4} + B_1 \qquad (8.67)$$

$$\rho_2 = R_{LG} + \rho_{neut} + \delta\rho_{neut} - \frac{q-\varepsilon}{f_2^2} - \frac{1}{2}\frac{v}{f_2^3} - \frac{1}{3}\frac{w}{f_2^4} - \frac{\beta}{f_2^4} + B_2 \tag{8.68}$$

where

$$q = \frac{1}{2}\int_0^{R_{LG}} f_p^2 dx = (40.3)\int_0^{R_{LG}} n_e dx = 40.3\ TEC \tag{8.69}$$

$$v = \int_0^{R_{LG}} f_g f_p^2 \cos\theta_B\, dx = 7527\ c\int_0^{R_{LG}} n_e B_o \cos\theta_B\, dx \tag{8.70}$$

$$w = (2437)\int_0^{R_{LG}} n_e^2 dx + 4.74\times 10^{22} \int_0^{R_{LG}} n_e B_o^2 \left(1+\cos^2\theta_B\right) dx \tag{8.71}$$

$$\varepsilon = 40.3\times 10^{-6}\int_0^{R_{LG}} \frac{dx}{x^2}\left[\left(\int_o^x dx'x'\frac{\partial n_e}{\partial z}\right)\left(\int_o^x dx'x'\frac{\partial N_{neut}}{\partial z}\right)+\left(\int_o^x dx'x'\frac{\partial n_e}{\partial y}\right)\left(\int_o^x dx'x'\frac{\partial N_{neut}}{\partial y}\right)\right] \tag{8.72}$$

$$\beta = \frac{(40.3)^2}{2}\int_0^{R_{LG}} \frac{dx}{x^2}\left[\left(\int_o^x dx'x'\frac{\partial n_e}{\partial z}\right)^2 + \left(\int_o^x dx'x'\frac{\partial n_e}{\partial y}\right)^2\right] \tag{8.73}$$

All integrations in eqs. (8.69) through (8.73) are along the straight line connecting the transmitter and the receiver; units are assumed to be in meter kilogram second (MKS). The quantity R_{LG} is the direct geometrical distance between the transmitter and the receiver; ρ_{neut} is the extra delay (including bending) along the L1 path due to the neutral atmosphere; $\delta\rho_{neut}$ is the difference between the ρ_1 and ρ_2 extra phase delays induced by the neutral atmosphere; f_i is the carrier frequency; TEC is the total electron columnar content along the path of integration; and ε and β are curvature correction terms. B_i is the sum of other biases and error sources such as thermal noise, multipath, and clock drift; these are not the subject of our study in this section and so these terms will be dropped in the following analysis.

The term $\delta\rho_{neut}$ is the result of the splitting of the two signals that is caused by the ionosphere; this splitting causes the two signals to travel different paths in the neutral atmosphere (Figure 8-15). The term ε completely cancels out as we form the L1 and L2 traditional "ionospheric free" combination as explained below. The term β/f_i^4 corresponds to the extra delay which is caused by bending in the ionosphere.

The following linear combination removes the $1/f^2$ ionospheric terms

$$\left[\frac{f_1^2}{f_1^2 - f_2^2}\right]\rho_1 - \left[\frac{f_2^2}{f_1^2 - f_2^2}\right]\rho_2 = R_{LG} + \rho_{neut} - 1.546\ \delta\rho_{neut} + \frac{1}{2}\frac{v}{f_1 f_2 (f_2 + f_1)} + \frac{1}{3}\frac{w}{f_1^2 f_2^2} + \frac{\beta}{f_1^2 f_2^2} \quad (8.74)$$

The coefficients multiplying ρ_1 and ρ_2 correspond to 2.546... and 1.546..., respectively. This linear combination removes the $1/f^2$ ionospheric term; R_{LG} is estimated from the POD-based ephemerides of the LEO and GPS satellites. The desired observable is ρ_{neut} from which the neutral refractivity can be obtained. The third, fourth, fifth, and sixth terms on the RHS of eq. (8.74) will act as error sources to the inversion process; in what follows they will be referred to as "splitting," "second-order," "third-order," and "bending" terms, respectively. Below we estimate these residuals and suggest ways of reducing them.

Simulation of the Ionosphere

To proceed with the computation of the higher order and bending terms, we have to assume models for the electron density n_e and the Earth's magnetic field B_0. (See Appendix D for a treatment of the higher order ionospheric effects using a uniform spherical shell model.) For the daytime electron density, we generated a two-dimensional grid along 20° east geodetic longitude using the parameterized ionospheric model (PIM) (Figure 8-16) (Anderson 1993). This two-dimensional grid corresponds to the simulated electron density at 12:00 Universal Time; September 26, 1992. The density is representative of the ionosphere during the day time and for a year near solar maximum. The PIM is a worldwide electron density profile model that is based on a parameterized version of the Utah State University model, which is a complete first principles, ionospheric model using as inputs the solar ultraviolet (UV) flux wavelength profile, the Hedin neutral wind model, the MSIS-86 neutral atmosphere model, and an electric field model obtained from all available experimental data. The two-dimensional density grid has a bulge near 20° north latitude. A vertical profile at that latitude is shown in Figure 8-17. The two peaks correspond to the F2 peak at a 400-km altitude and the E peak at an ~100-km altitude. In our simulation, the occultation is taking place in the same 20° longitude plane, with a tangent point at 20° north latitude.

For the nighttime electron density, a Chapman model is used with an F-peak density of

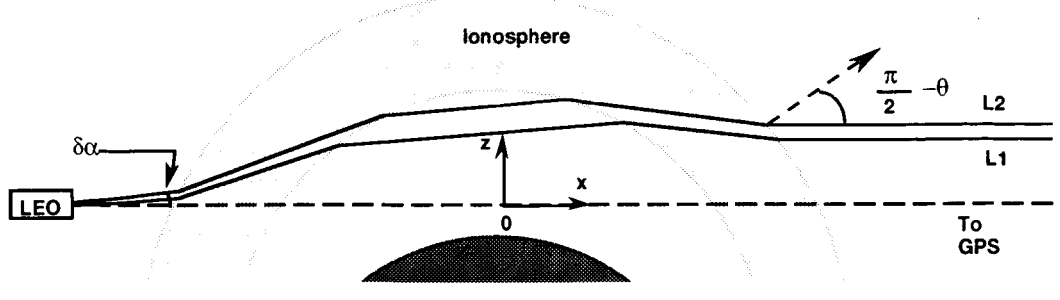


Figure 8-15. L1 and L2 ray splitting due to dispersive ionospheric refraction using a simple spherically symmetric shell model. The emitting GPS satellite is assumed to be at an infinite distance. Differential bending through the atmosphere by the L1 and L2 signals is suppressed for clarity. Spherical geometry results in non-parallel paths shown in the atmosphere. The z = 0 line corresponds to the ionospheric-free stationary phase path for L1 and for L2.

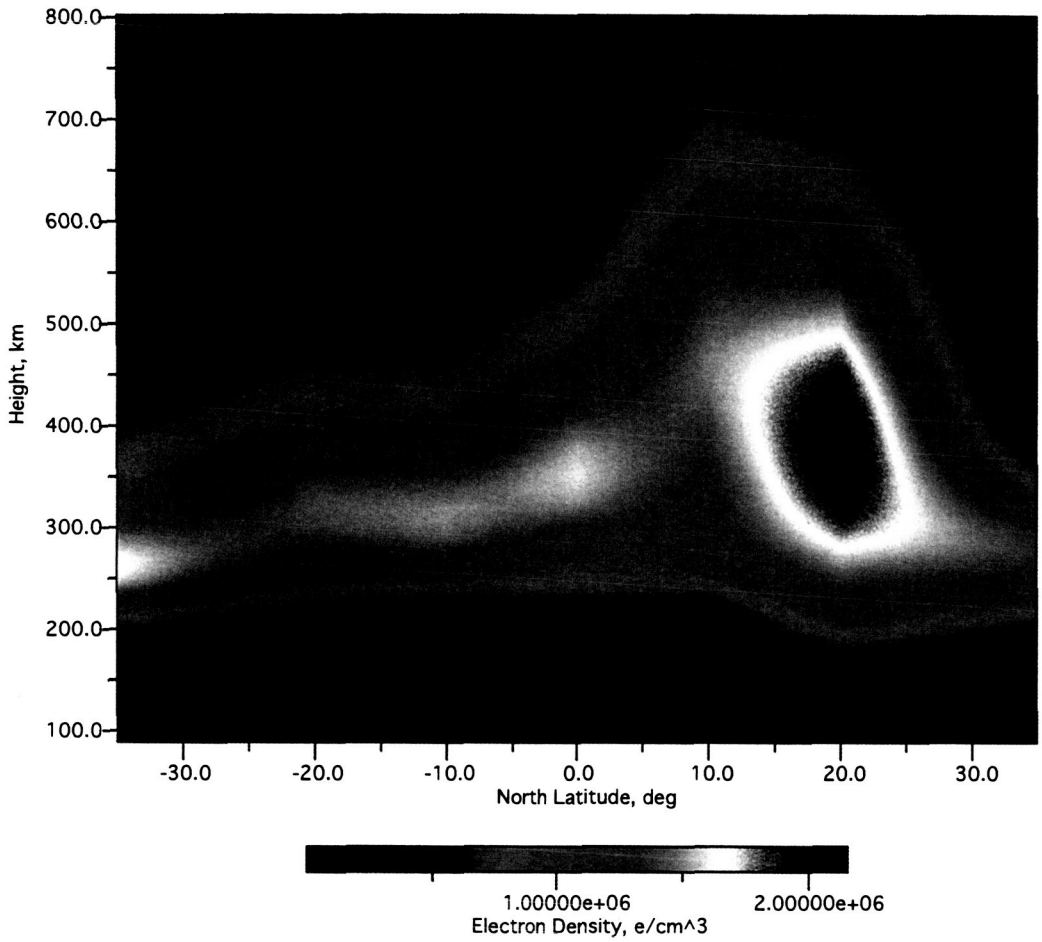


Figure 8-16. PIM electron density model (electron/cm^3)used to estimate higher order delay and bending terms.

10^5 (e/cm^3) at 350 km altitude and an E-peak density of 7×10^3 (e/cm^3) at 100 km altitude.

Next we need to model the Earth's magnetic field. A first-order approximation to the geomagnetic field near the surface of the Earth that we will use is an Earth-centered dipole with its axis tilted to intersect the Earth at 78.5° north latitude, 291.0° east longitude, which corresponds to the geomagnetic north pole; and at 78.5° south latitude, 111.0° east longitude, which corresponds to the geomagnetic south pole (Figure 8-18). In general, such an approximation is only accurate to about 75%; an approximation that is 90% accurate can be obtained by taking the dipole to be eccentric. An even more accurate value of the magnetic field can be obtained by using the International Geomagnetic Reference Fields; however, our aim is to get a first-order assessment of the higher order and bending terms. These terms are small, but can be limiting sources of error in the temperature profile retrieval in the upper stratospheric heights. The complexity of the models used to estimate these errors need only be consistent with the required accuracies.

Based on these models, we estimate the residual terms of eq. (8.74). Figure 8-19 shows the propagation delay residuals due to the second- and third-order terms as functions of the occultation-link tangent-point height. The daytime second-order term is ~13–17 mm for our specific example. More generally, this term may vary between 0 and 20 mm for a given occultation. It is

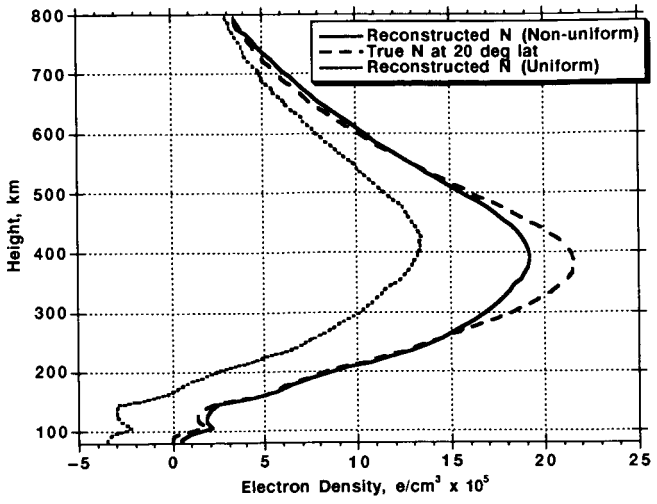


Figure 8-17. True and reconstructed electron density profiles at 20 °north latitude.

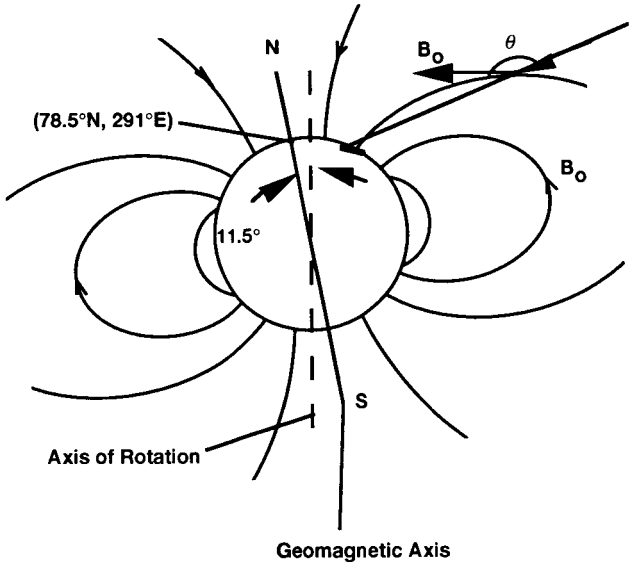


Figure 8-18. The Earth's magnetic field modeled as an Earth-centered dipole aligned along the geomagnetic axis.

possible to further reduce this error by a factor of 10 through modeling (Brunner and Gu 1991; Bassiri and Hajj 1993). The nighttime second-order term is of order 0–3 mm. The third-order term is sub-millimeter and can be ignored.

The daytime bending term, shown in Figure 8-20, ranges between 3–17 cm at a 40–100 km altitude. This term becomes particularly large at

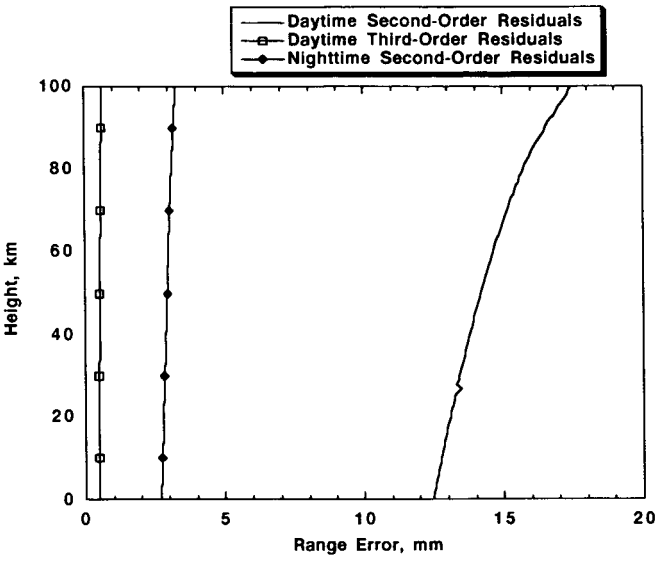


Figure 8-19. Modeled second-order ionospheric residuals for daytime and nighttime ionosphere and third-order daytime ionospheric residuals as functions of the occultation-link tangent height.

a 100-km altitude, which corresponds to the relatively sharp vertical refractivity gradient of the bottom side of the E layer. The effect of the ionospheric bending is understood further as we examine the amount of bending experienced by each signal as shown in Figure 8-20 (right-hand scale). The L1 and L2 signals experience a maximum bending of 0.015 and 0.025 deg respectively, near the E layer. The positive bending indicates that the signal paths curve toward the Earth, namely in the direction of increasing ionospheric refractivity. The nighttime bending term is of the order of 1 mm and can be neglected. Bending due to the neutral atmosphere becomes dominant below 40 km.

Because the bending of the L1 and L2 signal paths within the ionosphere differs, we can estimate the bending term as follows. Treating the L1 and L2 data separately, and assuming spherical symmetry, we can determine the angles of arrival γ_1, δ_1 for L1 and γ_2, δ_2 for L2 and the corresponding asymptote geometric ray length g_1, d_1 and g_2, d_2 (see Figure 8-21 for a depiction of these param-

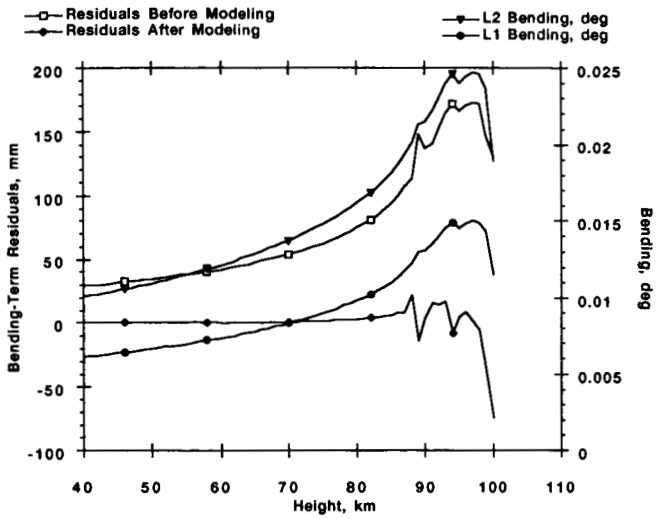


Figure 8-20. Modeled bending of L1 and L2 signals (right-hand scale) for the ionosphere of Figure 8-16. Left-hand scale shows the residual ionospheric bending term before and after calibration.

eters). The bending term is then estimated to first order by

$$\text{Estimated Bending Term} = \tag{8.75}$$
$$2.546\left[\tfrac{1}{2}\left(g_1\gamma_1^2 + d_1\delta_1^2\right)\right] - 1.546\left[\tfrac{1}{2}\left(g_2\gamma_2^2 + d_2\delta_2^2\right)\right]$$

Using the ionospheric model described above, we evaluate eq. (8.75) and subtract it from the actual bending term (shown in Figure 8-20). The difference is also shown in Figure 8-20. The bending-term error is now reduced to the millimeter level below a 70-km altitude.

We now turn our attention to the splitting term ($-1.546\,\delta\rho_{neut}$ in eq. (8.74)). This term is the result of the separation of the L1 and L2 signals as they travel through the neutral atmosphere. Figure 8-22 shows the vertical distance separating the tangent heights of the L1 and L2 ray paths (left-hand scale). This ranges from roughly 20 m near the surface of the Earth to about 500 m near a 100-km altitude for the daytime; it is about an order of magnitude smaller for the nighttime. The right-hand scale of Figure 8-22 shows what this separation would correspond to in terms of

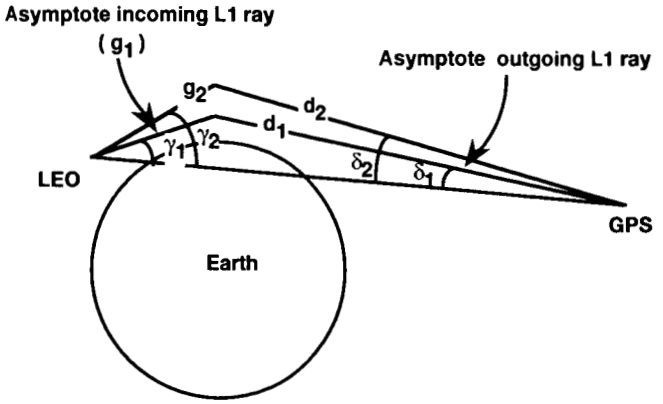


Figure 8-21. Asymptote rays and bending angles for the L1 and L2 signals.

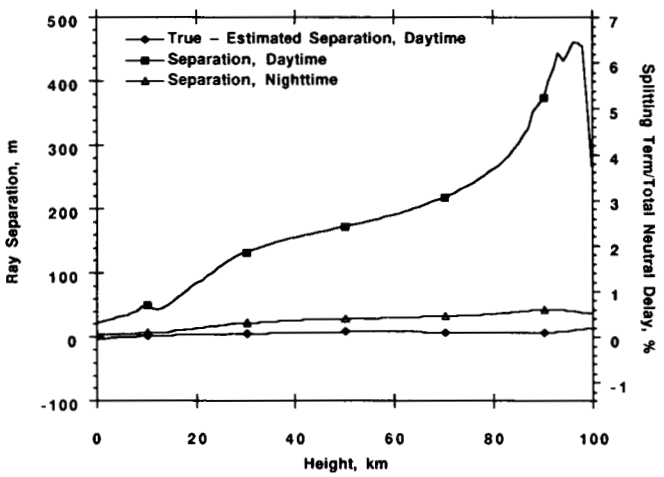


Figure 8-22. Modeled daytime and nighttime ray separation (left-hand scale) and the corresponding ratio of residual error due to the splitting of signals over the total atmospheric effect (right-hand scale) as a function of tangent height. Calibrated daytime effects are also shown.

percent error. This percent error is derived by assuming the neutral refractivity to be an exponentially decaying function of height with a constant scale height of 7 km. It is clear from this example that this error would be a substantial fraction of the total effect at all heights (0.5–11%) for the daytime. During nighttime, this splitting term is again about a factor of 10 smaller. As an approximation, a 1% error in refractivity would map into a 2–3 K error in temperature. Therefore, in order to obtain sub-kelvin temperature accu-

racy, we must model this term in order to remove it from the observables. This can be done by obtaining a first-order estimate of the neutral refractivity profile, and then estimating the ray separation according to the following equation:

$$\text{Estimated Rays' Separation} = g_2\gamma_2 - g_1\gamma_1 \quad (8.76)$$

where the variables on the RHS are defined in Figure 8-21. We can then use the estimated separation and the first-order solution of the neutral atmospheric refractivity to estimate and remove the $\delta\rho_{neut}$ term. A refined solution of the neutral atmospheric refractivity can be obtained by iterating the procedure a few times until convergence. Figure 8-22 shows the difference between the estimated ray separation and the true one for the daytime example. The 0.5–11% error is reduced to a ~0–0.2% error, which will map to sub-kelvin-level temperature errors.

Equations (8.75) and (8.76) assume local spherical symmetry within the ionosphere. These can be further improved if additional information about the ionosphere is included, thereby allowing departure from this assumption. To see how this can be done, an example of the solution obtained under the assumption of local spherical homogeneity is shown in Figure 8-17 (labeled "Uniform"). A better reconstruction is obtained when a horizontal gradient in the electron density was assumed (labeled "Non-uniform" in Figure 8-17). This horizontal gradient is obtained from ground zenith TEC measurements and distributed equally across all the ionospheric layers (Hajj et al. 1994b). It is important to notice that the point-to-point structure of the E layer is recovered in the profile, but the absolute level is off. This can be improved by applying the constraint that the electron density must fall off to zero near 90 km, and not exceed a certain limit within the E region. Given the more accurate electron density profile and the horizontal gradient, more accurate estimates of the ray separation (and therefore the splitting term), the second-order term, and the bending terms are possible.

CHAPTER

9 SIMULATIONS[1]

Based on order-of-magnitude estimates, the atmosphere can be divided into several altitude and climatic regions where different error sources become significant. (See Figure 2-2.) Below 25 km, where random measurement error is a small fraction of the total refractivity, systematic errors in the assumed shape of the contours of constant refractivity are expected to dominate. Refractivity, temperature, and pressure errors will grow gradually with increasing altitude until the decrease in refractivity in the thinning atmosphere causes random errors and residual ionospheric effects to dominate. Short-term phase instability in the clocks used in the measurements, thermal noise in the receivers, multipath off the surrounding transmitters and receivers, ephemeris errors, and uncertain tropospheric delays are examples of random measurement errors. The dominant error sources also depend on the strategy that is being used during an occultation.

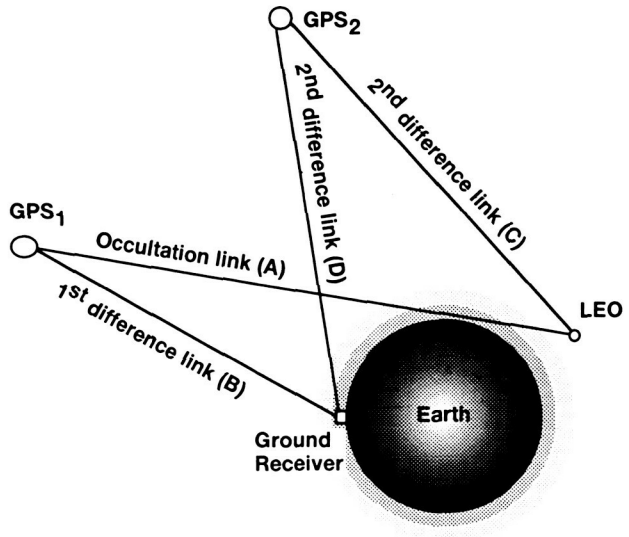


Figure 9-1. Single- and double-differencing geometries.

Three different approaches are possible when analyzing the occultation data. First is the undifferenced technique where measurements from link A, which is the link connecting GPS_1 to LEO shown in Figure 9-1, is used alone. The second approach is the single-differencing technique. In this technique the carrier phase of link A is subtracted from that of B (the link between GPS_1 and a ground antenna), thereby differencing out the GPS_1 clock error.[2] The third approach is the double-differencing technique, where the following linear combination is formed:

(link A – link B) – (link C – link D)

[1] Certain material in this chapter appeared in a more complete form in Hardy et al. 1993; Kursinski et al. 1993a; and Kursinski et al. 1993b.

[2] The analysis in this section assumes the LEO to have an oscillator that is 30 times more stable than the GPS oscillator. When the LEO oscillator is less stable than that of the GPS, such as the case with GPS/MET, it is (link A – link C) that should be used in the single-differencing technique. This difference will change some of the discussion that follows.

Link C is between GPS_2 and LEO; link D is between GPS_2 and the same ground antenna, as shown in Figure 9-1. The double-differencing technique cancels out all clocks involved, namely, those of GPS_1, GPS_2, LEO, and the ground antenna. However, the single- and double-differencing techniques introduce new errors caused by atmospheric effects induced on the signals between the GPS satellite(s) and the ground receiver. With six GPS ground stations distributed evenly around the globe (such as the ones used for orbit determination for TOPEX/POSEIDON), there is sufficient coverage to do double differencing for about 85% of the occultations.

Table 9-1 shows the different sources of error for the three different techniques. The "x" refers to the presence of the noise source listed on the left for the particular technique. The capital and bold **"X"** means the error is dominant. For instance, the GPS cesium oscillator is a dominant error source in the undifferenced technique, whereas it is totally absent in the single- and double-differencing techniques. On the other hand, a LEO quartz oscillator is about 30 times more stable than the GPS cesium and has an effect in the no-differencing as well as the single-differencing technique. In the double-differencing technique, this error is eliminated along with the ground hydrogen-maser clock, assuming that data are taken at the same rate from both the LEO and ground receivers. Also indicated is the type of the errors assumed and, when relevant, the time constants that are associated with these errors. The factor of (2) next to an "x" refers to the noise that appears twice, since it is a double-differencing technique.

Figure 9-2 shows the typical noise levels due to different sources. Within the 200 s shown, the largest error is due to the LEO quartz oscillator as seen from a ground receiver; therefore, it includes the inexact specification of the troposphere. The

Table 9-1. Possible error sources for three different techniques used to analyze occultation data.

Noise Origin	τ (s)	Type	Undifferenced	Single-Differencing	Double-Differencing
Oscillators:					
GPS cesium	1	FF	**X**		
LEO quartz	1	FF	x	x	
Ground H-maser	1	FP		x	
SNR:					
GPS to LEO	0.3	WP	x	x	x(2)
GPS to Ground	0.3	WP		x	x(2)
Troposphere:					
GPS to Ground	1	FF		x	x(2)
Multipath:					
GPS to LEO		SIM	**X**	**X**	**X**(2)
GPS to Ground				x	x(2)
Ionosphere:					
GPS to LEO			See Chapter 8	x	x
GPS to Ground				Appears in SNR	Appears in SNR
GPS_2 to LEO					Appears in SNR

FF = Flicker Frequency FP = Flicker Phase WP = White Phase SIM = Simulation

stability of the LEO oscillator is assumed to be 1 part in 10^{13}. The following error sources are also shown: (1) a hydrogen maser; (2) the LEO SNR after taking the linear combination of L1 and L2, whereby most of the ionospheric effect is subtracted out, but this linear combination increases the phase jitter by a factor of 3; (3) the ground SNR; (4) a 1-mm sine curve with a period of 300 s, which represents the ground multipath; (5) a simulated error due to having the multipath off the LEO body; and (6) a velocity bias that is due to uncertainties in the LEO and GPS velocities and is taken to be 0.05 mm/s. The error sources shown in Figure 9-2 do not include the residual ionospheric errors that have been discussed previously in Chapter 8. Also not shown in the Figure is the GPS cesium clock drift. A typical GPS clock error would look like that of the LEO quartz oscillator shown in Figure 9-2, but would be a factor of 30 greater. Therefore, when an occultation link is used without differencing with any other link (no-differencing technique), then the GPS clock oscillator is by far the most dominant error source.

In order to quantify the effect of these error sources on the recovery of the refractivity, pressure, and temperature, the simulation procedure outlined in Figure 9-3 was developed. The first step is to generate atmospheric temperature and pressure profiles, which are obtained from the

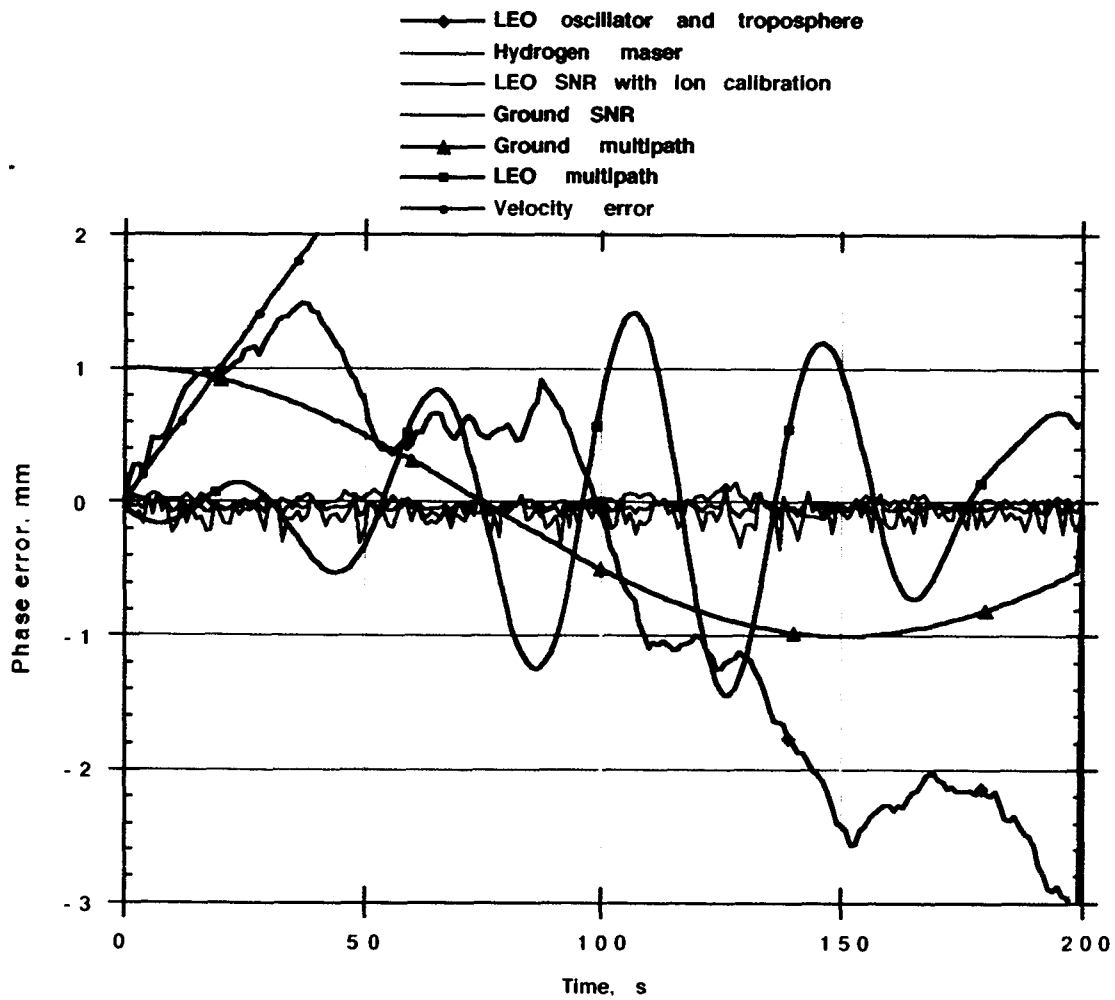


Figure 9-2. The estimated noise levels for seven different error sources.

U.S. Standard Atmosphere (Champion et al. 1985). Ignoring water vapor, refractivity is then generated using the first term of eq. (7.36).

Observables for a given GPS–LEO occultation geometry are generated by tracing the ray signal through the assumed atmosphere. Noise is added to the simulated radio signatures and converted to Doppler shift measurements, which in turn are inverted using the Abel transform approach to obtain the refractivity versus height. Density, pressure, and temperature are obtained from the refractivity profile through the procedures described by eqs. (7.32) and (7.39). The resultant temperature solution is then subtracted from the initial temperature profile used in the simulation in order to obtain an estimate of the error.

Figure 9-4 shows the error in the reconstructed temperature versus height for the single-differencing technique where the GPS clock is eliminated. Two different cases are plotted in Figure 9-4; in one case the inversion started at 69 km, whereas in the other it started at 59 km. Each case corresponds to the RMS of four runs where the data noise shown in Figure 9-2 is divided into four parts: (1) 0–50 s, (2) 50–100 s, (3) 100–150 s, and (4) 150–200 s. Each part is run independently.

Figure 9-4 shows that the temperature error is ~1 K at 50 km and drops to a few tenths of a kelvin in the middle and lower stratosphere. Although several error sources are included in the simulations, possible significant errors due to the ionosphere as well as systematic errors due to inhomogeneity in the horizontal structure of the refractivity have been excluded. The results shown in Figure 9-4 thus represent a lower bound of the temperature errors that are due to random measurement errors. Moreover, the effect of water vapor has not been incorporated into the simulation results, and, for most atmospheres, the presence of water vapor obscures the interpretation of the refractivity below about 7 km. The retrieval of moisture information in the lower troposphere has been discussed in the previous chapter. Further discussion of simulations of temperature and

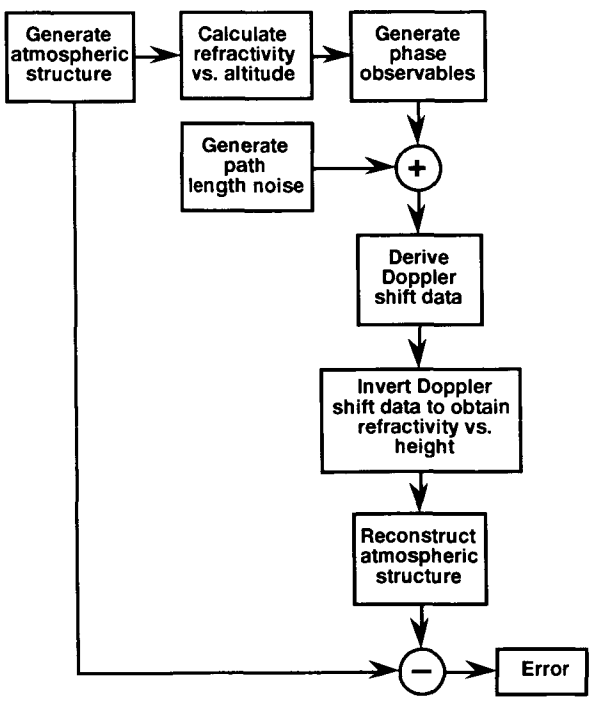


Figure 9-3. Procedure for simulation of some error sources that contribute to the retrieval of atmospheric profiles.

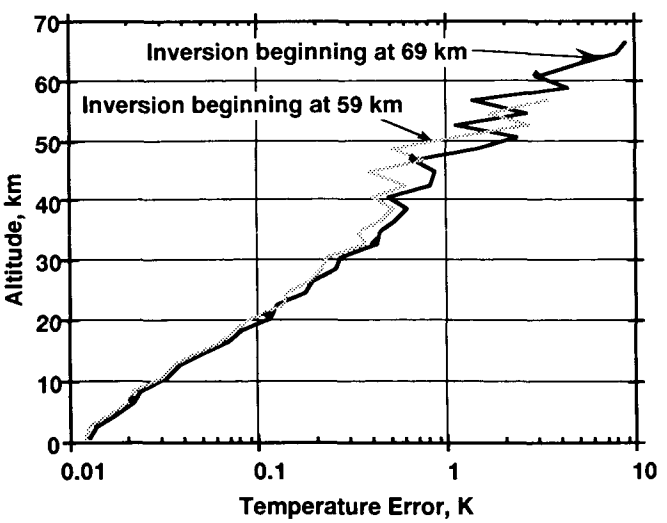


Figure 9-4. Temperature error due to the sum of error sources shown in Figure 9-2. (Water vapor in the troposphere is ignored.)

water vapor reconstruction in the troposphere is given in the following section.

Tropospheric Simulations

Diffraction limits the resolution achievable in the application of geometric optics theory used to interpret and invert the radio occultation measurements. However, this limit is not due to the receiving aperture. Instead, because the signals are coherent, the resolution is set by the diffraction limit of the geometric optics approximation; namely, the vertical and the cross-beam dimensions of the sampling volume are given approximately by the diameter of the first Fresnel zone. In the absence of bending, the first Fresnel zone is given approximately by the diameter of a cone with its apex at the receiver, its axis being the distance from the receiver to the limb of the Earth, and the rays forming the cone having a length of this same distance plus one-half the radio wavelength (see Chapter 11). With a receiver in orbit at an altitude of 700 km, which puts the limb at a distance of 3000 km from the receiver, and a wavelength of 20 cm, this diameter is about 1.5 km. The vertical resolution also improves as the ray descends through the atmosphere because the vertical refractivity gradients increase in magnitude causing the ray path to bend and the vertical dimension of the first Fresnel zone to decrease (Fjeldbo and Eshleman 1968). An example of this is shown in Figure 9-5.

This example represents a simulated observation from a GPS receiver in LEO at a 700-km altitude using the temperature and humidity data from the radiosonde out of Hilo, Hawaii, on July 11, 1991, at 12:00 UTC. Near the top of the stratosphere, the diameter of the first Fresnel zone is about 1.4 km and decreases to about 0.45 km near the surface. The gradual decrease in size is due to the exponentially increasing density—and therefore refractivity and refractivity gradient—with decreasing altitude. The sharp changes apparent in the troposphere are generally associated with changes in water vapor density. As can be seen, the Fresnel zone at times drops abruptly several hundred meters in its height component. These cases are associated with very sharp vertical refractivity gradients at an inversion layer boundary near a 3.8-km altitude and near the surface. Given these vertical resolutions, the along-path length of the sampling volume is typically near 200 km. However, it can decrease to 100 km in the lower troposphere when very sharp vertical refractivity gradients are encountered. The parameters recovered using this technique, namely, N, ρ, P, T, and, in some cases, water vapor, represent averages across the sampling volume.

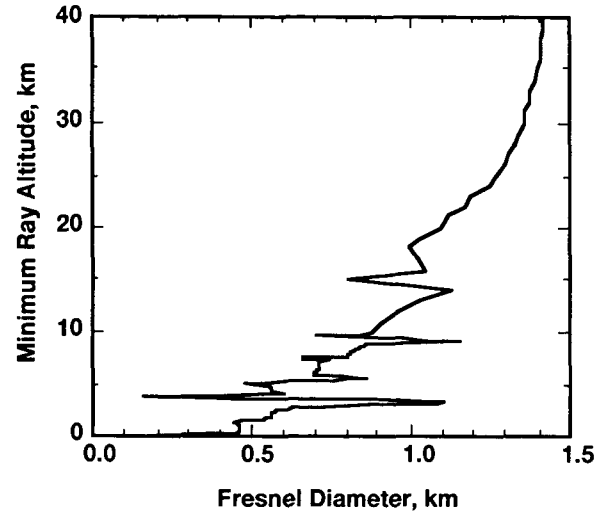


Figure 9-5. Vertical resolution versus altitude.

Both temperature and moisture contribute to the refractive index in the lower troposphere (see eq. (7.36)); in general, it is difficult to separate the two contributions. A special case exists in the tropics where the temperature profiles are relatively uniform over large distances and for long periods. This situation can be used to provide estimates of moisture profiles. During July of

1991, soundings over a 12-day period were available from Hilo, Hawaii. For the 23 soundings from this period, a single average vertical profile of the temperature and the dry term of eq. (7.36) was computed. For each individual radiosonde sounding, a profile of the refractive index (dry plus moist term) is calculated; this profile illustrates the type of information that will be available from occultation measurements. A profile of the observed refractive index and the average dry term for Hilo, Hawaii, on July 7, 1991, at 00.00 UTC is shown in Figure 9-6. In the altitude range between 8 and 10 km, the refractive index profile is very close to the mean dry term, and moisture is of little consequence. Below 8 km, the total refractive index begins to deviate from the mean dry term and exhibits the structure caused by moisture. To obtain the contribution of moisture to the refractive index, the average dry-term profile, representing the average of all 23 soundings, is subtracted from the computed refractive index for this particular sounding. Then from the moist term of eq. (7.36), the water vapor pressure is calculated. The result of this calculation and the observed water vapor pressure are also shown in Figure 9-6. The error in the retrieved water vapor pressure expressed as a percentage is shown in the right side of the figure. Below the inversion at about 3 km, the retrieved water vapor has an error of less than 10%. Above the inversion, the moisture decreases rapidly and the percentage errors are large. However, the trend of the water vapor profile is still preserved up to 7 km.

The moisture profiles in the tropics often have very large vertical gradients. The profiles in Figure 9-6 are plotted at 400-m height intervals, but the "observed N" profile has been smoothed to 1.2 km, which is close to the upper limit of the

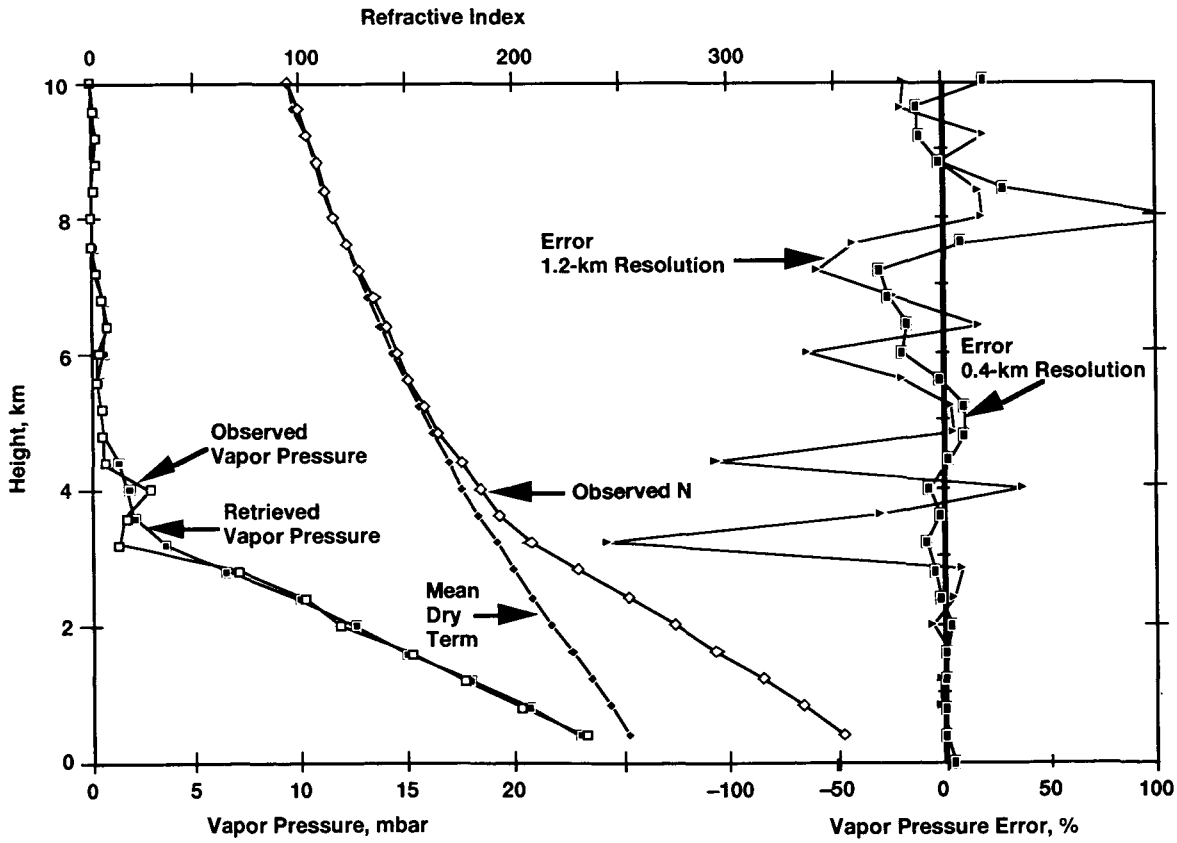


Figure 9-6. Simulated moisture retrieval for Hilo, Hawaii, on July 7, 1991, 00.00 UTC.

vertical resolution for the occultation measurements. The plots of the water vapor profiles are for the observed data at 0.4 km, but the retrieved estimates were computed from profiles smoothed to 1.2-km. The importance of the vertical resolution on the error is clear in comparing the magnitude of the error of the 1.2 km smoothed results with those at 0.4 km. It is apparent that the small-scale vertical structure of the water vapor profile is often lost in the retrieved profile. If the measurement of the basic refractive index profile has a vertical resolution of 0.4 km, then the errors above the inversion are reduced considerably. It should be understood that the results in Figure 9-6 are optimistic in the sense that they do not include any error in refractivity observations, and they ignore any horizontal gradients in the refractivity. It also should be noted that the height of the inversion layer should be retrievable to high accuracy assuming the layer extends for at least 200 km, which is approximately the horizontal resolution for limb sounders. As indicated in Figure 9-5, where the vertical resolution shrank to approximately 200 m in the vicinity of the inversion layer boundary near 3.8 km, the vertical resolution is extremely good in the vicinity of marine inversion layers. This fact combined with the very sharp changes in the gradient of refractivity should make detection and characterization of these features a natural product of these observations. Also Fresnel diffraction techniques can be used to sharpen the vertical resolution of the layer (see Chapter 11).

Another special situation occurs during the polar winter, where water vapor content is small at the surface of the Earth due to the extremely cold weather. In order to estimate how well the dry and wet terms in this region can be separated, a selected set of the TIGR radiosonde data is used as follows: From the temperature, humidity, and pressure measurements of a particular TIGR radiosonde data, the vertical refractivity profile can be computed. The computed refractive index is then modified by adding a white Gaussian noise value having a mean of zero and a standard deviation that varies with height. The standard deviation is assumed to be 50% of the refractivity at 70 km and decreased exponentially to 0.005% of refractivity near the surface. These values are consistent with a measurement error of about 0.1 cm in path delay near 70 km and 1 cm near the surface. The noise-modified refractivity profile is used as a measurement series; we attempt to recover either temperature or moisture from these measurements.

The result of a retrieval for a cold polar profile is shown in Figure 9-7. The surface temperature for this sounding is cold at 248 K, and the dew point is about 4 K lower than the temperature, up to about 8 km. Although the air has a low water content, it helps to assume that it is not completely dry. Consequently, a relative humidity of 50% was used to represent the air below 7.5 km. We also assume the temperature profile in order to compute the contribution of moisture to the total refractive index. For the sole purpose of estimating the contribution of moisture to the

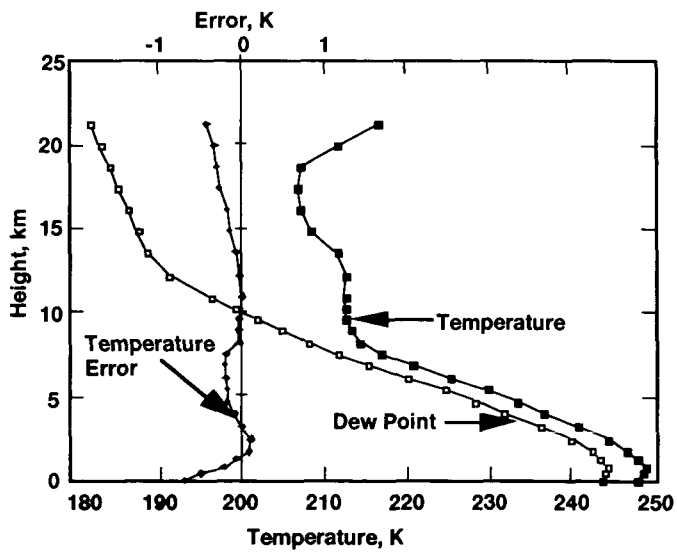


Figure 9-7. Temperature error for TIGR polar-2 sounding #1201.

refractivity, a mean lapse rate temperature of 6 K/km was used from 7.5 km down to the surface. The temperature from 70 km down to 7.5 km was retrieved. Below 7.5 km, the estimated moisture term of the refractivity was subtracted from the noise-modified refractivity profile in order to get the temperature or density contribution. Then the temperature is retrieved applying the same process used for dry air. The resulting temperature error is within 0.7 K, and for much of the sounding, the error is less than 0.3 K. This error level is typical of the type of temperature retrieval possible when the temperature near the surface or at any other level is colder than 250 K. When the surface temperature approaches 270 K, then the unknown contribution of the moisture increases the error in the retrieved temperature profile.

Simulation of a Weather Front

When a GPS is occulted by the Earth, its radio path traverses about 2000 km of the atmosphere along a path that is nearly tangent to the Earth's surface. The simulations shown above assume a horizontally uniform atmosphere over this 2000 km. Under these conditions, the Abel transform provides a very accurate reconstruction of the refractive index profile. However, the Earth's atmosphere often has considerable structure in the horizontal dimension. (The effect of variable horizontal structure was discussed in Chapter 8.) In an attempt to estimate the upper limit of the impact of horizontal structure of the atmosphere on the retrieval of profiles from GPS data, we have simulated an occultation through the center of a strong frontal surface. We believe this represents the upper limit because (1) an extreme weather front is assumed, (2) the occultation ray paths are orthogonal to the front line, and (3) on average, the tangent point of the ray path falls along the center position of the front (distance = 1000 km in Figure 9-8), where the horizontal gradient is largest. It should be understood that if the ray paths were parallel to the front, then the horizontal variations would be much smaller and retrievals would be very accurate.

The temperature structure for the cross section is shown in Figure 9-8; it was adapted from a cross section presented in Palmen (1969). Only data up to a height of 14 km are shown in Figure 9-8, but for the occultation simulations, both temperature and moisture information from the surface up to a height of 70 km was generated. The cross section shows the strong front extending from the surface at a distance of about 1500 km up to a height of about 12 km at a distance of 700 km. The change across the 2000 km of the cross section represents a very pronounced front that might be found at mid-latitudes over continents during the winter. The structure in the cross section of the stratosphere and the mesosphere also contains features that are consistent with those we would expect to find between a very cold and a very warm air mass.

The refractive index was computed from the temperature and moisture cross sections. The change in the refractivity from one end of the cross

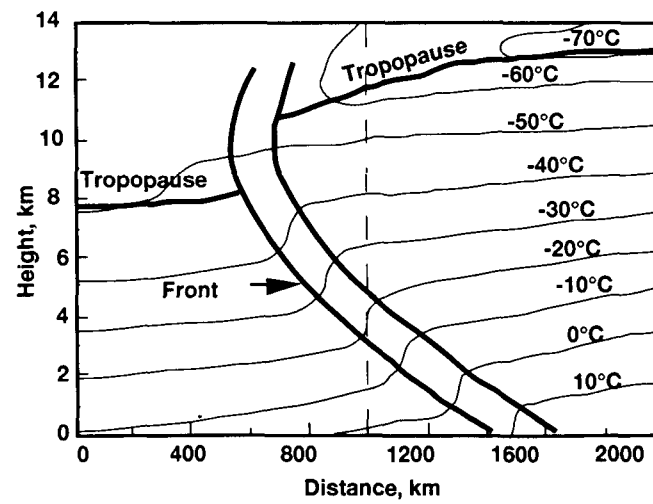


Figure 9-8. Vertical cross section of temperature for a strong frontal surface.

C-2

section to the other ranged from about 50% at 70 km to about 10% near the surface. Typical variability of the refractivity within the troposphere of the cross section was about 15%.

Using this weather front structure, the accuracy of tropospheric refractivity retrievals under somewhat extreme conditions can be assessed. Applying the simulation system described above (and in Figure 9-3), signal paths from the GPS satellite to a receiver in LEO were traced through the weather front shown in Figure 9-8 to generate a simulated data set. This was inverted to recover the apparent vertical refractivity structure at the center of the cross section at 1000 km assuming local spherical symmetry (the Abel transform solution). If the atmospheric structure for the cross section exhibits only a linear change along the ray path, the Abel transform still provides an excellent solution for the center profile. The reason for this is that although an excess in the refractive index will cause an increase in the observed phase change from one side of the center, a corresponding deficiency in the index of refraction on the other side of the center will offset the phase change. On the other hand, the refractive index structure that is symmetric about the center profile, either as a relative minimum or maximum value at the center, will introduce errors when the Abel transform solution is used (see Chapter 8). Thus the retrievals provide an initial estimate of the effect of symmetric atmospheric structure relative to the center (at 1000 km). The results are shown in Figure 9-9. Except at the very highest levels and near the surface, the errors can be less than 1%, and for many heights the error is less than 0.6%, even with the local spherical symmetry assumption. The Abel transform solution conveniently provides initial estimates of the effect of a non-uniform atmosphere on the retrieval of the refractive index profiles. However,

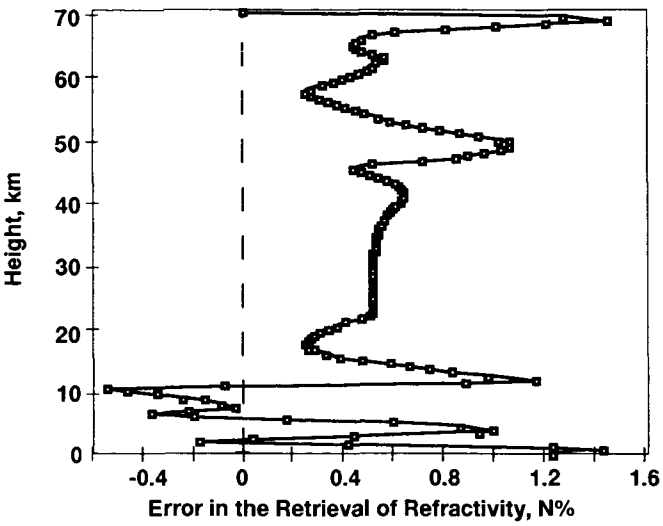


Figure 9-9. Errors in the retrieval of refractive index at the center of the cross section shown in Figure 9-8.

other techniques have already been developed to handle a non-uniform atmosphere (Lindal 1992). When these are applied, the errors in the recovery of refractive index are expected to be generally smaller than those shown in Figure 9-9.

The pressure profile is computed by integrating the refractive index profile beginning at the top of the profile, or 70 km in our example. The integration can proceed down to a height where moisture starts contributing to the refractive index (which in the profile of Figure 9-8 is about 8 km). In order for the integration to proceed to the surface, an average lapse rate in temperature of 6 K/km is assumed from 8 km to the surface. This step leads to an estimate of the first term of eq. (7.36). By subtracting this estimate from the simulated refractive index, information on the moisture profile is obtained.

The results of this retrieval procedure are shown in Figure 9-10. Since the pressure at the top of the profile would, in general, not be known, an error of 5% in the pressure at a height of 70 km is introduced. The effect of the unknown pressure at the start of the integration is largely removed after descending about two scale heights to near

60 km. For much of the stratosphere, the error in the temperature is less than 0.5 K. The variable structure in the cross section near the tropopause leads to an increased error in the temperature profile at heights between 9 and 12 km. Below 8 km, the error is a reflection of how well the assumed average lapse rate matches the actual profile. The large errors in the temperature estimates in this altitude region also cause large errors in the water vapor estimates. As shown previously (see Chapter 8), the retrieval of moisture in the lowest 5 km of the troposphere depends very strongly on how well the temperature can be estimated independent of the occultation observations. The results for the lower troposphere shown in Figure 9-10 should be viewed as upper limits because they do not take advantage of additional information on the temperature and pressure structure, which would generally be available from other sensors and weather models.

The meteorological situation of the front is in marked contrast to the situation encountered with marine boundary layer inversions where the refractive index profile is useful for identifying the height of the inversion because both temperature and moisture contribute to the existence of a very strong refractivity gradient at the top of the boundary layer. In the case of a front, the moist and dry effects tend to cancel each other when the point of tangency is centered on the front. On the front's cold side, the air is denser but contains less water, so the dry contribution is larger and the moist term is smaller than the relative contributions of their counterparts on the warm side. The net result is that remarkably little microwave refractivity contrast exists across the front when the tangent point of the occultation is centered on the front.[3] Therefore, for a highly variable atmosphere in the mid-latitudes, the optimum use of the occultation measurements may be in using the refractive index profile as a limit for temperature and moisture contributions.

---

[3] For an arbitrarily placed occultation tangent point in the plane of Figure 9-8, the phase delay due to the front is given approximately by

$$\delta(\Delta\rho(h_a)) = 10^{-6} \int_{-\infty}^{\infty} \delta N(h_a, s) W(s) ds$$

where δN is the change in refractivity due to the front and $W(s)$ is the weighting function for limb sounders, which is given by a normalized Gaussian distribution for an exponential atmosphere (see eq. (10.3)) and shown in Figure 10-3. Here, s is the path length along the ray measured from the point of tangency and h_a is the altitude of the ray at that point. The quantity δN depends on changes in both the local air density and the water vapor density as a result of the temperature change across the front. If δN can be constrained by a meteorological model, so that the degrees of freedom in the model describing δN and the location of the front versus altitude are no more than the number of independent $\Delta\rho(h_a)$ measurements made at different altitudes h_a during the occultation, then the above convolution integral should be invertible. Using a local reference (front-free) model for the refractivity profile to form the $\delta(\Delta\rho(h_a))$ measurement sequence, the values of the model parameters should be recoverable. This would enable one to recover both the profile of the front and the variation of δN with altitude.

For example, when δN may be treated as a height-dependent step function (and therefore antisymmetric) across the front that is located at $s_o(h_a)$, the extra phase delay due to the front becomes

$$\delta(\Delta\rho(h_a)) \doteq 10^{-6}\delta N(h_a)\,\mathrm{sign}[s_o(h_a)]\,\mathrm{Erf}\left[|s_o(h_a)|/\sqrt{(2aH)}\right]/2$$

where H is the local scale height of the reference refractivity model and a is the geocentric radial distance of the point of tangency. Although this expression does yield a zero contribution from the front at $s_o = 0$ for the aforementioned symmetry reasons, it clearly shows a significant frontal signature if the location function for the front $s_o(h_a)$ causes s_o to slide across the $W(s)$ distribution as the altitude h_a of the occultation varies.

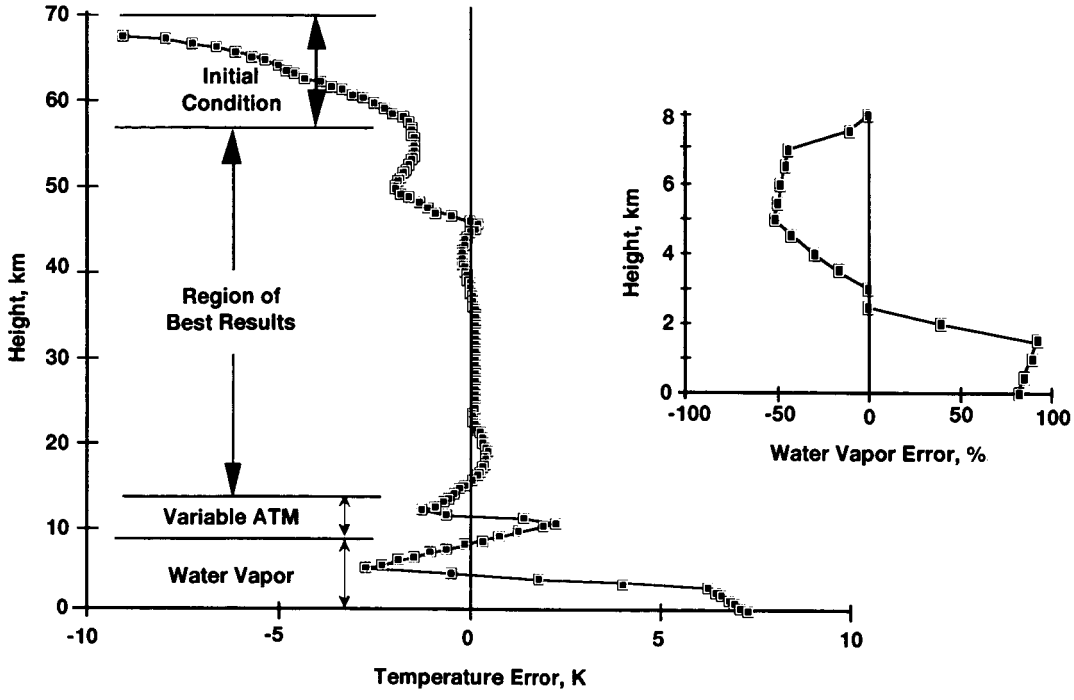


Figure 9-10. The error in the retrieval of temperature and water vapor at the center of the cross section shown in Figure 9-8.

CHAPTER

10 SPATIAL RESOLUTION TOPICS FOR LIMB SOUNDERS

Introduction

A characteristic of limb sounders is relatively fine vertical resolution but relatively coarse horizontal resolution. Along-track resolution is essentially controlled by the vertical resolution (if local horizontal isotropy prevails), which in turn is limited by a combination of the measurement sampling interval and Fresnel diffraction. Cross-track resolution is purely Fresnel-limited. On the other hand, horizontal resolution can be sharpened by increasing the density of sounder measurements through deployment of multiple LEOs and/or by trading off temporal resolution versus spatial resolution.

Along-Track Resolution

If Δh is the vertical resolution, then one measure of the along-track resolution ΔL is given by

$$\Delta L = 2\sqrt{2a\Delta h} \qquad (10.1)$$

which is merely the length of the chord that is tangent to the inner of two concentric circles whose radii differ by a small amount Δh. Here, a is the geocentric distance of closest approach of the ray to the Earth's surface—for grazing incidence at the Earth's surface and for $\Delta h = 1$ km, $\Delta L \approx 225$ km. Thus, each subsequent phase observation of an ingressing occultation includes an additional contribution to the path integral for α, which arises from the line elements along the ray spanning a spherical cap of chord length ΔL (see Figure 7-2). It follows that the along-track resolution is of the order of ΔL.

Alternatively, the kernel or weighting function in the integral equation for α, eq. (7.20), provides a measure of the relative contribution to α from each atmospheric element along the ray path. Let the refractivity $N(\mathbf{r})$ be defined by eq. (8.25) where f (defined in eq. (8.27)) describes the along-track variability other than the radial behavior. Then, eq. (7.20) becomes

$$\alpha(a) = \alpha_m(a)[1 + K(a)\int_{-\infty}^{+\infty} W(s)f(s)ds] \qquad (10.2)$$

where a change of variable from θ to s has been made in eq. (7.20) (assuming a thin atmospheric approximation) and where s is the arc length along the ray path measured from the point of closest approach. Here, α_m is the spherically symmetrical model-derived value for the bending angle (e.g., from eq. (7.24) or from eq. (8.38)), K(a) is a normalization factor for a given ray path, and W(s) is the weighting function, which is given by the Gaussian distribution (when an exponential model is used for $N(\mathbf{r})$) shown in Figure 10-1.

$$W(s) = \exp[-s^2/2aH] \qquad (10.3)$$

Thus, the standard deviation of W, $\sqrt{aH}$, is about equal to ΔL (for H = 8 km) when Δh is equal to 1 km. Figure 10-1 sketches W(s) in the context of the occultation geometry. For $s = 2\sqrt{aH} \approx$ 450 km, the contribution to α from a point on the ray path at that position is less than 15% of the contribution from an element within the central cap at s = 0.

Another measure of the degree of locality of the atmospheric information in α is given by the weighted mean path length through the atmosphere as defined by

$$\bar{L}(a) = 2 \int_a^{a_M} s(r)N(r)dr \Big/ \int_a^{a_M} N(r)dr \qquad (10.4)$$

where s(r) is the arc length along the ray path. For the Earth's atmosphere $\bar{L}(a)$ is about 360 km for all values of a except those very near a_M, where it linearly approaches zero. The upshot from any of these concepts is that the radio occultation technique for a single occultation recovers atmospheric refractivity profiles that are primarily derived from mesoscale-sized along-track averages.

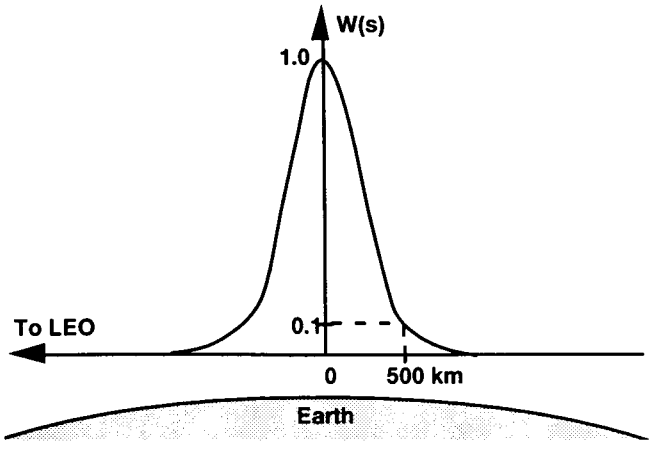


Figure 10-1. Along-track weighting function for atmospheric limb sounders.

Vertical Resolution

Figure 8-4 is a plot of the defocusing factor $\zeta(a)$ resulting from the refractivity gradient for an exponential atmosphere and is given by eq. (7.31) for dry air. For a fixed temporal sampling interval it also provides the corresponding relative vertical sampling interval. However, the resolution is usually limited by the size of the first Fresnel zone for temporal sampling intervals less than about 0.5 s. The rationale for choosing the first Fresnel zone is that the Fresnel diffraction theory shows that 50% of the radiation arriving at the LEO originates from directions within the small solid angle subtended by the first Fresnel zone; the remaining 50% is distributed over all other directions. The first Fresnel zone for a non-refracting medium has a radius of $\sqrt{\lambda D}$, where λ is the RF wavelength of the GPS signal (~20 cm) and D is the distance of the LEO from the point of maximum atmospheric immersion of the ray path (~3000 km). The zone subtends an angular radius at the LEO of $\sqrt{\lambda/D}$, or about 1 arcmin. For a refracting medium, the first Fresnel zone also depends on the refractive gradient. Using the thin screen model, we will show in Chapter 11 that the Fresnel zone is given by

$$\Delta h(a) = \sqrt{\lambda D \zeta(a)} \qquad (10.5)$$

which for an exponential atmosphere is about 750 m in the upper stratosphere and about 275 m in the lower troposphere; Δh can be considerably smaller if large water-vapor-induced refractivity gradients are present (see Figure 9-5). We will also show in Chapter 11 how Fresnel deconvolution techniques can be used to sharpen the vertical resolution past the Fresnel limit.

To summarize the diffraction-limited case, the occultation information content in a *single* phase measurement can be characterized as originating from those atmospheric elements that principally lie within a near-horizontal tube with dimensions of approximately 1.5 $\sqrt{\zeta(a)}$ km in height, 1.5 km in width, and about 300 km in length.

CHAPTER

11 TOPICS IN FRESNEL DIFFRACTION

Introduction

Although the atmospheric contributions to a single-phase delay measurement by the LEO principally originate from those directions within the first Fresnel zone (adjusted for the atmospheric refractivity gradient), the overall occultation event comprises amplitude and phase measurement sequences; these sequences involve overlapping Fresnel zones when the temporal sampling interval is less than $2\sqrt{(\lambda D / \zeta)}/\dot{b}$. The measurements for these overlapping zones may be synthesized or "deconvolved" using Fresnel transform techniques to obtain a determination of the vertical location of a local refractivity feature, such as an inversion layer that is significantly sharper than the Fresnel limit. This technique has been successfully used to study the spatial brightness distributions of radio sources occulted by the Moon (Scheuer 1962), the structural aspects of the rings of Saturn (Marouf et al. 1982; Marouf et al. 1986; Tyler 1987), and the determination of the magnitude and wavelength of gravity waves in the upper atmosphere of Uranus (Hinson and Magalhães 1991).

Here we give a brief summary of Fresnel diffraction theory. (For an in-depth treatment, see Goodman 1968 and Born and Wolf 1980.) We then apply Fresnel diffraction to a thin screen model, which represents the actual delay through the Earth's atmosphere; finally, we discuss several useful concepts in diffraction theory.

Strictly speaking, the GPS signal amplitude and phase measurements are not derived from monochromatic carriers but rather from spread spectrum, fully suppressed carriers with an effective bandwidth of the order of 10 MHz. Because this effective bandwidth is only about 1% of the carrier frequency, a monochromatic diffraction approach, which is readily generalized to the spread spectrum case through Fourier transform techniques (Goodman 1968), provides a close approximation to the actual waveform. Also, the L1 and L2 carriers, whose frequencies differ by 350 MHz, present slightly different diffraction properties to the LEO. For example, the radii of the first Fresnel zones for the two carriers differ by 13%.

As noted earlier, the effective path length $\bar{L}$ in the atmosphere over which refracting processes occur is a few hundred kilometers, which is small compared to the distances of the LEO ($D \approx 3000$ km) and the emitting GPS satellite (~25,000 km). The thin screen model's range of validity has been discussed in Hubbard et al. (1978). Two criteria for good results are the ratios $\bar{L}/D$ and $\alpha/\sqrt{(H/a)}$; both should be <<1. Here α is the bending angle due to refraction, a is the impact parameter of the ray, and H is the scale

height of the vertical refractivity profile. For the Earth's atmosphere, $\overline{L}/D \approx 0.1$ and at the Earth's surface, $\alpha/\sqrt{(H/a)} \approx 0.5$. The thin screen approximation improves as these ratios become smaller. Thus, near the Earth's surface the validity of the thin screen model tends to degrade, but the model still gives good qualitative results, particularly at higher altitudes. Results from a thin screen treatment lend insight into diffraction processes in the actual atmosphere. However, to achieve more accurate, enhanced resolution from the phase and amplitude measurements would require at least corrections to the thin screen model that reflect the actual signal propagation through a three-dimensional refracting medium and from a diffracting boundary surface presented by the Earth's limb. For a geometrical optics (stationary phase) treatment of diffraction effects from a boundary surface (such as the Earth's limb), the reader should consult Keller (1962) and (1958).

Fresnel Diffraction

The scalar diffraction approach that we use here is based upon the Rayleigh–Sommerfeld formulation of Fresnel diffraction. This particular formulation uses a Green function approach to solve the scalar differential equation for the electric intensity of the electromagnetic wave; this equation expresses the electrical intensity of the wave at a particular point that results from the distribution of electrical intensities of the wave over an arbitrary, closed, but distant surface surrounding the point. Let S denote the surrounding surface (Figure 11-1), and let $\mathcal{E}(\mathbf{r},t)$ be the electric field strength of the wave at position $\mathbf{r}$ on the surface and at time t. In the case where $k|\mathbf{r}-\mathbf{r}_L| >> 1$, where k is the wave number of the GPS signal ($2\pi/\lambda$), it can be shown using the Helmholtz–Kirchoff integral theorem (Goodman 1968) that the observed electric field strength of the wave at the position $\mathbf{r}_L$ is given by the surface integral

$$\mathcal{E}(\mathbf{r}_L,t) = \frac{1}{2\pi c}\iint_S \frac{\partial \mathcal{E}(\mathbf{r},t_r)}{\partial t}\,\frac{(\mathbf{r}-\mathbf{r}_L)\cdot d\mathbf{S}}{|\mathbf{r}-\mathbf{r}_L|^2} \qquad (11.1)$$

Here, $\mathbf{r}$ is the position vector of the surface element $d\mathbf{S}$; $d\mathbf{S}$ is directed outward along the normal vector to the surface at $\mathbf{r}$ (away from the LEO); and t_r is the retarded time associated with that surface element relative to the position $\mathbf{r}_L$, i.e., $t_r = t - |\mathbf{r}-\mathbf{r}_L|/c$. Thus, $t - t_r$ is the time delay required for the wavelet departing from the ele-

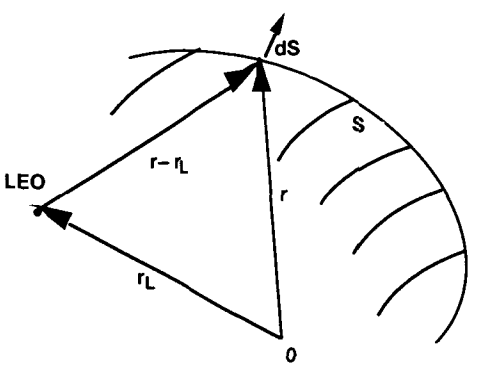


Figure 11-1. Surface of integration for Helmholtz–Kirchoff integral theorem.

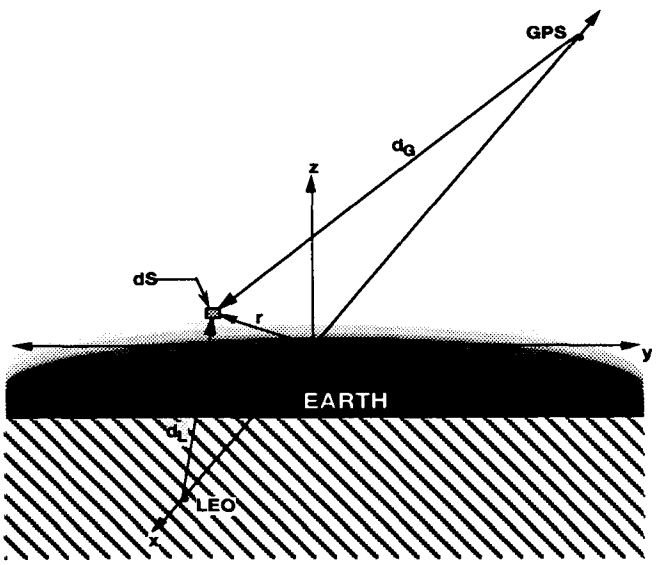


Figure 11-2. Thin screen model representing atmospheric phase delays.

ment dS to reach the position $\mathbf{r}_L$. The expression in eq. (11.1) also applies to the more general case where the wave is not necessarily monochromatic, but requires that each of its Fourier components satisfies the condition $k|\mathbf{r} - \mathbf{r}_L| >> 1$.

The Thin Screen Model

The surface integral formulation in eq. (11.1) enables us to express the phase and amplitude of the GPS signal received by the LEO in terms of the phases and amplitudes of a continuous ensemble of electromagnetic radiators embedded in a thin screen located above the Earth's limb between the LEO and the emitting GPS satellite. The Earth, treated as an opaque disk, lies in the yz plane of Figure 11-2, and the Earth's atmosphere is represented by the thin screen also in the yz plane. The LEO is located at $(x_L, 0, 0)$. The emitting GPS satellite is located behind the thin screen at $(x_G, 0, 0)$ and irradiates the back side of the screen. As the occultation event evolves with time, the thin screen and the Earth's disk move in only the z direction with a velocity $\dot{b}$ corresponding to the rate of ascent or descent of the undeflected ray in the actual case. The vector $\mathbf{r}$ denotes the position $(0, y, z)$ of a surface element of the area dydz in the thin screen. The point $(0, 0, 0)$ corresponds to an altitude in the actual atmosphere of h_b relative to the opaque surface of the Earth's limb. Note that h_b will be both positive and negative during the course of the occultation event depending on the value of the impact parameter a of the ray; that is

$$h_b \doteq h_a - x_L\alpha \tag{11.2}$$

When the GPS signal arrives at the thin screen, an additional phase delay $\Psi(y, z + h_b)$ is assumed to modulate the phase of the GPS signal as it departs from the thin screen, which reflects the actual phase delay experienced by the signal in traversing the Earth's atmosphere. This phase delay is given by

$$\Psi = k10^{-6} \int_{\mathscr{E}(h_a)} N(r)\, ds \tag{11.3}$$

where the integration is along the actual ray path $\mathscr{E}(h_a)$ through the Earth's atmosphere and has an impact parameter value of h_a in altitude. The thin screen approach also assumes local along-track isotropy in the refractivity ($f(s) = 0$). The criteria for using a straight-line approximation for $\mathscr{E}(h_a)$ have also been discussed in Hubbard et al. (1978).

The signal from the emitting GPS satellite $\mathcal{E}(\mathbf{r}, t_r)$ arriving at the thin screen at the position $\mathbf{r}$ and at the retarded time t_r (now assuming that the signal is monochromatic) is given by the real part of the expression

$$\mathcal{E}(\mathbf{r}, t_r) = \frac{\tilde{A}}{d_G} \exp\left[-i\omega\left(t_r - \frac{d_G}{c}\right)\right] \tag{11.4}$$

where $\tilde{A}$ is the real amplitude of the sinusoidal signal emitted by the GPS satellite and d_G, which accounts for the time delay for the emitted wave to arrive at the screen and for the space loss in signal strength, is the distance $|\mathbf{r} - \mathbf{r}_G|$ of the point $\mathbf{r}$ in the thin screen from the GPS satellite (Figure 11-2). Similarly, let the signal at the LEO, $\mathcal{E}(x_L, t)$, be measured by the GPS receiver in terms of a complex "voltage" to be denoted by $E\exp[i\phi - i\omega t]$. From eq. (11.1) and Figure 11-2, it follows that the phase and amplitude of the signal at the LEO is given by

$$E(h_b)\exp[i\phi(h_b)] = \frac{k}{2\pi i}\iint_S G(y,z)\frac{A(y, z+h_b)}{d_L d_G}\exp[ik(d_L + d_G)] \times \exp[i\Psi(y, z+h_b)]\, dy\, dz \tag{11.5}$$

where G(y,z) is a normalized gain function of the receiver's antenna, which is a function of the direction from the LEO to the surface element located at $\mathbf{r}$ in the thin screen, and the boresight direction of the antenna, $A(y, z + h_b)$ is a lumped amplitude coefficient that accounts for the magnitude $\tilde{A}$ of the emitted GPS signal, the obliquity factor $(\mathbf{r}-\mathbf{r}_L)\cdot d\mathbf{S}/|\mathbf{r}-\mathbf{r}_L|\,dS$ (which is essentially unity for the Earth's neutral atmosphere), and the cutoff boundary defined by the limb of the opaque Earth; thus, $A = 0$ for $z + h_b < 0$. We have suppressed the time-dependent part of the signal $\exp[-i\omega t]$, which shows up on both sides of eq. (11.5) for monochromatic waves.

Using the Fresnel approximation ($d \doteq x + (y^2 + z^2)/2x$) which assumes that the LEO and the emitting GPS satellite are very distant from the thin screen (more precisely, x_L, $|x_G| >>$ radius of the first Fresnel zone on the screen), eq. (11.5) can be reduced to

$$E(h_b)\exp[i\phi(h_b)] = \frac{\exp\left[ik(x_L + |x_G|)\right]}{x_L|x_G|} \times \frac{k}{2\pi i}\iint_S G(y,z)A(y,z+h_b)\exp\left[ik\left(\frac{y^2+z^2}{2D}\right)\right] \times \exp\left[i\Psi(y,z+h_b)\right]dy\,dz \qquad (11.6)$$

where D, to account for the effect of the finite distance of the emitting GPS satellite, is defined by a "reduced" distance given by

$$\frac{1}{D} = \frac{1}{x_L} + \frac{1}{|x_G|} \qquad (11.7)$$

which results in about a 10% reduction from the value of x_L. The factor $\exp[ik(x_L + |x_G|)]/x_L|x_G|$ is a function only of the spacecraft locations in the x direction. These locations are assumed to be constant in the thin screen model. This factor can therefore be dropped (or absorbed into the normalization constant K defined below).

We now reduce this formulation to a one-dimensional problem by assuming invariance in the y direction. (For resolution studies of possible out-of-plane refractivity gradients, eq. (11.6) would provide the starting point.) Also, the curvature of the Earth's limb in the y direction is negligible so that it can be treated as an opaque semi-infinite straightedge. Inasmuch as the radius of the first Fresnel zone is much smaller than the radius of the Earth, this is a valid approximation. In this case eq. (11.6) becomes

$$E(h_b)\exp[i\phi(h_b)] = K\int_{-\infty}^{\infty} G(z)A(z+h_b)\exp\left[i\left(k\frac{z^2}{2D} + \Psi(z+h_b)\right)\right]dz \qquad (11.8)$$

The quantity K is a complex normalization constant that would yield the observed voltage $E_0\exp[i\phi_0]$ at the LEO in the absence of the thin screen and the Earth's disk. That is, (with $G \equiv 1$)

$$K = \frac{E_0}{\sqrt{\lambda D}}\exp\left[i\left(\phi_0 - \frac{\pi}{4}\right)\right] \qquad (11.9)$$

Finally, a simple change of variable $h = z + h_b$ in eq. (11.8) results in

$$E(h_b)\exp[i\phi(h_b)] = K\int_0^{h_M} G(h-h_b)A(h)\exp\left[i\left(k\frac{(h-h_b)^2}{2D} + \Psi(h)\right)\right]dh \qquad (11.10)$$

where h is the altitude of the local point in the thin screen and h_b is the altitude of the LEO–GPS line at the point of intersection with the thin screen. The geometry for this expression is shown in Figure 11-3, where for simplicity we have assumed the emitting GPS satellite to be at an infinite distance.

If we assume zero optical depth for the thin screen, which is equivalent to assuming no attenuation of the amplitude of the GPS signal

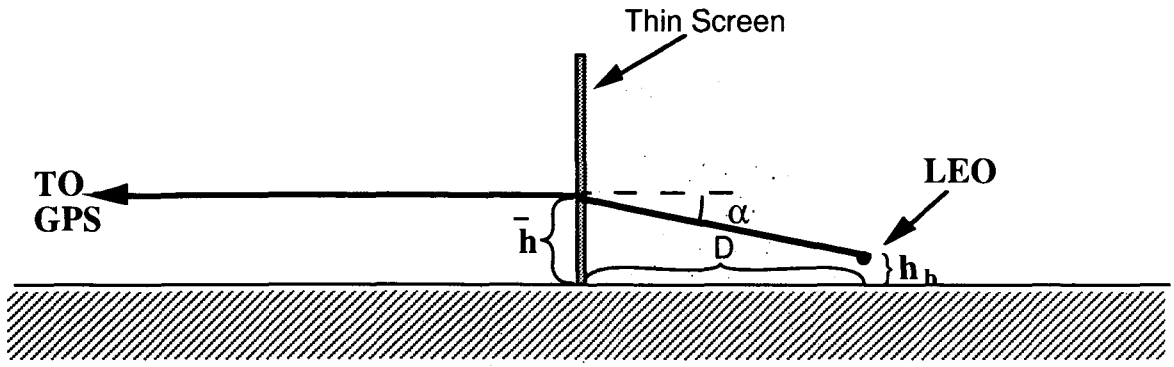


Figure 11-3. One-dimensional thin screen showing path of stationary phase for LEO at h_b.

due to absorption or scattering during its traverse of the Earth's atmosphere, then for $0 \leq h_a$, $A(h_a) = 1$ (no attenuation) and for $h_a < 0$, $A(h_a) = 0$; the value of the obliquity factor is set to unity. Also, for most GPS antennas, the gain function $G(h - h_b)$ is essentially flat over the angular range of interest here and can be set to unity.[1]

The phase modulation $\Psi(h)$ in the thin screen may be expressed as

$$\Psi(h) = \psi_o + \psi(h) + \varepsilon(h) \qquad (11.11)$$

where ψ_o is a constant phase resulting from the coherence of the assumed planar wave arriving at the left side of the screen from the emitting GPS satellite; it may be set to zero because the normalization constant K defined in eq. (11.9) already includes the phase delay due to $|x_G|$ as well as x_L. The quantity $\psi(h)$ is the "deterministic" component of the phase, which is controlled by the reference refractivity model used for the atmosphere. The quantity $\varepsilon(h)$ is the "unknown" phase component resulting from a corresponding "unknown" refractivity component in the atmosphere, which is to be analyzed using diffraction techniques. The quantity $\psi(h) + \varepsilon(h)$ represents the additional phase modulation imposed at the Earth's limb on the phase of the incoming GPS signal and, therefore, it is the additional phase on the signal as it emerges from the thin screen in the direction toward the LEO.

Stationary Phase Path

The vertical position $\bar{h}$ denotes the altitude in the thin screen of the incoming ray that yields a stationary phase path when the LEO is located at $(D, 0, h_b)$. Returning momentarily to a geometric optics approach, the ray traveling from the emitting GPS satellite through the thin screen at an altitude of $\bar{h}$ (equivalent to the impact parameter h_a in the actual case) and then continuing to the LEO is defined as a stationary phase path relative to all other neighboring paths satisfying the boundary conditions. From Figure 11-3, the phase residual at the LEO for an arbitrary ray arriving at the screen at altitude h from the GPS satellite (assumed to be at infinity) is given by

$$\phi(h, h_b) = \phi_o(h_b) + k\left[\sqrt{D^2 + (h - h_b)^2} - D\right] + \Psi(h) \qquad (11.12)$$

where $\phi_o(h_b)$ is the phase at the LEO in the absence of the thin screen and the Earth's limb. It follows from Figure 11-3 that a stationary phase condition ($\partial\phi/\partial h = 0$) on the path is given by

[1] Even for a ~15-dB "high gain" antenna (for example, a dish 30 cm in diameter), the 3-dB rolloff angle from boresight is about 0.3 rad, corresponding to an altitude difference in the thin screen of about 1000 km.

$$\left.\frac{d\Psi}{dh}\right|_{h=\bar{h}} = -k\sin\alpha\Big|_{h=\bar{h}} \doteq -k\alpha(\bar{h}) \qquad (11.13)$$

where $\alpha(h)$ is to be identified with the refractive bending angle in the actual case. If we let $\alpha(h)$ correspond to the bending defined by the reference refractivity model in eq. (7.25), then it follows that the reference phase modulation $\psi(h)$ is given by

$$\psi(h) = 10^{-6}kN_o\sqrt{2\pi aH}\exp\left[-\frac{h}{H}\right] \qquad (11.14)$$

The First Fresnel Radius in a Refracting Medium

Let $\Phi(h)$ be the reference component of the phasor defined in the integrand of eq. (11.10)

$$\Phi(h) = k\frac{(h-h_b)^2}{2D} + \psi(h) \qquad (11.15)$$

If we assume that $\Phi(h)$ is smooth and expand it in a Taylor series about $\bar{h}$, the point of stationary phase, then using eq. (11.13) we obtain

$$\Phi(h) = \frac{k}{2D}\left\{\left[1 - D\frac{d\alpha(\bar{h})}{dh}\right](h-\bar{h})^2 + (\bar{h}-h_b)^2\right\} + \Psi(\bar{h}) + O\left[(h-\bar{h})^3\right] \qquad (11.16)$$

The first Fresnel zone is centered at $\bar{h}$, and its radius $(h - \bar{h})_F$ is defined by the condition

$$\Phi(h) - \Phi(\bar{h}) = \pi \qquad (11.17)$$

from which it follows that the first Fresnel radius is given by

$$(h-\bar{h})_F = \sqrt{\lambda D\zeta(\bar{h})} \qquad (11.18)$$

where $\zeta(\bar{h})$ is the defocusing factor and given by

$$\zeta^{-1} = 1 - D\frac{d\alpha}{dh} \qquad (11.19)$$

For the exponential refractivity model, $\zeta(h)$ is given by eq. (7.31) and exhibited in Figure 8-4. An example of $\sqrt{\zeta(h)}$ from an actual profile of atmospheric refractivity is given by Figure 9-5 (scaled by the factor $\sqrt{\lambda D}$).[2]

Enhanced Resolution Using Fresnel Diffraction Analysis

The degree to which a feature in the refractivity profile may be resolved partially depends upon the complexity of the profile itself. Here we consider a fairly simple feature immersed in an otherwise smoothly varying medium. Suppose that an inversion layer is transected by the GPS signal and that its lower boundary is at an altitude h^*. We will assume the layer is horizontally laminar and that there is a small discontinuity in the temperature and, hence, in the refractivity that occurs at h^*. We will characterize this inversion layer by assuming

$$\varepsilon(h) = \begin{cases} 0 & \text{for } 0 \le h \le h^* \\ \varepsilon_o & \text{for } h^* < h \end{cases} \qquad (11.20)$$

Using the thin screen model, the perturbed observation made by the LEO at the altitude h_b, $\tilde{E}\exp[i\tilde{\phi}]$ (Figure 11-4) is obtained by evaluating eq. (11.10) (with $A = G \equiv 1$) using eq. (11.20) for $\varepsilon(h)$ and eq. (11.14) for ψ. Using a stationary phase approximation (Born and Wolf 1980), we obtain

$$\tilde{E}\exp[i\tilde{\phi}] = E\exp[i\phi] + \mathscr{H}[\exp[i\varepsilon_o] - 1]\int_{\frac{u^*-\bar{u}}{\sqrt{\zeta(\bar{h})}}}^{\infty}\exp\left[\frac{i\pi}{2}v^2\right]dv \qquad (11.21)$$

[2] For the two-dimensional case it can be shown using the stationary phase approximation that the first Fresnel zone is an ellipse in a refracting medium. Its semi-major axis of length $\sqrt{\lambda D}$ is oriented perpendicular to the local gradient of the refractivity. The length of its semi-minor axis is given by eq. (11.18).

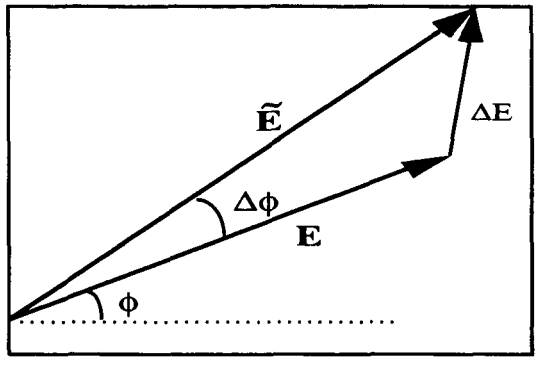


Figure 11-4. The E-vector perturbation by ε(h).

where u is a Fresnel variable given by

$$u = \sqrt{\frac{2}{\lambda D}}h, \qquad u_b = \sqrt{\frac{2}{\lambda D}}h_b \tag{11.22}$$

and where $\mathscr{A}$ is a complex constant for a given value of h_b and given by

$$\mathscr{A}(h_b) = K\sqrt{\frac{\lambda D\zeta(\bar{h})}{2}}\exp\left[i\frac{\pi}{2}(\bar{u}-u_b)^2 + i\psi(\bar{h})\right] \tag{11.23}$$

The quantity $\bar{u}$ (h_b) is given by

$$\bar{u} = \sqrt{\frac{2}{\lambda D}}\,\bar{h} = \sqrt{\frac{2}{\lambda D}}\left[h_b + \alpha(\bar{h})D\right] \tag{11.24}$$

with $\alpha(\bar{h})$ given by eq. (7.26). Thus, a transcendental relationship is to be solved for each value of h_b to obtain the stationary phase value of the Fresnel variable $\bar{u}$ (h_b).

The integral in eq. (11.21) may be expressed in terms of the Fresnel integrals (see Born and Wolf 1980), which are defined by

$$\mathscr{S}(v) = \int_0^v \sin\left[\frac{\pi}{2}y^2\right]dy \tag{11.25}$$

and

$$\mathscr{C}(v) = \int_0^v \cos\left[\frac{\pi}{2}y^2\right]dy \tag{11.26}$$

It is convenient to recast the Fresnel integrals into the form

$$\begin{cases} U(v) = \frac{1}{2}[\mathscr{C}(v) - \mathscr{S}(v)] \\ V(v) = \frac{1}{2}[1 + \mathscr{C}(v) + \mathscr{S}(v)] \end{cases} \tag{11.27}$$

where v is given by

$$v = (\bar{u} - u^*)/\sqrt{\zeta(\bar{h})} \tag{11.28}$$

which is a local Fresnel measure of the difference in altitude between the stationary phase point (for the deterministic component of refractivity) and h*, the altitude of the inversion layer.

Figures 11-5 and 11-6 illustrate the fringe behavior of the functions V(v) and U(v), which are recognized as the real and imaginary components, respectively, of the Fresnel diffraction effect in the observed variations of the phase and amplitude of a monochromatic wave grazing a straightedge (Born and Wolf 1980).

If we ignore, for the moment, Fresnel diffraction effects from the Earth's limb, the expression in eq. (11.21) for the diffracted wave can be rewritten as

$$\tilde{E}\exp[i\tilde{\phi}] = E\exp[i\phi]\left[1 + [\exp[i\varepsilon_o] - 1][V(v) - iU(v)]\right] \tag{11.29}$$

Using Figure 11-4 and eq. (11.29), we obtain the change in the observed phase, $\Delta\phi = \tilde{\phi} - \phi$, as a result of the presence of ε_0 in the thin screen at an altitude h*. It can be shown that

$$\tan(\Delta\phi) = \frac{V(v)\sin\varepsilon_o + 2U(v)\sin^2\frac{\varepsilon_o}{2}}{1 + U(v)\sin\varepsilon_o - 2V(v)\sin^2\frac{\varepsilon_o}{2}} \tag{11.30}$$

Note that for large v, $\Delta\phi \to \varepsilon_0$ (modulo 2π), as it must. This can be further simplified when the phase perturbation ε_0 is small. In this case

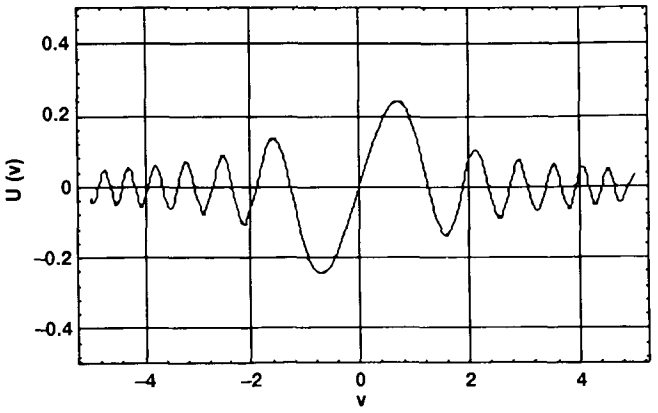


Figure 11-5. Imaginary component of the Fresnel diffraction effect from a straightedge upon a grazing monochromatic wave.

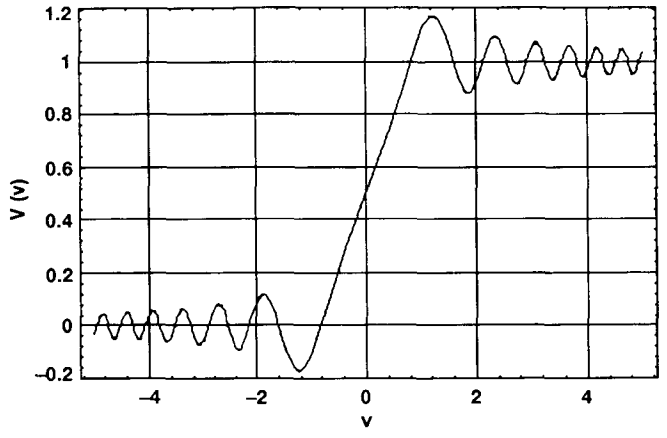


Figure 11-6. Real component of the Fresnel diffraction effect from a straightedge upon a grazing monochromatic wave.

$$\Delta\phi = \varepsilon_0 V(v) + O(\varepsilon_0^2) \qquad (11.31)$$

Thus, for small perturbations ε_0 produces a diffraction pattern that mimics the real component of the straightedge diffraction pattern (Figure 11-6) centered at h*.

The location u* would follow by determining the epoch and, hence, the value of $\bar{u}$, where $\Delta\phi/\varepsilon_0 = 1/2$ and setting $u^* = \bar{u}$ for that epoch. The sensitivity of this determination, $\partial(\Delta\phi)/\partial u^*$, would be maximized at this point and equal to $\varepsilon_0/2$. Roughly, the precision of this determination of the height h* of the inversion layer is given by

$$\sigma_{h^*} \approx \sqrt{(2\lambda D\zeta)}(\sigma_\phi/\varepsilon_0) \qquad (11.32)$$

where σ_ϕ in this case would be the ionosphere-corrected and double-differenced phase measurement error, which for 1-s measurement averages is about 30 mrad using a TurboRogue tracking the P-codes to recover carrier phase. This expression shows that for large "SNR" values, $(\sigma_\phi/\varepsilon_0)^{-1}$, the vertical resolution of the inversion feature can be much smaller than the diameter $2\sqrt{(\lambda D\zeta)}$ of the first Fresnel zone. Moreover, by equating the vertical sampling interval to σ_{h^*} (and noting that $\sigma_\phi = \sigma_{\phi_0}/\sqrt{\Delta t}$, where σ_{ϕ_0} is the SNR error for a 1-s average), we can arrive at an "optimal" temporal sampling interval Δt and therefore a vertical sampling interval that is given by

$$\dot{a}\Delta t = \left[\zeta\sigma_{\phi_0}/\varepsilon_0\right]^{2/3}\left[\left|2\lambda D\dot{b}\right|\right]^{1/3} \qquad (11.33)$$

where $\dot{a}$ is the velocity of the impact parameter.

Similarly, ε_0 would be determined by best matching the amplitude of the Fresnel phase pattern that occurs in the vicinity of u*. Figure 11-6, which illustrates the behavior of the phase fringe pattern with h_b for small ε_0, clearly shows the stationary phase pattern near $\bar{u} = u^*$. A best fit to this theoretical curve would yield a determination of the values of h* and ε_0.

The difference in amplitudes of the perturbed and unperturbed electric intensities also provides useful information. Because the GPS receiver "tracks" the phase of the GPS signal, the nominal value of the *residual* phase within the receiver tracking loop is zero. Therefore, the primary phasor information in the digital signal processor is derived from excursions from zero of the sine correlator output. The amplitude information is complementary with the phase information and primarily carried by the output of the cosine correlator, which nominally would be essentially constant with time under normal signal condi-

tions. However, when spatial inhomogeneities in the refractivity are encountered along the ray path, interference between signal components can occur, leading to deep scintillations and, thus, to strong variations in both the received phase and amplitude. The analog to eq. (11.30) for the amplitudes is

$$\left(\frac{\tilde{E}}{E}\right)^2 - 1 = 2U(v)\sin\varepsilon_o + 4\left[U^2(v) + V^2(v) - V(v)\right]\sin^2\frac{\varepsilon_o}{2} \tag{11.34}$$

which for small values of ε_o reduces to

$$\Delta E(h_b) = \tilde{E} - E = EU(v)\varepsilon_o + O(\varepsilon_o^2) \tag{11.35}$$

For small ε_o, the fringe behavior of $\Delta E(h_b)$ with h_b is given by the imaginary component U(v) of the straightedge diffraction effect and shown in Figure 11-5.

The accuracy of the determination of the inversion layer parameters from the amplitude variations versus those from the phase depends on the relative precisions of the measurements $\sigma_E E^{-1}$ and σ_ϕ. For the TurboRogue, a 20-ms digital correlation product at nominal SNR conditions (6-dB antenna gain) has a value of about 50,000, giving a precision of about 20 ppm. Thus, the measurement precision is the same for both data types, and therefore, it is SNR-limited, about 2% using a 50-Hz sampling rate at nominal SNR conditions. An optimal estimation strategy, of course, would use the best from both the phase and amplitude measurements.

We could also improve the determination of h* by correcting for the deficiencies in the thin screen model as a result of the finite effective path length $\bar{L}$ and the finite bending angle α. For example, the divergence Δ between the radial position of the intersection of the incoming and outgoing asymptotes and the radial position of the point of closest approach by the ray is given by the expression

$$\Delta \doteq a\left[N\times 10^{-6} + \frac{\alpha^2}{8}\right]\Bigg|_{h=h_a} \tag{11.36}$$

which is about 1.9 km for dry air at grazing incidence with the Earth's limb ($a = R_E$) and less than 500 m for closest approach at an altitude of 10 km.

Fresnel Diffraction From the Earth's Limb

The imposition of the Earth's limb itself in the ray path results in a straightedge-like diffraction pattern that is offset and distorted by tropospheric refractivity. (See, for example, Linfield 1992, for a treatment of the Fresnel offset and fringe distortion caused by the Venusian atmosphere in occultation measurements of extragalactic radio sources made by VLBI.) The Earth's diffraction pattern (using the straightedge model) is readily computed from eqs. (11.10) and (11.14) and by using the stationary phase approximation. It is given by

$$\tilde{E}(h_b)\exp\left[i\tilde{\phi}(h_b)\right] = \tag{11.37}$$
$$E_o\exp\left[i\left(\phi_o + \Psi(\bar{h}) + \frac{k}{2D}(\bar{h} - h_b)^2\right)\right]\sqrt{\zeta(\bar{h})}\left[V(v) - iU(v)\right]$$

where $v = \bar{u}/\sqrt{\zeta(\bar{h})}$ and $\bar{u}$ is the stationary phase altitude in Fresnel units and obtained from eq. (11.24); $E_o\exp[i\phi_o]$ provides the observed phase and amplitude unperturbed by the Earth's limb and its atmosphere. Figures 11-7 and 11-8 show the Fresnel fringe patterns in carrier phase and signal amplitude for the vacuum case ($\Psi \equiv 0$), which corresponds to the classic straightedge fringe pattern. The figures also show the fringes for a refracting atmosphere (note the change in the abscissa scale); these figures are offset (because of refractive bending) and distorted (because of the vertical gradient in refractive bending). Here an exponentially distributed atmospheric refractivity was assumed and therefore Ψ is given by eq. (11.14). For the refracting case, the phase delay term $\Psi(\bar{h}) + \frac{k}{2D}(\bar{h} - h_b)^2$ in eq. (11.37) should

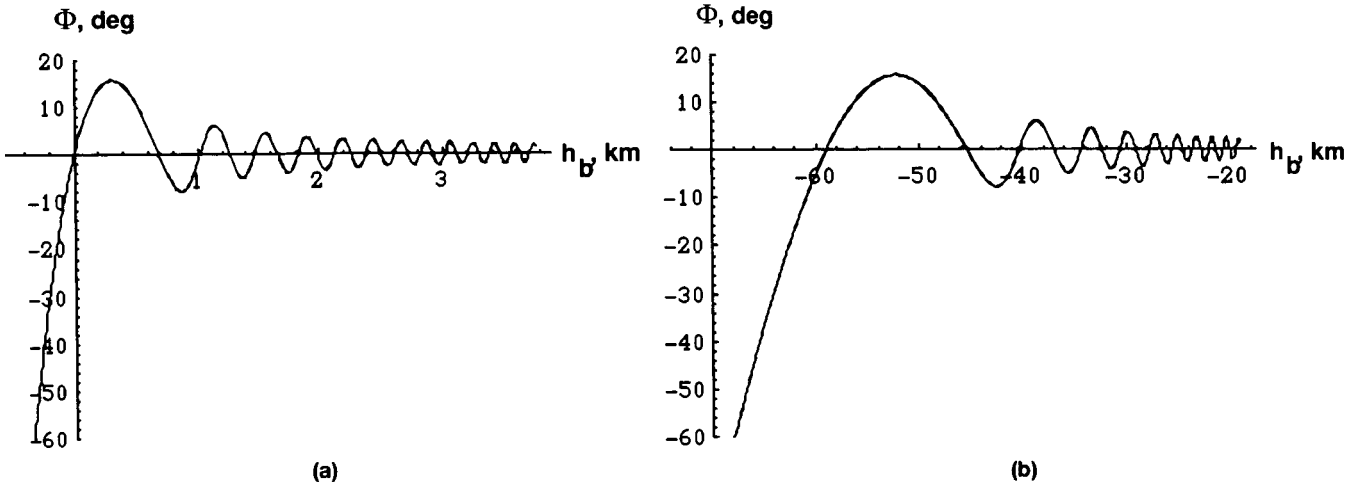


Figure 11-7. Fresnel fringes in carrier phase due to the Earth's limb for two cases: (a) propagation in a vacuum and (b) propagation through an exponentially refracting atmosphere with geometric optics term $2\pi\Delta\rho/\lambda$ suppressed. A scale height of 8 km and a surface refractivity of 260 have been assumed.

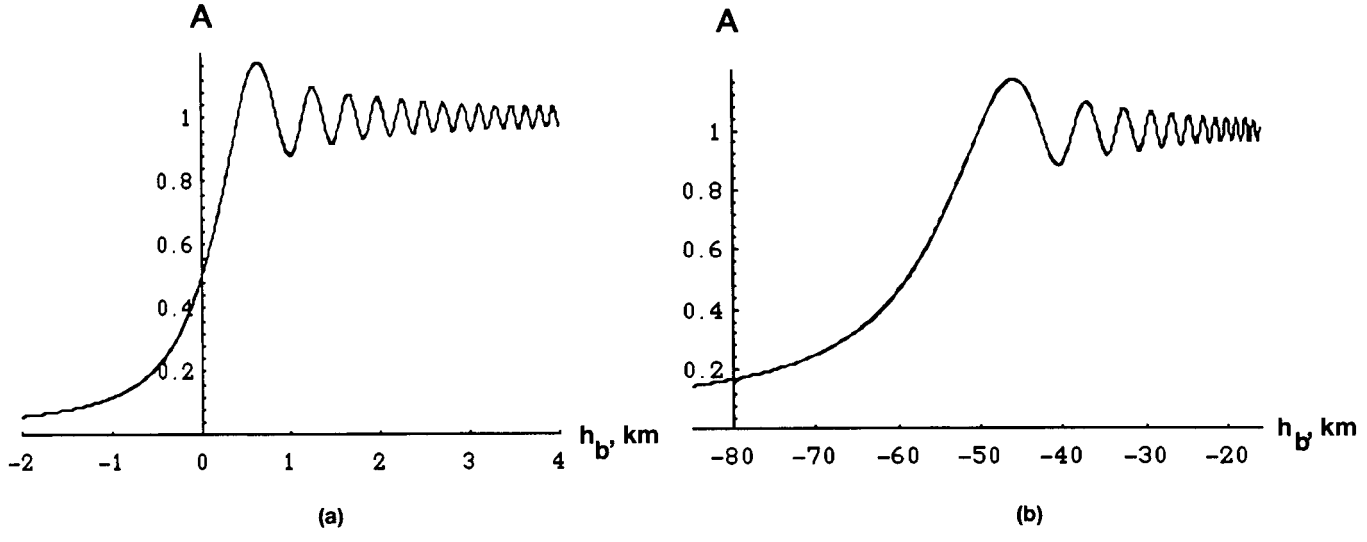


Figure 11-8. Fresnel fringes in signal amplitude due to the Earth's limb for two cases: (a) propagation in a vacuum and (b) propagation through an exponentially refracting atmosphere with the defocusing factor $\sqrt{\zeta}$ suppressed. A scale height of 8 km and a surface refractivity of 260 have been assumed.

be recognized as the same as $2\pi\Delta\rho(h_b)/\lambda$, where $\Delta\rho$ is the residual delay from atmospheric refractivity and is obtained from geometric optics (see eq. (7.27b)). This term has been suppressed in Figure 11-7(b).

Figures 11-7 and 11-8 show that Fresnel diffraction effects from the Earth's limb are quite significant and must be accounted for in the recovery of atmospheric refractivity. The envelopes (and the intervals between zero crossings) of the Fresnel diffracted phase and amplitude variations decrease asymptotically[3] (see footnote next page) only as $1/\bar{h}$. Even for $\bar{h}$ = 5 km (h_b = –26 km), which is nearly 10 Fresnel units above the Earth's limb (23 in integrated local Fresnel units), the amplitude of the Fresnel variation on the phase is

about 0.8 mm (with a local frequency of 5 Hz), which is comparable to the SNR-limited precision of the ionospheric-corrected, double-differenced phase measurement.

Fresnel Transform Techniques

We note that eq. (11.10) (with $G \equiv 1$) involves a well-defined Fourier transform of a modified integrand; therefore, an inverse Fourier transform must exist. During the occultation, if the sampling interval $\Delta\tau$ is sufficiently short so that the displacement between samples $\dot{a}\Delta\tau$ of the stationary phase position in the thin screen is small compared to the radius of the first Fresnel zone, then the phase and amplitude measurements of the GPS signal by the LEO can be considered to be continuous. The quantity $\bar{E}\exp[i\Delta\phi]$ defined by

$$\bar{E}\exp[i\Delta\phi] = \tilde{E}\exp[i\tilde{\phi}] / E_o \exp[i\phi_o] \qquad (11.38)$$

is the perturbation in the actual received signal normalized by the signal that would have been received in the absence of the thin screen and the Earth's opaque limb. Thus, $\bar{E}$ is the normalized variation in amplitude due to the Earth and its atmosphere, and $\Delta\phi$ is the variation in phase.

Then, using eq.(11.9), eq.(11.10) can be rewritten in Fresnel variables as

$$\bar{E}(u_b)\exp[i\Delta\phi(u_b)] = \qquad (11.39)$$
$$\frac{1}{1+i}\int_{-\infty}^{\infty} A(v)\exp[i\Psi(v)]\exp\left[\frac{i\pi}{2}(u_b - v)^2\right] dv$$

Here $u_b = h_b\sqrt{2/\lambda D}$ and $v = h\sqrt{2/\lambda D}$.

Multiplying eq.(11.39) by $\exp\left[-\frac{i\pi}{2}(u_b - v')^2\right]$ and integrating on u_b, we obtain

$$A(v)\exp[i\Psi(v)] = \qquad (11.40)$$
$$\frac{1}{1-i}\int_{-\infty}^{\infty} \bar{E}(u_b)\exp[i\Delta\phi(u_b)]\exp\left[-\frac{i\pi}{2}(u_b - v)^2\right] du_b$$

Thus, eqs. (11.39) and (11.40) form a Fresnel transform pair.

In theory eq. (11.40) is a formal solution, providing $A(v)$ and $\Psi(v)$ in terms of a convolution integral with the observed amplitude and phase perturbation series. In practice, the kernel in eq. (11.40) is slowly converging (a change of integration variable, $u_b - v = \sqrt{t}$, shows that the amplitude of the kernel goes as $1/\sqrt{t}$); so, care must be exercised in establishing practical limits for the integrations. Of course, the boundary conditions: $\bar{E}(u_b)\exp[i\Delta\phi(u_b)] = 1$ for $u_b > u_{max}$, (where u_{max} is the adopted cutoff in altitude), $A(v) = 0$ for $v < 0$ and $\Psi(v) = 0$ for $v > v_{max}$ can be used to preserve accuracy in the integrations.

One can evaluate the quantization error in the recovered products in eq. (11.40) as a result of assuming a continuous measurement sequence for $E(u_b)\exp[i\phi(u_b)]$. Here, one would use a finite sampling approach in conjunction with the Fresnel transform that is analogous to use of the Whittaker–Shannon sampling theorem in Fourier analysis for band-limited signals.

---

[3]The U(v) and V(v) functions can be written in the form

$$U(v) = f(v)\left[\cos\frac{\pi}{2}v^2 + \sin\frac{\pi}{2}v^2\right] - g(v)\left[\cos\frac{\pi}{2}v^2 - \sin\frac{\pi}{2}v^2\right]$$

$$V(v) = 1 - f(v)\left[\cos\frac{\pi}{2}v^2 - \sin\frac{\pi}{2}v^2\right] - g(v)\left[\cos\frac{\pi}{2}v^2 + \sin\frac{\pi}{2}v^2\right]$$

and the asymptotic expansions for f(v) and g(v) are given by

$$f(v) = \frac{1}{2\pi v} - \frac{3}{2\pi^3 v^5} + \cdots,\quad g(v) = \frac{1}{2\pi^2 v^3} - \frac{5}{2\pi^4 v^7} + \cdots$$

Limitations in Resolution

Spatial Band-Limiting

Equation (11.40) would suggest that the resolution of a refracting feature in the thin screen could be made arbitrarily fine, but in actuality that is not the case. To see this we will follow the approach taken in Marouf et al. (1986), where a delta function $\delta(v - v^*)$ is introduced into the integrand of eq. (11.39) in place of $A\exp[i\Psi]$. The impulse response function I is given by

$$I(u_b, v^*) = \frac{1}{1+i}\int_{-\infty}^{\infty}\delta(v-v^*)\exp\left[\frac{i\pi}{2}(u_b-v)^2\right]dv$$

$$= \frac{1}{1+i}\exp\left[\frac{i\pi}{2}(u_b-v^*)^2\right] \qquad (11.41)$$

When $I(u_b, v^*)$ is then inserted into eq. (11.40) in place of $\bar{E}\exp[i\Delta\phi]$ but with a finite integration limit, $\pm U_b$, to reflect the actual spatial limits of the measurement sequence[4], we will not recover the original delta function but rather an approximation to it that is spread as a result of the band-limiting arising from the finite spatial integration limits. The recovered impulse function $R(v,v^*)$ is given by

$$R_{U_b}(v,v^*) = \frac{1}{1-i}\int_{-U_b}^{U_b} I(u_b,v^*)\exp\left[-\frac{i\pi}{2}(u_b-v)^2\right]du_b$$

$$= U_b\frac{\sin[\pi(v-v^*)U_b]}{\pi(v-v^*)U_b}\exp\left[\frac{i\pi}{2}(v^{*2}-v^2)\right] \qquad (11.42)$$

Although several measures of resolution are derivable from this result, a straightforward one is the span across the zero crossings of the first lobe of the sin(x)/x function in eq. (11.42). This quantity is

$$\Delta v_{U_b} = 2/U_b \qquad (11.43)$$

An alternative measure is given by

$$\Delta v = \int_{-\infty}^{\infty}|R(v,v^*)|^2 dv \Big/ |R(v^*,v^*)|^2 \qquad (11.44)$$

For $R(v, v^*)$ given by eq. (11.42), this version of Δv gives just half of the value provided by eq. (11.43), and is therefore a less conservative result.

Equation (11.43) suggests that we can make Δh an arbitrarily small fraction of the radius of the first Fresnel zone just by increasing the integration limit, U_b. But we should keep in mind the various approximations that have been made that can degrade the resolution: the Fresnel approximation itself, the assumption that the distance D is exactly known, the dual-band ionospheric correction, and of course the representation of the actual signal traverse through the real atmosphere by a thin screen. The horizontal location and extent of a refracting feature in the atmosphere can easily be uncertain by 100 km, which amounts to a few percent of D. This error significantly impacts the resolution by altering the phase of the kernel in eq. (11.40); the degradation can be exacerbated for large values of U_b. Any effect that alters the phase of the kernel in an unknown manner can potentially degrade the resolution.

Erroneous Fresnel Scale

To calculate the spreading in resolution that results from an error, $\delta D/D << 1$, in the assumed value of the horizontal distance of the LEO from the thin screen (Figure 11-3), we use the impulse response function in eq. (11.41). Here $I(u_b,v^*)$, which is considered as the true or correct response to the input delta function, is inserted into eq. (11.40), which is modified to reflect the error in D. Thus, the kernel is modified

$$\exp[-i\pi(u_b-v)^2/2] \Rightarrow \exp[-i\pi(u_b-v)^2/2(1+\delta D/D)]$$

[4] The integration limits need not be equally spaced about zero. From Figure 11-7(b), differential refractive bending would suggest that offset limits for U_b might be more appropriate. Here, for simplicity we keep the limits equally spaced.

With this change, we obtain to first order in $\delta D/D$ the recovered impulse function

$$R_{\delta D}(v,v^*) = \frac{1}{2}\int_{-U_b}^{U_b} \exp\left[\frac{i\pi}{2}\left(\gamma^2(u_b - v)^2 + (v + v^* - 2u_b)(v^* - v)\right)\right] du_b \tag{11.45}$$

where $\gamma^2 = \delta D/D$. Using the definition for Δv given by eq. (11.44), it can be shown (assuming that $U_b >> v^*$) that

$$\Delta v_{\delta D} = \frac{\gamma}{1-\gamma^2}\ \frac{x}{\mathscr{C}^2[x] + \mathscr{S}^2[x]} \geq \frac{1.25\gamma}{1-\gamma^2} \tag{11.46}$$

where $x = \gamma U_b$. The function $x/[\mathscr{C}^2(x) + \mathscr{S}^2(x)]$ behaves asymptotically as $1/x$ for small values of x and its mean value goes as $2x$ for large values; it reaches a minimum value of 1.25 at $x = 1.02$. (See Figure 11-9.) Hence, the optimal choice for the integration limits is $U_b = \gamma^{-1}$; it depends inversely on the square root of the error in the assumed value of D. For the resolution, we obtain

$$\Delta h_{\delta D} \geq 1.25\sqrt{\frac{\lambda \delta D}{2}} \tag{11.47}$$

which is at least 125 m for $\delta D = 100$ km. For the effect of other error sources on resolution, such as the Fresnel approximation and the effect of the finite RF bandwidth, the reader should consult Marouf et al. (1986).

Stochastic Considerations

We can also use eq. (11.40) to evaluate the expected error and the statistical correlation of the recovered values of A(v) and $\Psi(v)$ resulting from thermal noise in the phase and amplitude measurements. Let the outputs from the sine and cosine correlators in the GPS receiver be designated by I and Q, respectively. Then $\hat{E}$ and $\hat{\phi}$, the measurements of E and ϕ, respectively, are given by

$$\hat{E} = \left[I^2 + Q^2\right]^{1/2} \tag{11.48}$$

$$\tan\left[\hat{\phi} - \phi_o\right] = I/Q \tag{11.49}$$

where ϕ_o is the reference phase from the in-receiver model. For band-limited (approximately 10-MHz, single-sided) Gaussian white noise, the errors in I and Q are uncorrelated. Defining $\delta\hat{E}$ and $\delta\hat{\phi}$ to be the resulting errors in the measurements, it follows from eqs. (11.48) and (11.49) that the expected values

$$\left\langle \delta\hat{E}\delta\hat{\phi} \right\rangle = 0 \tag{11.50}$$

and that

$$\left\langle \delta\hat{E}^2 \right\rangle = \left\langle \delta\hat{\phi}^2 \right\rangle E^2 = W\sigma^2 / \tau \triangleq \hat{\sigma}^2 \tag{11.51}$$

where W is the single-sided bandwidth ($\times 2\pi$) of the RF spectrum, σ is the noise spectral density, and τ is the correlator averaging time.[5] We can also treat the errors in the measurement sequences, $\delta\hat{E}$ and $\delta\hat{\phi}$, as white-noise processes because $\tau >> 2\pi/W$. From eq. (11.40) it follows that $\hat{A}(v)$ and $\hat{\Psi}(v)$ are given by

$$\begin{Bmatrix} \hat{A}(v)\sin\hat{\Psi}(v) \\ \hat{A}(v)\cos\hat{\Psi}(v) \end{Bmatrix} = \frac{1}{2}\int_{-U_b}^{U_b} \overline{E}(u_b)\{\cos[\bullet] \pm \sin[\bullet]\} du_b \tag{11.52}$$

where

$$[\bullet] = \Delta\varphi - \frac{\pi}{2}(u_b - v)^2 \tag{11.53}$$

Taking the variation of $\hat{\Psi}$ in terms of the $\delta\hat{E}$ and $\delta(\Delta\phi)$ sequences in the integrand of eq. (11.52), and then squaring and taking the expected value, we obtain for $\delta\hat{\Psi}(v)$

[5] Here, σ should be multiplied by ~3 to compensate for the ionospheric delay correction, and by 2 to account for double differencing.

$$\left\langle \delta\hat{\Psi}^2 \right\rangle = \frac{4U_b\hat{\sigma}^2 / E_o^2}{\left[\int_{-U_b}^{U_b} \overline{E}(u_b)\{\cos[\bullet] - \sin[\bullet]\}du_b\right]^2 + \left[\int_{-U_b}^{U_b} \overline{E}(u_b)\{\cos[\bullet] + \sin[\bullet]\}du_b\right]^2} \tag{11.54}$$

If $\overline{E}$ is set equal to 1 and use is made of the stationary phase approximation for the arguments of the sine and cosine functions, we can evaluate eq. (11.54). Here we expand the argument $[\bullet]$ about the stationary phase value of $u_b, \bar{u}_b(v) = v + \Psi'(v)/\pi$ (see eq. (11.12))[6] to obtain

$$[\bullet] = \Psi(v) - \frac{\pi}{2}\zeta(v)\left[u_b - \bar{u}_b\right]^2 + O\left[u_b - \bar{u}_b\right]^3 \tag{11.55}$$

where $\zeta(v)$ is the defocusing factor and given by

$$\zeta(v) = \frac{1}{1 + \pi^{-1}\Psi''} \tag{11.56}$$

where $\Psi''(v)$ is the second derivative of Ψ with respect to v. For an exponential atmosphere Ψ is given by eq. (11.14). It can be shown using the stationary phase approximation that eq. (11.54) becomes

$$\left\langle \delta\hat{\Psi}^2(v) \right\rangle = \frac{\hat{\sigma}^2}{2E_o^2}\sqrt{\zeta(v)}\ \Gamma(X,Y) \tag{11.57}$$

where $\Gamma(X,Y)$ is given by

$$\Gamma(X,Y) = \frac{4X}{\left[\{\mathscr{C}(X+Y) + \mathscr{C}(X-Y)\}^2 + \{\mathscr{S}(X+Y) + \mathscr{S}(X-Y)\}^2\right]} \tag{11.58}$$

and where

$$X(v) = U_b\sqrt{\zeta(v)} \tag{11.59}$$

and

$$Y(v) = \bar{u}_b(v)\sqrt{\zeta(v)} \tag{11.60}$$

The function $\Gamma(X,Y)$ is shown in Figure 11-9 parametrically in Y. When Y = 0 we obtain the same function that arose earlier in the discussion of resolution limits resulting from an erroneous adopted value for the Fresnel scale.

The function $\Gamma(X,0)$ also applies when the integration limits in eq. (11.54) are equally spaced about Y(v). Here, X is reinterpreted as the one-sided spacing about Y(v). Figure 11-9 shows that opting for an overly large (or small) value for the integration limits can be too much (or too little) of a good thing; the "optimal" spacing about Y(v) is 1. However, choosing this value for X would compromise the resolution. We note from Figure 11-9 that a kind of uncertainty principle applies to the band-limited resolution Δv_{U_b} and to $\left\langle \delta\hat{\Psi}^2 \right\rangle$ when $X > 1$: the larger X is, the better the resolution, but with poorer accuracy (on average) in the recovery of Ψ. For small X ($X<1$)

$$\left\langle \delta\hat{\Psi}^2(v) \right\rangle \doteq \frac{\hat{\sigma}^2}{2XE_o^2}\sqrt{\zeta(v)} \geq \frac{1.25\hat{\sigma}^2}{2E_o^2}\sqrt{\zeta(v)} \tag{11.61}$$

Thus, we gain both accuracy and resolution for small X by making the spacing of the integration limits in altitude about the point of stationary phase at least as wide as $2/\sqrt{\zeta(v)}$ times the Fresnel scale $\sqrt{(\lambda D/2)}$. Choosing a value for X greater than 1 might be desirable to achieve finer resolution at the expense of some accuracy, but our choice should be moderated by the uncer-

[6] By reciprocity, we also have, using eq. (11.12), the stationary phase condition $\partial\varphi/\partial h_b = 0$, (see eq. (7.15)). From this it follows that the condition $d\varphi_o/dh_b = d\psi/dh$ must hold on a path of a stationary phase.

tainty in the horizontal distance of the refracting feature (through eq. (11.46)), by the effect of unknown gradients in ionospheric electron density (see eq. (11.75)) and by other approximations in this thin screen Fresnel analysis that limit resolution.

From eq. (11.52), we can also derive the autocorrelation function for $\hat{\Psi}$, which is defined by

$$\rho_{\Psi} = \frac{\left\langle \delta\hat{\Psi}(v)\delta\hat{\Psi}(v^*)\right\rangle}{\sqrt{\left\langle \delta\hat{\Psi}^2(v)\right\rangle\left\langle \delta\hat{\Psi}^2(v^*)\right\rangle}} \qquad (11.62)$$

It can be shown that ρ_{Ψ} is given by

$$\rho_{\Psi} = \frac{\sin[\pi U_b(v-v^*)]}{\pi U_b(v-v^*)}\cos\frac{\pi}{2}\left(v^{*2}-v^2\right) \qquad (11.63)$$

Thus, the estimates of Ψ decorrelate for $v-v^* >> 1/U_b$. Equation (11.63) should be recognized as the real part ($\div U_b$) of the band-limited resolution function $R_{U_b}(v, v^*)$ given in eq. (11.42).

Finally, using the correspondence between $\Psi(v)$ in the thin screen model and the refractivity in the actual atmosphere, as given by eq. (11.3), we can obtain $\left\langle \delta\hat{N}^2\right\rangle$. Here, the thermal noise-to-signal ratio $\hat{\sigma}/\hat{E}$ corresponds to $2\pi\sigma_\rho/\lambda$, where σ_ρ is the ionosphere-corrected and double-differenced phase measurement in range units. For two successive determinations of Ψ separated by a short-time $\Delta\tau$ (but with $\Delta\tau >> 2\pi/W$), $\Delta\Psi$ may be expressed in terms of N by

$$\Delta\Psi \doteq \frac{4\pi}{\lambda}\sqrt{2a\dot{a}\Delta\tau}N(v)10^{-6} \qquad (11.64)$$

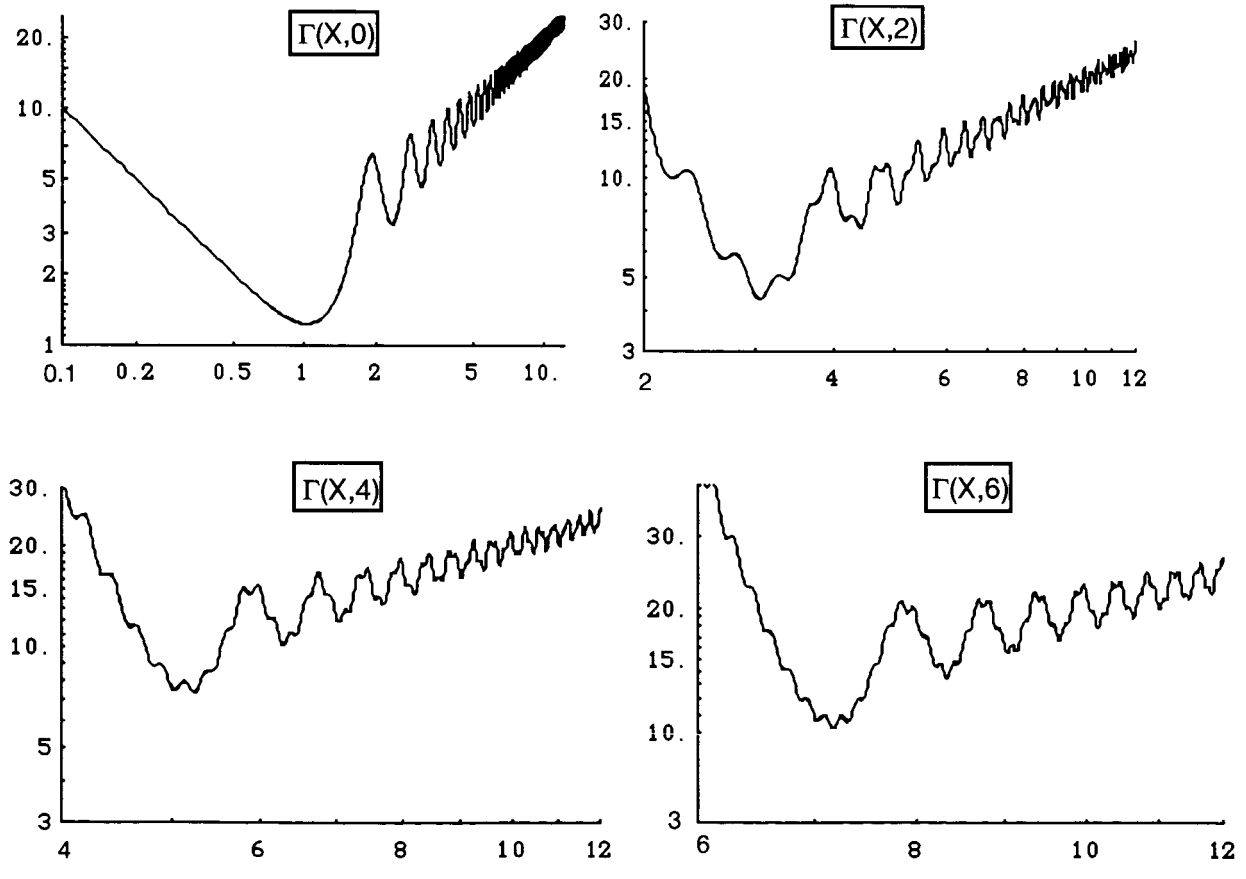


Figure 11-9. The function Γ(X,Y), from eq. (11.57), versus X.

where a is the impact parameter (~6400 km) and $\dot{a} = \dot{b}\zeta(v) \approx 3\zeta(v)$ km/s. It follows that

$$\left\langle \delta\hat{N}^2(v) \right\rangle^{1/2} = 10^6 \sigma_\rho \sqrt{\frac{\Gamma(X,Y)}{8a\dot{b}\Delta\tau\sqrt{\zeta(v)}}}$$
$$\approx 2\times 10^{-3} \sqrt{\frac{\Gamma(X,Y)}{\Delta\tau\sqrt{\zeta(v)}}} \qquad (11.65)$$

where the last relation applies when σ_ρ = 1 mm. Then, with the result from eq. (11.65) and assuming that the errors in the recovered refractivity form a white-noise sequence, we can use the same procedure starting from eq. (8.13) to obtain a similar expression for $\left\langle \delta\hat{T}^2(v) \right\rangle$. We will leave it to the reader to show that $\left\langle \delta\hat{T}^2(v) \right\rangle^{1/2}$ obtained by this Fresnel-diffraction/thin screen approach (recalling that $\sigma_\rho = \sigma_{\rho_o}\sqrt{1\,\mathrm{sec}/\Delta\tau}$) yields essentially the same SNR-based accuracy profile as shown in Figure 8-3.

Dealing With the Ionosphere

Because the correction for ionospheric delay involves two RF carriers with different Fresnel scales, the accuracy and resolution of refracting features in the atmosphere will be adversely impacted. Let $\Delta\phi_k$ be the observed carrier phase residual from the *k*th carrier, and let $\Delta\phi_k^T$ be the true phase that would have been observed in the absence of an intervening ionosphere. Here, thermal errors are omitted from discussion. It follows (see eqs. (8.67) and (8.68)) that the linear combination of the dual-band phase data that eliminates the first-order ionospheric delay term (and ignoring the higher order terms) yields, in the Fresnel case, the condition

$$f_2\Delta\phi_2 - f_1\Delta\phi_1 = f_2\Delta\phi_2^T - f_1\Delta\phi_1^T \qquad (11.66)$$

where f_1/f_2 = 77/60, the ratio of the L1 and L2 carrier frequencies. In a geometric optics context, $\Delta\phi_1^T / f_1$ will equal $\Delta\phi_2^T / f_2$ and eq. (11.66) provides a means to recover an ionosphere-free phase measurement. In the Fresnel diffraction approach, $\Delta\phi_1^T / f_1$ and $\Delta\phi_2^T / f_2$ will differ because their convolution integrals in eq. (11.39) involve Fresnel scales that differ by 13%. Referring to Figure 11-3, let N_e be the integrated columnar electron content along a ray path between the LEO and the GPS satellite. In general N_e will be a function of h and h_b. Suppose N_e can be separated into two functions, $N_e(h,h_b) = \upsilon(h) + \eta(h_b)$, which might be a good representation considering the large distances of the LEO and the GPS satellite from the Earth's neutral atmosphere and the short duration of the occultation. Here, $\upsilon(h)$, which can be considered as the cumulative effect of the ionosphere along the ray path between the GPS satellite and the Earth's atmosphere, and therefore behind the thin screen, can be lumped into the thin screen phase delay $\Psi(h)$. Thus,

$$\Psi_k(h) \Rightarrow \Psi(h) - q\upsilon(h)f_k^{-1} \qquad (11.67)$$

where $q = 8.58 \times 10^{-7}$ and the units of $\upsilon(h)$ are electrons/m^2. For this component of the ionosphere, we would carry out the convolution integral in eq. (11.40) separately for L1 and L2 before applying the two-frequency correction. From the recovered $\hat{\Psi}_1$ and $\hat{\Psi}_2$ sequences, we would then form

$$\hat{\Psi}_3 = \left[f_2\hat{\Psi}_2 - f_1\hat{\Psi}_1\right]/(f_2 - f_1) \qquad (11.68)$$

which would be free of the $\upsilon(h)$ term.

Dealing with the $\eta(h_b)$ term is more problematic. This term is dependent on the position of the LEO and is therefore embedded in the convolution integral in eq. (11.40). In this case eq. (11.40) becomes

$$\hat{A}_k(h)\exp\left[i\hat{\Psi}_k(h)\right] =$$

$$\frac{1}{1-i}\sqrt{\frac{2}{\lambda_k D}}\int_{-\infty}^{\infty}\bar{E}\exp[i(\Delta\phi_k^T - q\eta/f_k) - \frac{i\pi}{\lambda_k D}(h_b - h)^2]dh_b \tag{11.69}$$

To evaluate this integral, we expand $\eta(h_b)$ in a power series about the ionosphere-free stationary phase value for h_b, $\bar{h}_b(h)$, (see "Stationary Phase Path" earlier in this chapter) and we use the stationary phase approximation for the integral. Then

$$\eta(h_b) =$$

$$\eta(\bar{h}_b) + \eta'(\bar{h}_b)(h_b - \bar{h}_b) + \frac{1}{2}\eta''(\bar{h}_b)(h_b - \bar{h}_b)^2 + \cdots \tag{11.70}$$

Letting A = 1, and carrying out the integration in eq. (11.69), for the change in $\hat{\Psi}_k(h)$ due to the presence of the $\eta(h_b)$ term, we obtain

$$f_k\delta\hat{\Psi}_k(h) = -q\eta(\bar{h}_b) + \frac{q^2\lambda_k D\eta'^2}{4\pi\zeta(\bar{h}_b)f_k - 2q\lambda_k D\eta''} \tag{11.71}$$

where ζ is the defocusing factor. When the η'' can be neglected, the error in $\hat{\Psi}_3$ becomes

$$\delta\hat{\Psi}_3 = \frac{q^2\eta'^2}{2\pi c\zeta(\bar{h}_b)}\sqrt{\frac{\lambda_1 D}{2}}\sqrt{\frac{\lambda_2 D}{2}}\left(\frac{f_1+f_2}{f_1 f_2}\right) \approx \frac{3.9\times10^{-32}}{\zeta(\bar{h}_b)}\eta'^2(\bar{h}_b) \tag{11.72}$$

where in the last expression the units of η' are electrons/m^{2}km. For $\eta' = 10^{15}$, the resulting phase error is comparable to the SNR error in $\hat{\Psi}_3$.

The resolution is also degraded by the presence of the $\eta(h_b)$ term. To calculate its effect we will again use the impulse response function $I(u_b,v^*)$ given by eq. (11.41). However, before inserting it into eq. (11.42), we will corrupt it by advancing its phase to account for the $\eta(h_b)$ term. Thus

$$I(u_b,v^*) \Rightarrow I(u_b,v^*)\exp[-i\varepsilon(u_b)] \tag{11.73}$$

where

$$\varepsilon(u_b) = q\eta(u_b)/f_k \tag{11.74}$$

and where $u_b = h_b\sqrt{(2/\lambda_k D)}$. Also the integration limits in eq. (11.42) are changed from $\pm U_b$ to $\bar{u}_b(v) \pm \Omega$, and therefore equally spaced about the point of stationary phase (when $\varepsilon \equiv 0$). To evaluate eq. (11.42), ε is expanded about $\bar{u}_b(v)$ as given in eq. (11.70). For the case where $\Omega << |2\varepsilon'/\varepsilon''|$, or equivalently, $|\varepsilon''| << \varepsilon'^2$, the quadratic term can be neglected. Using the definition for Δv given by eq. (11.44) and inserting the modified eq. (11.42) into eq. (11.44), we obtain after carrying out the integrations[7]

$$\Delta v = \left|\varepsilon'\left(\bar{u}_b\right)\right|\frac{x}{\sin^2 x} \tag{11.75}$$

where $x = |\varepsilon'|\Omega$. The minimum value of $x/\sin^2 x$ is 1.38 at $x = 1.17$. Thus, the optimal value of Ω is $1.17/\left|\varepsilon'\left(\bar{u}_b\right)\right|$ and

$$\Delta h \geq 1.38\left|\frac{d\varepsilon}{dh_b}\right|\frac{\lambda_k D}{2} \tag{11.76}$$

---

[7] A heuristic but more intuitive method of determining the magnitude of Δv, which avoids the above-cited integrations required by eq. (11.44), is based on the idea of "loss of coherence." It is well known in correlator operations that little or no additional information is gained by extending the integration limits past the point where the operands in the convolution integral lose coherence. Let U_b^* be the integration point where the unknown variation in phase error in the integrand of eq. (11.40) has grown to $\pi/2$ rad. Then U_b^* is effectively the spatial band limit and we set $\Delta v = 2/U_b^*$. For the ionosphere error in eq. (11.69), we obtain $\Delta v = 4|\varepsilon'|/\pi$. For the error in the Fresnel scale $\gamma 2$, we obtain $\Delta v = 2\gamma$. For the Fresnel approximation itself, we obtain $\Delta v = [2\lambda/D]^{1/4}$.

Equation (11.76) predicts that an unknown transverse gradient in N_e on the segment of the ray path between the LEO and the Earth's atmosphere of 10^{15} electron/m^{2}km (see Figure 8-16) would degrade the resolution to about 225 m on the L1 carrier. The optimal integration span in this case would be about 4 Fresnel scales centered at the stationary phase point.

When the ε'' term has to be taken into account, it can be shown that

$$\Delta v = \sqrt{\frac{|\varepsilon''(\bar{u}_b)|}{\pi}}\Gamma(X,Y) \qquad (11.77)$$

where

$$X = \sqrt{\frac{|\varepsilon''(\bar{u}_b)|}{\pi}}\Omega, \quad Y = \sqrt{\frac{|\varepsilon''(\bar{u}_b)|}{\pi}}\frac{\varepsilon'(\bar{u}_b)}{\varepsilon''(\bar{u}_b)} \qquad (11.78)$$

and where $\Gamma(X,Y)$ is defined by eq. (11.58).

Correspondence Between the Thin Screen Model and the Real Atmosphere

The total delay measured by the LEO arises from a geometric bending term, $\sim D\alpha^2/2$ (see eq. (7.27b), and an additional atmospheric delay term resulting from the reduced speed of light in the atmosphere. The latter term corresponds to the phase delay in the thin screen, $\psi(h_a)$, which is given by eq. (11.3). From eq. (7.17), it follows that the correspondence between $\psi(h_a)$ and the atmospheric delay term is

$$\frac{\lambda}{2\pi}\psi(h_a) \leftrightarrow -2\int_{r_o}^{\infty}\frac{n'}{n}\sqrt{n^2r^2 - a^2}\,dr \qquad (11.79)$$

The proof of eq. (11.79) follows from differentiating both sides with respect to a and using eq. (7.20) and the stationary phase condition in eq. (11.13). It follows from the Abel transform that

$$\ln[n(a)] \leftrightarrow \frac{-\lambda}{2\pi^2}\int_a^{\infty}\frac{\psi'(\xi)}{\sqrt{\xi^2 - a^2}}d\xi$$

$$= \frac{\lambda}{2\pi^2}\frac{d}{da}\left[\int_a^{\infty}\frac{a\psi(\xi)}{\xi\sqrt{\xi^2 - a^2}}d\xi\right] \qquad (11.80)$$

where $\xi = nr$. Equation (11.80) enables us to transfer, at least qualitatively, the Fresnel deconvolution properties derived from the thin screen model to the recovered refractivity series.

The Height of the Tropopause

The height of the tropopause is an important parameter in the study of global climate change (see Chapter 3). Long-term monitoring of globally distributed tropopausal height and temperature variations should be valuable for studies of possible anthropogenic global warming. This region is characterized by a local minimum in the temperature profile and, in some cases, an abrupt change in lapse rate. The discussion earlier in this chapter suggests that Fresnel deconvolution techniques can achieve vertical resolutions of the recovered atmospheric refractivity products that are significantly sharper than the diameter of the first Fresnel zone, perhaps as small as 10%. We can estimate the accuracy with which the height of the tropopause can be determined from the series of recovered refractivity and/or temperature profiles.

Using the gas law and the assumption of hydrostatic equilibrium (see Chapter 7), it can be shown (for the case of dry air only) that

$$\ln\left[\frac{NT}{N_oT_o}\right] = -\frac{1}{H_o}\int_{h_o}^{h}\frac{T_o}{T}\,dh \qquad (11.81)$$

where H_o is the local scale height of the refractivity at h_o. At tropopausal altitudes, the temperature is sufficiently cold ($T_o \approx$ 210–220 K) that

virtually all water vapor has been wrung out; the residual relative abundance at these altitudes is less than 0.03% (in number density) on average. Therefore, water vapor contributes less than 0.5% to the total refractivity.

For the purpose of estimating the accuracy with which h_0 can be determined, the temperature profile in the vicinity of the tropopause can be modeled, for example, by a discontinuous lapse rate, or by a quadratic model, such as

$$T = T_0 + \frac{1}{2}T''(h_0)(h - h_0)^2 \qquad (11.82)$$

where h_0 is the height of the tropopause, which is defined in this model by the condition $T'(h_0) = 0$. Although Figure 2-1 is an idealized representation of the temperature profile, it gives a rough magnitude for T" of about 1 K/km^2. The expression in eq. (11.81) can be integrated to yield

$$N(h) = N_0 \frac{T_0}{T(h)} \exp\left[-\sqrt{2T_0/T''H_o^2}\,\tan^{-1}\left[\sqrt{T''/2T_0}(h - h_0)\right]\right] \qquad (11.83)$$

Equation (11.83) can be expanded to yield, since $(T''/2T_0)(h - h_0)^2 << 1$ near the tropopause,

$$N(h) \doteq N_0 \exp\left[-\frac{h - h_0}{H_0}\right]\exp\left[-\frac{T''}{2T_0}(h - h_0)^2\right] + O(h - h_0)^3 \qquad (11.84)$$

The unknowns, N_0, H_0 (or T_0), $T''/2T_0$, and h_0, are to be determined in a least-squares fit to the time series $\hat{N}(h)$, which is the recovered profile of the refractivity.

Although calculating the covariance matrix from the variations of N(h) (given by eq. (11.84)) with respect to the unknowns is straightforward, the recovered profile of the temperature provides a simpler way to obtain an estimate of the accuracy of the recovery of h_0. From Chapter 8, eqs. (8.14) through (8.17), it follows that the thermal errors on the temperature series are generated by an integrated white-noise stochastic process. If we perform a minimum variance least-squares fit of dT/dh to the recovered temperature series over a spanned interval $h_2 - h_1$ about h_0, it can be shown that the accuracy of the determination of h_0 is given by

$$\sigma_{h_o} = \frac{6\sqrt{6}}{\pi}\frac{2T_0}{T''}\frac{10^6}{N(h_0)}\sqrt{\frac{\zeta_o}{a_0(h_2 - h_1)}}\frac{1}{\dot{a}_0\Delta\tau}\sqrt{\frac{1\,\text{sec}}{\Delta\tau}}\sigma_{\rho_o} \qquad (11.85)$$

where σ_{ρ_o} is the standard deviation of the thermal error in the L1 carrier phase measurement over 1 s and $\Delta\tau$ is the temporal sampling interval. For a least-squares span of 10 km and a sampling interval of 1 km, eq. (11.85) predicts an accuracy of roughly 100 m for the determination of h_0, which is well within the first Fresnel zone.

We can also estimate the Fresnel diffraction effects that would result from an assumed temperature variation in the vicinity of the tropopause. The phase perturbation term ε in the thin screen model (see eq. (11.11)) would be given by

$$\varepsilon(h, h_0) = \begin{cases} k\int_h^{h_0} \delta\alpha(r)\,dh, & h \le h_0 \\ 0, & h \ge h_0 \end{cases} \qquad (11.86)$$

where $\delta\alpha(r)$ is the departure of the bending angle profile from its nominal value due to the change in the temperature profile at the tropopause. The reference phase $\psi(h)$ in the screen model would be obtained from the nominal bending angle profile (see eq. (11.14), for example). If $\psi + \varepsilon$ is inserted into eq. (11.39) in place of Ψ, we obtain a Fresnel diffraction pattern that is a function of the height h_b of the LEO, the altitude h_0 of the tropopause, and a function of the parameters in the adopted model for the temperature profile. We wish to determine the Fresnel pattern that results from the perturbation in the temperature profile and to use it to recover the values of these parameters.

A simple approach is to consider the tropopause as marked by a discontinuity in the lapse rate. Then for $h < h_0$, $\delta\alpha(h)$ is given by

$$\delta\alpha(h) = \alpha^-(r) - \alpha^+(r) - 2a^- \int_{r_0}^{\infty} \left[\frac{dn^+/dr'}{n^+\sqrt{(n^+r')^2 - (a^-)^2}} - \frac{dn^-/dr'}{n^-\sqrt{(n^-r')^2 - (a^-)^2}} \right] dr' \quad (11.87)$$

where $a^- = rn^-(r)$; the + or - signs on n(r) denote the refractivity profile that corresponds to the lapse rate regime above or below the altitude h_0; $\alpha^{\pm}(r)$ is the bending that would be obtained if the corresponding refractivity profile prevailed over the entire atmosphere, and $\alpha^+(r)$ is the nominal bending angle profile. Therefore, $\delta\alpha(r) = 0$, $r \geq r_o$. The details for calculating $\delta\alpha(r)$ from eq. (11.88) for the case of a discontinuous lapse rate will be found in Appendix E. The free parameters in this model are $\Delta\gamma$, which is the discontinuity in lapse rate, and h_0. Using the stationary phase approximation it can be shown from eq. (11.39) that $\tilde{E}\exp[i\delta\phi]$, the thin screen Fresnel diffraction pattern due to $\delta\alpha$ and ε, is given by

$$\tilde{E}\exp[i\delta\phi] \doteq \sqrt{\frac{\zeta^-(u^-)}{\zeta^+(u^+)}} \exp\left[i\Delta^-\right] \left[V(w^-) - iU(w^-)\right] + V(-w^+) - iU(-w^+), \quad u_b \leq u_b^* \quad (11.88)$$

where U(w) and V(w) are the Fresnel integral forms defined in eq. (11.27); their argument w is defined by

$$w^{\pm} = (u_0 - \bar{u}^{\pm})/\sqrt{\zeta^{\pm}(\bar{u}^{\pm})} \quad (11.89)$$

and where the altitudes of the stationary phase paths (see eqs. (11.12) and (11.13)) are given by

$$\bar{u}^{\pm} = \begin{cases} \bar{u}^+(u_b) = u_b + D'\{\alpha^+[\bar{u}^+(u_b)]\} \\ \bar{u}^-(u_b) = u_b + D'\{\alpha^+[\bar{u}^-(u_b)] + \delta\alpha[\bar{u}^-(u_b)]\} \end{cases} \quad (11.90)$$

Here $D' = D\sqrt{2/\lambda D}$. The quantity $\zeta^{\pm}$ is the defocusing factor

$$\zeta^{\pm} = \begin{cases} \zeta^+ = \left[1 - D'\,d\alpha^+/du\right]^{-1} \\ \zeta^- = \left[1 - D'\,d(\alpha^+ + \delta\alpha)/du\right]^{-1} \end{cases} \quad (11.91)$$

The quantity Δ^- is the geometric optics phase delay due to $\delta a(r)$ and $\varepsilon(h)$; it is given by

$$\Delta^- = \varepsilon(\bar{u}^-) + kD\zeta^+(\bar{u}^-)\delta\alpha^2(\bar{u}^-)/2 + O[\delta\alpha^3] \quad (11.92)$$

There is a caustic near r_0 when $\Delta\gamma > 0$, which results in multiple stationary phase points below r_0. The quantity u_b^* marks the point in u_b-space where $\zeta^-[\bar{u}^-(u_b)] \to \infty$; above this point there are no stationary phase altitudes below u_0. Similarly, u_b^o, where $\bar{u}^+ = u_0$, marks the point below which there are no stationary phase points above u_0. The range $u_b^o < u_b < u_b^*$ defines a caustic zone where two stationary phase points may occur below u_0 and one above. Eq. (11.39) should be evaluated by using all three points; accordingly, eq. (11.88) is still valid in this zone but it should have a third term appended, which is evaluated at the extra stationary phase point.

For $u_b \geq u_b^*$ eq. (11.39) becomes

$$\tilde{E}\exp[i\delta\phi] \doteq \sqrt{\frac{\zeta^-(u_o)}{\zeta^+(u_o)}} \exp[i\Delta_0] \left[V(w_o^-) - iU(w_o^-)\right] + V(-w^+) - iU(-w^+), \quad u_b \geq u_b^* \quad (11.93)$$

where

$$w_o^- = w^+\sqrt{\zeta^-(u_o)/\zeta^+(\bar{u}^+)} \quad (11.94)$$

and where the phase delay term Δ_0 is given by

$$\Delta_0 = \frac{\pi}{2}\left[(w^+)^2 - (w_o^-)^2\right] + O[w^3] \quad (11.95)$$

As u_b increases above u_b^* the Fresnel forms $V(w_o^-)$ and $U(w_o^-)$ attenuate rapidly; therefore, the quadratic approximation in eq. (11.95) should suffice. Whenever $\Delta\gamma$ is large enough (> ~1 K/km) to cause the geometric optics delay term Δ^- in eq. (11.89) to grow by roughly $\pi/2$ radians over a distance of 1 local Fresnel scale, significant Fresnel modulations in amplitude and phase will result. These are shown in Figures E-3 and E-4 for a range of values of $\Delta\gamma$.

CHAPTER

12 SPECIAL TOPIC: RECEIVER OPERATIONS IN THE LOWER TROPOSPHERE

Obtaining high-precision phase measurements through the lower troposphere is difficult because of three factors. First, the signal amplitude is greatly reduced due to defocusing. Second, the signal experiences significant acceleration in this region. Finally, multipath in the lower troposphere results in amplitude and phase scintillations.

One of the approaches that is being considered and would allow tracking through this region involves having the receiver carrier tracking loop transfer from a data-residual driven loop to a data-plus-model driven loop. In other words, when the minimum altitude of the ray reaches a predetermined lower level, or the SNR reaches a lower limit, the conventional tracking loop will be supplemented by a model based on an expected variation of the integrated index of refraction with time. This model would have its phase, phase-rate, and phase-acceleration initialized by the values in the conventional loop. Additional lag channels would be used to receive multipath signals with more than one P-chip of delay. The sample rate would also be increased to capture the full expected range of frequencies contained in the scintillation spectrum. These samples could then be analyzed by post-processing, where multiple passes through the data could be used to search for signals.

The approach described above is obviously only applicable to setting occultations. In the case of rising occultations, we would have to begin with a model loop using still higher sample rates to accommodate the uncertainty in a model based on predictions of the signal Doppler and phase acceleration, until the signal had been acquired by a tracking loop.

This approach will help track lower SNR signals because the final data can be extracted in post-processing. There is no SNR limit below which data cannot be gathered; that is, a complete record of the bandwidth being observed is obtained independent from the ability of the signal to track the low SNR signal. The high accelerations will be accommodated through the use of models to aid the tracking loop. Finally, the simultaneous acquisition of data from multiple signals originating from multipath conditions is accomplished by increasing the output data rate.

CHAPTER

13 REFERENCES

Ahnert, P. R. (1991), "Precision and compatibility of National Weather Service upper air measurements," *Preprints, Seventh Symp. on Meteorological Observations and Instrumentation,* New Orleans, Louisiana, 221–227.

Anderson, D. N. (1993), "The development of global, semi-empirical ionospheric specification models," *Proc. of the 7th Intl. Ionospheric Effects Symp.*, J. Goodman, ed., Alexandria, Virginia.

Bassiri, S. and G. A. Hajj (1993), "Higher-order ionospheric effects on the GPS observables and means of modeling them," *Manuscripta Geodætica*, 18:280–289, 1993.

Bauer, P. (1992), *GLIMPSE progress report*, Status report submitted to CNES Headquarters, Paris, France, January.

Blewitt, G., M. B. Heflin, F. H. Webb, and J. F. Zumberge (1992), "Global coordinates with centimeter accuracy in the International Terrestrial Reference Frame using the Global Positioning System," *Geophys. Res. Lett.*, **19**, 853–856.

Born, M. and E. Wolf (1980), *Principles of Optics*, Sixth Edition, Pergamon Press, Oxford.

Brunner, F. K. and M. Gu (1991), "An improved model for the dual frequency ionospheric correction of GPS observations," *Manuscripta Geodætica,* **16**.

Champion, K. S. W., A. E. Cole, and A. J. Kantor (1985), "Standard and reference atmospheres" in the *Handbook of Geophysics and the Space Environment*, A. S. Jursa (ed.), Air Force Geophysics Laboratory, Bedford, Massachusetts.

Chandrasekhar, S. (1952), "A statistical basis for the theory of stellar scintillation," *M.N.,* **112**, 475–483.

Chiu, Y. T., K. R. Hardy, G. L. Tyler, and D. P. Hinson (1991), "Sensing tropospheric temperature and ionospheric structures with space-borne GPS receivers," *Proc. U.S–Taiwan Workshop on Space*, Taipei, Taiwan, April 15–18.

Eshleman, V. R. (1973), "The radio occultation method for the study of planetary atmospheres," *Planet. Space Sci.*, **21**, 1521–1531.

Finger, F. G. and F. J. Schmidlin (1991), "Upper-air measurements and instrumentation workshop," *Bull. Amer. Meteor. Soc.*, **72**, 50–55.

Fjeldbo, G. and V. R. Eshleman (1968), "The atmosphere of Mars analyzed by integral inversion of the Mariner IV occultation data," *Planet. Space Sci.*, **16**, 1035–1059.

Fjeldbo, G. (1964), *Bistatic-radar methods for studying planetary ionospheres and surfaces,* Sci. Rpt. No. 2, NsG-377, SU-SEL-64-025, Stanford Electronics Laboratories, Stanford, California.

Fjeldbo, G., A. J. Kliore, and V. R. Eshleman (1971), "The neutral atmosphere of Venus as studied with the Mariner V radio occultation experiments," *Astron. J.*, **76**, 2, 123–140.

Fritts, et al. (1984), "Research statistics and recommendation from the Alaska workshop on gravity waves and turbulence in the middle atmosphere," *Bull. Am. Met. Soc.*, **65**, 149–158.

Gaposchkin, E. M. and A. J. Coster (1993), *GPS L1-L2 bias determination*, Technical Report 971, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts.

Gary, B. L. (1992), *A novel method for monitoring global warming using unmanned aerial vehicles and GPS receivers*, White Paper (internal document), December 8, Jet Propulsion Laboratory, Pasadena, California.

Geller, M. (1983), "Dynamics of the middle atmosphere," *Space Sci. Rev.*, **34**, 359.

Gille, J. (1991), *Product Development Plan for Operational Satellite Stratospheric Products for the NOAA Climate and Global Change Program*, NOAA Climate and Global Change Program Special Report No. 5, 19–30, NOAA, Washington, D.C.

Goodman, J. W. (1968), *Introduction to Fourier Optics*, McGraw–Hill, New York, New York.

Gorbunov, M. E. and S. V. Sokolovskiy (1993), *Remote sensing of refractivity from space for global observations of atmospheric parameters*, Report 119, Max-Planck-Institut für Meteorologie, Hamburg.

Gu, M. and F. K. Brunner (1990), "Theory of the two frequency dispersive range correction," *Manuscripta Geodætica*, **15**, 357–361.

Hajj, G. A., E. R. Kursinski, W. I. Bertiger, and L. J. Romans (1994a), *Assessment of GPS occultations for atmospheric profiling, Proc. of the 7th Conf. on Satellite Meteorology and Oceanography*, Paper J4.9, American Meteorological Society, Monterey, California.

Hajj, G. A., R. Ibañez–Meier, E. R. Kursinski, and L. J. Romans (1994b), Imaging the Ionosphere with the Global Positioning System, *International Imaging Systems and Technology,* in print.

Hardy, K. R., D. P. Hinson, G. L. Tyler, and E. R. Kursinski (1992), "Atmospheric profiles from active space-based radio measurements," *Proc. of the Sixth Conference on Satellite Meteorology and Oceanography,* January 5–10.

Hardy, K. R., G. A. Hajj, and E. R. Kursinski (1993), "Accuracies of atmospheric profiles obtained from GPS occultations," *Proc. of the Institute of Navigation-GPS '93 Conf.*, Salt Lake City, Utah, September 22–24.

Hauchecorne, A., et al. (1993), *Scientific Feasibility of GLIMPSE*, Progress report submitted to CNES, Paris, France, March 11.

Hazard, C. (1962), "The method of lunar occultations and its application to a survey of the radio source 3C212," *Mon. Not. R. Astro. Soc.*, **124**, 343–357.

Hines, C. (1960), "Internal atmospheric gravity waves at ionospheric heights," *Can. J. Phys.*, **38**, 144–1481.

Hinson, D. P. and J. Magalhães (1991), "Equatorial waves in the stratosphere of Uranus," *Icarus*, **94**, 64–91.

Houghton, J. T., G. J. Jenkins, and J. J. Ephraums, eds. (1990), *Climate Change: The IPCC Scientific Assessment*, Cambridge University Press, New York, New York.

Hubbard, W. B., J. R. Jokipii, and B. A. Wilking (1978), "Stellar occultations by turbulent planetary atmospheres: A wave-optical theory including a finite scale height," *Icarus*, **34**, 374–395.

Keller, J. B. (1958), "Calculus of variations and its applications," *Proceedings of Symposia in Applied Math*, L. M. Graves (ed.), McGraw–Hill Co., New York, New York.

Keller, J. B. (1962), "Geometrical theory of diffraction," *J. Opt. Soc. Am.*, **52**, 2, 116–130.

Kelley, M., et al. (1981), "Gravity wave initiation of equatorial spread F: A case study," *J. Geoph. Res.*, **86**, 9087–9099.

King, J. (1992), "GPS/MET: OSC view of program approach," *DOE Proc. CEES Workshop*, Ft. Belvoir, Virginia, February 11–13.

Kliore, A. J., D. L. Cain, G. S. Levy, V. R. Eshleman, G. Fjeldbo, and F. D. Drake (1965), "Occultation experiment: Results of the first direct measurement of Mars' atmosphere and ionosphere," *Science*, **149**, 1243–1248.

Kliore, A. J., T. W. Hamilton, and D. L. Cain (1964), *Determination of some physical properties of the atmosphere of Mars from changes in the Doppler signal of a spacecraft on an Earth occultation trajectory*, Technical Report 32-674, Jet Propulsion Laboratory, Pasadena, California.

Klobuchar, J. A. (1985), "Ionospheric total electron content (TEC)," *Handbook of Geophysics and the Space Environment*, A. S. Jursa (ed.), Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts.

Kursinski, E. R., G. A. Hajj, and K. R. Hardy (1993a), "Temperature or moisture profiles from radio occultation measurements," *Proc. of the 8th Symp. on Meteorological Observations and Instrumentation*, Amer. Internat. Soc. for Optical Eng., Anaheim, California, J153–158.

Kursinski, E. R., G. A. Hajj, and K. R. Hardy (1993b), "Atmospheric profiles from radio occultation measurements of GPS satellites," *SPIE Conference*, Orlando, Florida, 116–127.

Lindal, G. F., G. E. Wood, H. Hotz, D. N. Sweetnam, V. R. Eshleman, and G. L. Tyler (1983), "The atmosphere of Titan: An analysis of the Voyager 1 radio occultation measurements," *Icarus*, **53**, 2, 348–363.

Lindal, G. F. (1992), "The atmosphere of Neptune: An analysis of radio occultation data acquired with Voyager 2," *Astron. J.*, **103**, 3, 967–982.

Lindzen, R. S. and M. Fox-Rabinovitz (1989), "Consistent vertical and horizontal resolution," *Monthly Weather Review*, **117**, 2575–2583.

Linfield, R. P. (1992), "Occultation of a compact radio source by Venus," *J. Astron.*, **104**, 2, 880–890.

Lusignan, B., G. Modreel, A. Morrison, J. Pomalaza, and S. G. Ungar (1969), "Sensing the Earth's atmosphere with occultation satellites," *Proc. IEEE*, **57**, 4, 458–467.

Marouf, E. A., G. J. Tyler, and V. R. Eshleman (1982), "Theory of radio occultation by Saturn's rings," *Icarus*, **49**, 161–93.

Marouf, E. A., G. J. Tyler, and P. A. Rosen (1986), "Profiling Saturn's rings by radio occultation," *Icarus*, **68**, 120–166.

McMillin, L. M., M. E. Gelman, A. Savyal, and M. Sylva (1988), "A method for use of satellite retrievals as a transfer standard to determine systematic radiosonde errors," *Mon. Wea. Rev.*, **116**, 1091–1102.

Meehan, T. K., J. M. Srinivasan, D. J. Spitzmesser, C. E. Dunn, J. Y. Ten, J. B. Thomas, T. N. Munson, and C. B. Duncan (1992), *The TurboRogue GPS receiver*, Paper presented at the Sixth International Geodetic Symposium on Satellite Positioning, Columbus, Ohio, March 18.

Melbourne, W. G. (1976), "Navigation between the planets," *Sci. Am.*, **234**, 6, 58–74.

Melbourne, W. G., T. P. Yunck, L. E. Young, B. H. Hager, G. F. Lindal, C. Liu, and G. H. Born (1988), *GPS geoscience instrument for EOS and Space Station*, GGI Proposal submitted to NASA, July 15.

Melbourne, W. G., B. D. Tapley, and T. P. Yunck (1994), "The GPS flight experiment on TOPEX/POSEIDON," *Geophys. Res. Lett.*, in press.

Newman, G., G. Schubert, A. J. Kliore, and I. R. Patel (1984), "Zonal winds in the middle atmosphere of Venus from Pioneer Venus radio occultation data," *J. Atmos. Sci.*, **41**, 1901–1913.

Moritz, H. (1961), "Zur reduktion elektronisch gemessener strecken und beobachteter winkel wegen refraktion," *Zeitschrift für Vermessungungswesen*, **86**, 7, 246–252.

Palmen, E. and C. W. Newton (1969), *Atmospheric Circulation Systems*, Academic Press, New York, New York.

Papas, C. H. (1965), *Theory of Electromagnetic Wave Propagation*, McGraw-Hill, New York, New York.

Phinney, R. A. and D. L. Anderson (1968), "On the radio occultation method for studying planetary atmospheres," *J. Geophys. Res*, **73**, 5, 1819–1827.

Reale, A. L., D. G. Gray, and M. W. Chalafant (1988), "TOVS operational soundings without tropospheric microwave data," *Proc. of the 3rd Conf. on Satellite Meteor. and Oceanography*, Amer. Meteorol. Soc., Anaheim, California, February, 30–34.

Salpeter, E. E. (1967), "Interplanetary scintillations: I. Theory," *Astron. J.*, **147**, 433–448.

Scheuer, D. A. G. (1962), "On the use of lunar occultations for investigating the angular structure of radio sources," *Australian Journal of Physics*, **15**, 3, 333–343.

Schutz, B. E., B. D. Tapley, P. A. M. Abusali, and H. J. Rim (1993), "Dynamic orbit determination using GPS measurements from TOPEX/POSEIDON," *Geophys. Res. Lett.*, in press.

Schmidlin, F. J. (1991), "Derivation and applications of temperature corrections for the United States radiosonde," *Preprints, Seventh Symp. on Meteorological Observations and Instrumentation,* New Orleans, Louisiana, 227–231.

Smith, E. K. and S. Weintraub (1953), "The constants in the equation for atmospheric refractive index at radio frequencies," *Proc. IEEE*, 1035–1037.

Spencer, R. W. and J. R. Christy (1990), "Precise monitoring of global temperature trends from satellites," *Science*, **247**, 1558–1562.

Tetens, O. (1930), "Über einige meteorologische begriffe," *Zeitshrift für Geophysik*, **6**, 297.

Thayer, D. (1974), "An improved equation for the radio refractive index in air," *Radio Science*, **9**, 803–807.

Tyler, G. J. (1987), "Radio propagation experiments in the outer solar system with Voyager," *Proc. IEEE*, **75**, 10, 1404–1431.

Uddstrom, M. J. and D. Q. Wark (1985), "A classification scheme for satellite temperature retrievals," *J. of Climate and Appl. Meteorol.*, **24**, 1, 16–29.

Ware, R. (1992), "GPS sounding of Earth's atmosphere," *GPS World*, **3**, 8, 56–57.

Williams D. C. (1975), "A theory of the curvature correction in electromagnetic distance measurement," *Survey Review,* **23**, 166–172.

Wu, S. C., J. T. Wu, and W. G. Melbourne (1993), "An optimal GPS data processing technique for precise positioning," *IEEE Trans.*, **31**, 1, 146–152.

Yates, H., D. Wark, H. Aumann, N. Evans, N. Phillips, J. Susskind, L. McMillin, A. Goldman, M. Chahine, and L. Crone (1989), "Simulation studies of improved sounding systems," *NOAA Tech. Rpt. NESDIS 42*, 109, NOAA, Washington, D.C.

Yeh, K. and C. Liu (1974), "Acoustic gravity waves in the upper atmosphere," *Rev. Geophys. Space Phys.*, **12**, 193–216.

Yuan, L. L., R. A. Anthes, R. H. Ware, C. Rocken, W. D. Bonner, M. G. Bevis, and S. Businger (1993), "Sensing climate change using the Global Positioning System," *J. Geophys. Res.*, 98:14925–14937.

Yunck, T. P., G. F. Lindal, and C. H. Liu (1988), "The role of GPS in precise Earth observations," *Proc. IEEE: Position, Location and Navigation Symposium*, Orlando, Florida, November 29 to December 2.

Yunck, T. P. (1991), "Coping with the atmosphere and ionosphere in precise satellite and ground positioning," *Environmental Effects on Spacecraft Trajectories and Positioning, Proc. IUGG Symposium,* AGU Monograph, Vienna, Austria, in press.

Yunck, T. P., W. I. Bertiger, S. C. Wu, Y. Bar-Sever, E. J. Christensen, B. J. Haines, S. M. Lichten, R. J. Muellerschoen, Y. Vigue, and P. Willis (1993), "First assessment of GPS-based reduced dynamic orbit determination on TOPEX/POSEIDON," *Geophys. Res. Lett.*, in press.

CHAPTER

14 GLOSSARY

AGW	acoustic gravity wave
AMD	Applied Microwave Devices, Inc.
AO	announcement of opportunity
AOA	Allen Osborne Associates, Inc.
ARIANE	CNES-developed European launch vehicle
AS	antispoofing
ASIC	applications specific integrated circuit
C/A	coarse acquisition pseudorange (code) *and* clear access
CNES	Centre National d'Etudes Spatiales
CT	computerized tomography
DIS	data information system
DORIS	Doppler orbitography radio-positioning integrated by satellite
EOS	Earth Observation System
ERS-1	Earth Research Satellite
ESA	European Space Agency
FLINN	fiducial laboratories for an international natural science network
GGI	GPS geoscience instrument
GLIMPSE	a CNES-sponsored mission concept
GPS	Global Positioning System
HIRS/2	High-Resolution Infrared Sounder
IGS	International GPS Service for Geodynamics
IR	infrared
IRIDIUM™	Motorola's proposed low orbiting communications satellite network
LEO	low Earth orbiter
LHS	left-hand side
MET	meteorological
Mil. Spec.	Military Specification
MKS	meter kilogram second
MSU	Microwave Sounding Unit
MTBF	mean time between failures
NRE	nonrecurring engineering
NSF	National Science Foundation
ORBCOMM™	Orbital Science Corporation's proposed low orbiting communications satellite network
OSC	Orbital Science Corporation
PDOP	position dilution of precision
Pegasus XL	Orbital Science Corporation's air-launched rocket for launching payloads of ~300 kg into low Earth orbit
PIM	parameterized ionospheric model
POD	precision orbit determination

PRARE	precise range and range rate equipment
RF	radio frequency
RHS	right-hand side
RMS	root mean square
SA	selective availability
SCR	satellite consumption rate
SDL	satellite design life
SGLS	space–ground link system
SNR	signal-to-noise ratio
SST	satellite-to-satellite tracking
TEC	total (columnar) electron content
TID	traveling ionospheric disturbance
TIGR	TIROS initial guess retrieval
UCAR	University Corporation for Atmospheric Research
UNAVCO	University Navstar Consortium
USO	ultrastable oscillator
UTC	universal time coordinated
VLSI	very large scale integration

APPENDIX

A MODEL BUDGET FOR IMPLEMENTING A CONSTELLATION

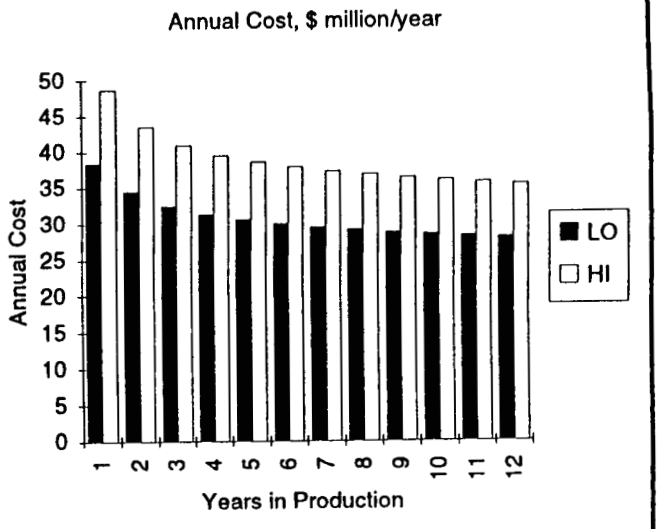
The cost model has two major components: (1) a non-recurring cost for developing the first single-purpose satellite flight system, including the ground systems to support the constellation, and (2) a recurring cost for manufacturing, deploying, operating, and maintaining a constellation of 20 to 30 microsatellites for producing refractivity profiles for the user community. The model for the baseline system is presented in Table A-1. Parameters are indicated in boldface and italic type. Computed values are in plain type. This "top-down" cost model shows the various budgetary trade-offs that are possible. The integrity of the individual estimates varies; some are based on reported contract values (such as the cost of a Pegasus launch) and others are program management estimates. The cost breakdown of the baseline system is presented in Figure A-1.

The cost breakdowns for other representative implementation concepts are presented in Figures A-2 and A-3. The concept represented in Figure A-2 assumes that the host satellite pays for everything except the incremental cost of integration with the satellite—hence the term "free rider."

The concept represented in Figure A-3 assumes that the dedicated satellite piggybacks on the Delta II. Launch costs are reduced to the incremental cost of integration of the satellite with the Delta II. To carry four piggyback satellites, the second stage of the Delta II may need to be requalified, hence the large uncertainty in the NRE costs for this option.

Table A-1. Cost model with baseline system parameters.

														LO	Per Satellite	3.366	2.813837	2.625	2.521	2.45	2.396	2.352	2.316	2.286	2.259	2.235	2.214
	GPSMET cost model													LO	Per Launch	9.5	9.122513	8.512	8.173	7.942	7.766	7.626	7.509	7.41	7.323	7.246	7.178
Operations Labor Rate			OLR=	0.12	per yr									LO	Per Year	3.933	3.933	3.933	3.933	3.933	3.933	3.933	3.933	3.933	3.933	3.933	3.933
Satellite Production Rate			SPR=	6.00	per year									HI	Per Satellite	4.712	3.939038	3.675	3.529	3.429	3.353	3.293	3.243	3.199	3.162	3.129	3.099
Satellites per Launch			SPL=	4.00										HI	Per Launch	10	9.802645	8.96	8.604	8.36	8.175	8.027	7.905	7.8	7.708	7.628	7.555
Learning Curve Parameter			LCP=	0.90										HI	Per Year	5.434	5.434	5.434	5.434	5.434	5.434	5.434	5.434	5.434	5.434	5.434	5.434
					3-year Development Phase			Per Satellite Cost	Per launch Costs	Per Year Costs		Per Satellite Cost	Per launch Costs	Per Year Costs													
			TOTAL		LO	HI		LO	LO	LO																
								CPS	CPL	CPY		CPS-HI	CPL-HI	CPY-HI													
					$9.73	$15.46		$3.37	$9.50	$3.93		$4.71	$10.00	$5.43		1	2	3	4	5	6	7	8	9	10	11	12
		LO	HI												LO	38	34	32	31	31	30	29	29	29	28	28	28
Development Management				HI		$0.80									HI	49	43	41	40	39	38	37	37	36	36	36	35
				Lo	$0.60					$0.20				$0.30													
		0.8	0.8																								
Engrng Improvement Program														$1.55													
● 0.1			Dev't Phase total/yr							$0.97																	
Payload				HI		$3.76	●					$0.95															
				Lo	$2.23		20%	$0.65																			
rcvr		1.9	2.8				of dev cost																				
ant		0.1	0.5				+$200k																				
ant I/F		0.03	0.08																								
USO		0.2	0.4																								
Satellite				HI		$4.70	●					$3.76															
				Lo	$3.40		80%	$2.72																			
Design		0.7	0.8				of dev cost																				
Subsys Acq		1.8	2.7																								
Sat. const		0.7	0.8																								
I&T		0.2	0.4																								
Launch				HI		$2.00							$10.00														
				Lo	$1.50				$9.50																		
I/F dev		1.5	2																								
Ground System				HI		$4.20																					
Development				Lo	$2.00																						
Comm sta		0.2	1																								
Ctrl Cntr		0.8	1																								
DIS h/w		0.3	0.4																								
DIS s/w		0.1	0.2																								
M ctrl		0.4	0.6																								
Data Prod s/w		0.5	1																								
Mission Operations						People				$2.76	HI =			$3.59													
Comm sta							2.00				LO+30%																
Con M&C							15.00																				
Ntwk M&C							1.00																				
DIS							1.00																				
Data Prod							2.00																				
MGT							2.00																				
						total staff	23.00																				



Annual Cost Breakdown	H	LO
NRE	15	10
Total/yr	49	38
Satellite	23	16
Launch	15	14
Payload	8	4
Mission Operations	4	3
Engrng Improv Prog	2	1
Development Mgt	0	0

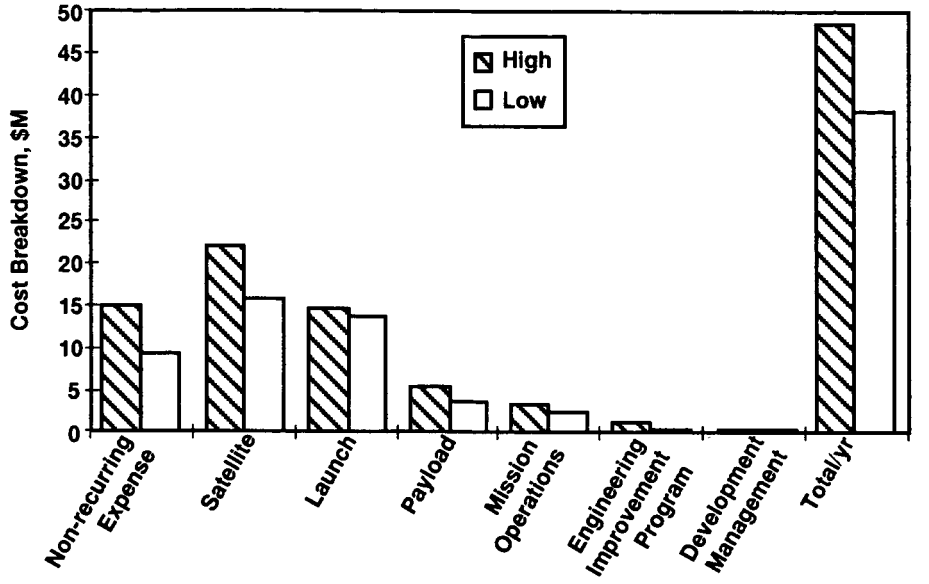


Figure A-1. Baseline System for a satellite system using Pegasus XL and deploying four satellites per launch.

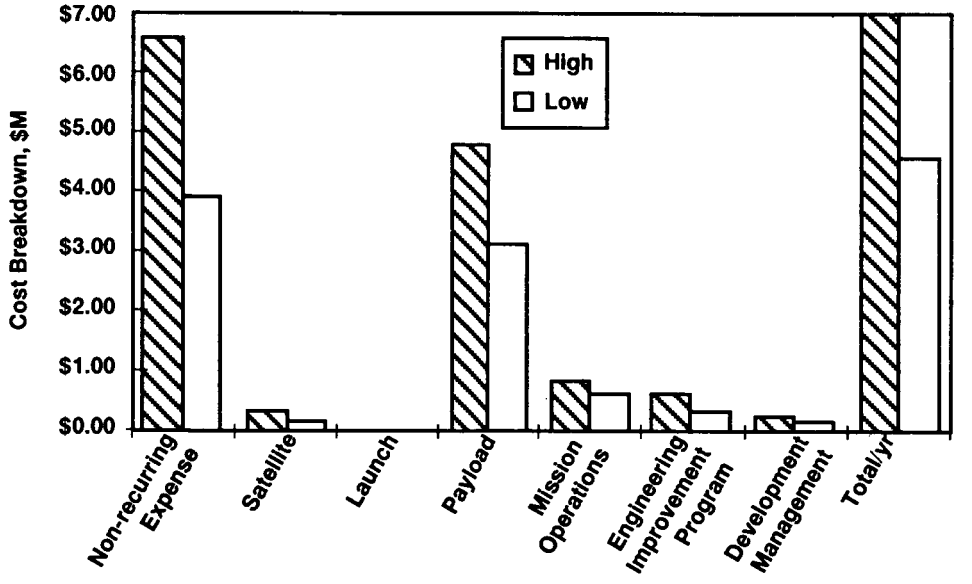


Figure A-2. Cost breakdown of free rider includes payloads for six satellites per year.

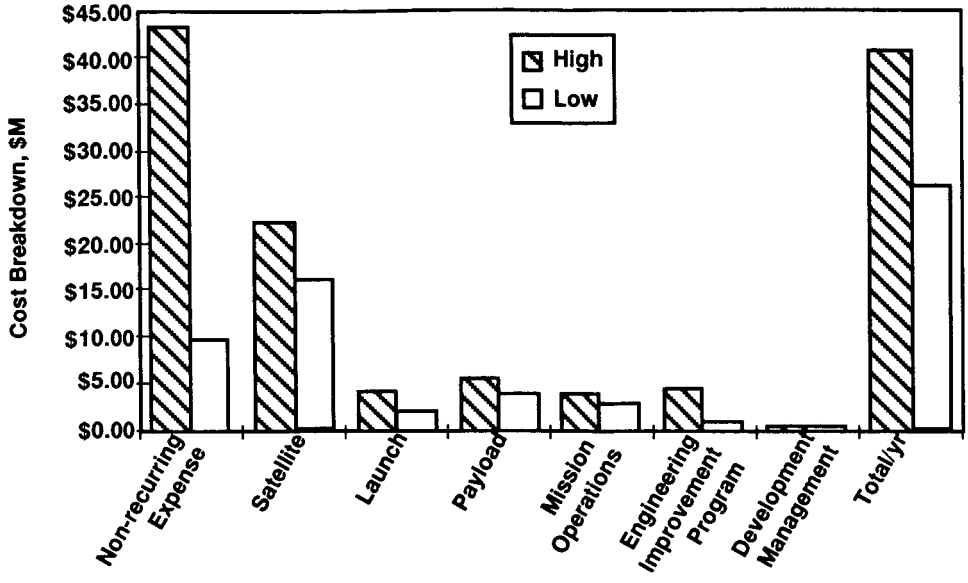


Figure A-3. Cost breakdown for piggyback on Delta II (four satellites per launch).

APPENDIX

B NOTES ON THE CALCULUS OF VARIATIONS

Stationary phase in geometric optics refers to how electromagnetic waves emanating from a point of origin A travel to the destination point B via a broad continuum of possible paths that spans (following some appropriate density distribution) the entire three-dimensional space in which A and B are embedded. Waves following different paths exhibit widely varying travel times, or time delays, because their phase velocities vary due to spatial refractivity differences and because the paths themselves are of different lengths. Therefore, the phases (modulo 2π) of nearly all of the waves when they arrive at B via these different paths will be uniformly distributed over 2π radians. Consequently, the waves in superposition will cause destructive interference or will cancel the combined electromagnetic disturbance at B. Only a special group of waves will constructively reinforce in superposition at B. These are waves that travel along paths that lie within a narrow sheath about the stationary phase path (essentially only waves arriving at B from directions lying within a solid angle subtended by the first Fresnel zone). Because these waves follow paths that result in only second-order changes in time delays, they will arrive at B virtually in-phase and, therefore, will constructively reinforce. The path of stationary phase is unique because it represents the centroid of this sheath, which contains those paths providing constructive reinforcement at B.

An alternative view of stationary phase is to consider the "number" of possible distinct paths arriving at B with time delays that lie within a small increment about a given time delay. We can construct the probability density distribution of this number versus its associated time delay. The path of stationary phase represents a *singularity* in this density distribution, usually, but not always, located at a global minimum point in time delay. A multipath configuration represents an example of multiple local minima. A caustic is an example of a conjugate point. Within a given neighborhood of the stationary phase path there are infinitely more paths (and hence more opportunities for reinforcement) per unit change in time delay than there are within a comparable neighborhood about a path of non-stationary phase.

The calculus of variations has been used in Chapter 7 to derive various refraction results from geometric optics using the principle of stationary phase. A sketch of the basic concepts from the calculus of variations is provided below.

The calculus of variations is a branch of mathematics that deals with a definite integral, or a functional in general, whose value depends on the path followed between the end points. Through the calculus of variations a path is found among a

suitably specified class of admissible or eligible paths. This path renders a stationary value for the integral. "Stationarity" has the same meaning as in ordinary calculus where setting the first derivative of a function to zero is used to isolate local minima, local maxima, or saddle points. Here, it means that the value of the integral when the optimal path is followed has no first-order dependence on the metric that describes the small differences in coordinates and slopes between a neighboring path and the optimal path.

It is convenient to represent the stationary phase integral in a generalized parametric form:

$$I = \int_{\tau_2}^{\tau_1} f(x^i, \dot{x}^i) d\tau \qquad \text{(B.1)}$$

where $(\dot{})$ denotes differentiation with respect to the parameter τ. The boundary conditions are given by

$$x^i(\tau_1) = x_1^i, \qquad x^i(\tau_2) = x_2^i, \qquad i = 1,\ 2,\ 3 \qquad \text{(B.2)}$$

The problem for the calculus of variations, then, is to find the path, described in parametric form by the coordinates $x^i(\tau)$, that renders a stationary value for the integral I. In the simplest version of the calculus of variations, which applies to the stationary phase problem, the functional forms for the $x^i(\tau)$ and their derivatives are unconstrained by side conditions other than the boundary conditions and certain continuity conditions.

For a calculus of variations problem posed in a parametric form, as given in eq. (B.1), the value of the integral I must be invariant to the choice of parameter τ for a given set of boundary conditions. The value of I must depend only on the path followed and the spatial environment in which the path is embedded. The parameter τ is merely an indexing convenience denoting the position on the path; it bears no relevance to the physical problem, and we should be able to replace it with an alternate parameter without changing the results. If we replace τ with η through the relationship $\eta = \eta(\tau)$ and define $\kappa = \dot{\eta}$, then the integrand of eq. (B.1) becomes

$$f(x^i, \dot{x}^i) d\tau = f(x^i, \kappa x'^i) d\eta / \kappa \qquad \text{(B.3)}$$

where $(')$ means differentiation with respect to η. For invariance to hold, it follows that not only must τ be explicitly absent as an argument in f, but f must also be a homogeneous function of order one in the $\dot{x}^i$. This means that f must satisfy the condition

$$f(x^i, \kappa \dot{x}^i) = \kappa f(x^i, \dot{x}^i) \qquad \text{(B.4)}$$

Differentiating both sides of eq. (B.4) with respect to κ and setting $\kappa = 1$ yields the equivalent condition

$$\dot{x}^i f_{\dot{x}^i} = f \qquad \text{(B.5)}$$

where the Einstein summation rule on the index i has been used. Certainly the path delay integrand

$$f = n(x, y, z)\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} \qquad \text{(B.6)}$$

satisfies this condition.

The value of the integral I for the path delay is clearly dependent on the choice of paths that is followed between x_1^i and x_2^i. We wish to find the path that yields a stationary value for I, which means that to first order the variation in the value of I is zero as a result of following a neighboring path that is close to the stationary phase path. Let the parametric representation of the stationary phase path $\mathscr{E}$ be denoted by the coordinates $x^i(\tau)$. Let the small differences in coordinates and slopes between a neighboring path and the optimal path be denoted by $\delta x^i(\tau)$ and $\delta \dot{x}^i(\tau)$ as shown in Figure B-1. The functional forms of the variations $\delta x^i(\tau)$ are arbitrary, other than being

required to vanish at the end points. Also, we assume that $f(x^i, \dot{x}^i)$ is a well-behaved function, typical of those found in most physical problems; thus, we assume that the $\dot{x}^i(\tau)$ are at least piecewise continuous along the stationary phase path. The same continuity conditions apply to $\delta\dot{x}^i(\tau)$. Then, a necessary condition for the integral I in eq. (B.1) to take on a stationary value, with respect to the values of I obtained from following any other admissible neighboring path satisfying the boundary conditions, is that its first variation must be zero. This is obtained by expanding $f(x^i, \dot{x}^i)$ in a Taylor's series in powers of δx^i and $\delta\dot{x}^i$ and integrating each term over the reference path. The change in I as a result of following the alternative path is given by

$$\Delta I = \int_{\tau_1}^{\tau_2} \left[f_{x^i}\delta x^i + f_{\dot{x}^i}\delta\dot{x}^i \right] d\tau + O\left[\delta x^i(\tau), \delta\dot{x}^i(\tau)\right]^2 \tag{B.7a}$$

where the symbol $O[\bullet]^2$ denotes those integrations involving second- and higher order terms in the variations of $x^i(\tau)$ and $\dot{x}^i(\tau)$ and their cross products. We assume that the magnitudes of $\delta x^i(\tau)$ and $\delta\dot{x}^i(\tau)$ can be made sufficiently small so that the second- and higher order terms can be safely ignored relative to the first-order terms. The first variation δI includes only the first-order terms in $\delta x^i(\tau)$ and $\delta\dot{x}^i(\tau)$ and is given by

$$\delta I = \int_{\tau_1}^{\tau_2} \left[f_{x^i}\delta x^i + f_{\dot{x}^i}\delta\dot{x}^i \right] d\tau \tag{B.7b}$$

Here, the arguments of the partial derivatives of f, f_{x^i}, and $f_{\dot{x}^i}$ are evaluated on the reference path. Integrating the second term by parts and invoking the boundary conditions that $\delta x^i(\tau_1) = \delta x^i(\tau_2) = 0$, we obtain

$$\delta I = \int_{\tau_1}^{\tau_2} \left[f_{x^i} - \frac{d}{d\tau}\left(f_{\dot{x}^i}\right) \right] \delta x^i d\tau \tag{B.8}$$

For I to assume a stationary value, it follows that δI must be zero for all admissible paths neighboring the path of stationary phase; otherwise, we could find a neighboring path that would provide a first-order change in the integral and, hence, a smaller or larger value for the integral. Inasmuch as the $\delta x^i(\tau)$ and their derivatives are not restricted for the stationary phase problem by additional constraints, their functional forms are quite arbitrary other than their values being required to vanish at the end points; also their derivatives must be at least piecewise continuous. In this case, it follows that the bracketed term $[\bullet]$ in eq. (B.8) must be zero at every point along the optimal path, provided that the operation $d(f_{\dot{x}^i})/d\tau$ is admissible (see below). For if $[\bullet]$ were not zero, even over a very small but finite interval on $\mathscr{E}$, we could find a corresponding variation $\delta x^i(\tau)$ over that interval that would render δI either positive or negative; this would violate the proposition that I assumes a stationary value when the path defined by $\mathscr{E}$ is followed. Thus, the second-order differential equations

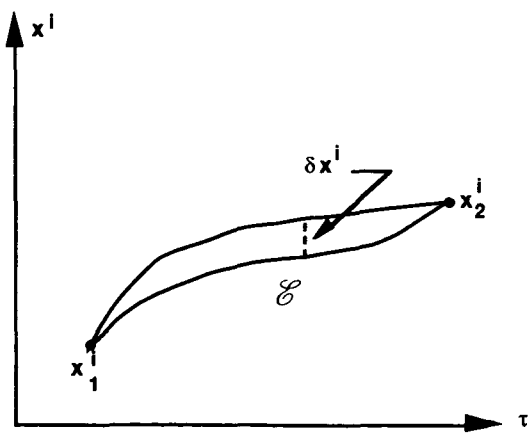


Figure B-1. Optimal phase path $\mathscr{E}$ and its neighbor satisfy the same boundary conditions.

$$f_{x^i} - \frac{d}{d\tau}\left(f_{\dot{x}^i}\right) = 0, \quad i = 1, 2, 3 \quad \text{(B.9)}$$

must be satisfied at all points along the stationary phase path (except possibly those points where the second derivatives do not exist). The equations used in eq. (B.9) are the famous Euler equations and are necessary conditions for the path to provide a stationary phase value for I.

The differentiation by parts to convert eq. (B.7b) to eq. (B.8) requires that the second derivatives of x^i exist so that the operation $d(f_{\dot{x}^i})/d\tau$ is valid. Equation (B.8) implicitly assumes that $\dot{x}^i$ are at least piecewise continuous and requires the customary care in defining the right-hand and left-hand limits around a discontinuity point in $\dot{x}^i(\tau)$. If the $\dot{x}^i$ are piecewise continuous on the path of stationary phase, their discontinuities must satisfy certain continuity conditions, which in the calculus of variations are known as corner conditions. Integrating the first term in eq. (B.7b) by parts does not involve any differentiability assumptions on the $\dot{x}^i$. Carrying out this integration and invoking the boundary conditions where δx^i must vanish at the end points obtains

$$\delta I = \int_{\tau_1}^{\tau_2} \left[f_{\dot{x}^i} - \int_{\tau_1}^{\tau} f_{x^i} dt \right] \delta\dot{x}^i \, d\tau \quad \text{(B.10)}$$

Since the functional forms for the $\delta\dot{x}^i$ are arbitrary—other than subject to the constraint that their integrals must vanish at the end points—and since they have been assumed to be at least piecewise continuous, we conclude, using similar arguments to those used in eq. (B.9), that if δI is to be zero the bracketed term [•] in eq. (B.10) must be invariant with τ and, indeed, can be at most a constant. Therefore, we obtain the conditions

$$\int_{\tau_1}^{\tau} f_{x^i} dt - f_{\dot{x}^i}(\tau) = K_i, \quad i = 1, 2, 3 \quad \text{(B.11)}$$

where the K_i's are constants along the path of stationary phase $\mathscr{E}$. From eq. (B.11), we obtain the continuity conditions that must be satisfied by $\mathscr{E}$, namely, the $f_{\dot{x}^i}$ must be continuous at all points on $\mathscr{E}$, even at points of discontinuity in $\dot{x}^i$. The integral version of Euler's equations in eq. (B.11) is more general in the sense that it holds at all points on $\mathscr{E}$, even at points where the second derivative of x^i may not exist and therefore at points where the differentiation operation $d(f_{\dot{x}^i})/d\tau$ may not be valid. These conditions may be used to derive Snell's refraction law at corners of $\mathscr{E}$, where the slope is discontinuous as a result of n(x,y,z) being discontinuous.

Applying eq. (B.9) to the path delay integrand in eq. (B.6) gives the set of differential equations that in conjunction with the boundary conditions must be solved to yield the path of stationary phase. We obtain in vector form

$$\frac{d(n\mathbf{T})}{d\tau} = \nabla n \, (\dot{x}^2 + \dot{y}^2 + \dot{z}^2)^{\frac{1}{2}} \quad \text{(B.12)}$$

where $\mathbf{T}$ is the unit tangent vector to the path and given by

$$\mathbf{T} = \frac{\boldsymbol{i}\dot{x} + \boldsymbol{j}\dot{y} + \boldsymbol{k}\dot{z}}{(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)^{\frac{1}{2}}} = \frac{\dot{\mathbf{r}}}{|\dot{\mathbf{r}}|} \quad \text{(B.13)}$$

Without loss of generality (because of the homogeneity property of $f = n(x,y,z)\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}$ given by eq. (B.4)), we can set the parameter τ equal to arc length "s" along the path and eq. (B.12) becomes

$$\frac{d(n\mathbf{T})}{ds} = \nabla n \quad \text{(B.14)}$$

From differential geometry, the radius of curvature vector $\mathscr{R}(\mathbf{r})$ of a point on a path at the position $\mathbf{r}$ is given by

$$\frac{\mathscr{R}}{\mathscr{R}^2} = \frac{d\mathbf{T}}{ds} \qquad (B.15)$$

from which it follows that eq. (B.14) becomes

$$n\frac{\mathscr{R}}{\mathscr{R}^2} = \nabla n - (\nabla n \cdot \mathbf{T})\mathbf{T} \qquad (B.16)$$

which is identical with eq. (7.3b).

Suppose now that the right-hand end point of $\mathscr{E}$ is moving at a velocity given by the vector **V**. After a short-time interval Δt_2, the end-point coordinates will have been displaced by a small amount $\Delta x^i(t_2)$ as shown in Figure B-2. Now there are two stationary phase paths starting from the point $x^i(t_1)$; the small separation in coordinates between them at a specified time t is denoted by the quantities $\delta x^i(t)$. At time t_2, it follows from Figure B-2 that $\Delta x^i(t_2)$ and $\delta x^i(t_2)$ differ because of the slopes $\dot{x}^i$ at t_2 and are related by the conditions

$$\delta x^i(t_2) = \Delta x^i(t_2) - \dot{x}^i \Delta t_2 + O(\Delta t_2^2), \quad i = 1, 2, 3 \qquad (B.17)$$

The change in the stationary phase integral as a result of this end-point displacement is

$$\Delta I = \int_{t_1}^{t_2} [f_{x^i}\delta x^i + f_{\dot{x}^i}\delta \dot{x}^i]\, dt + [f\Delta t]\Big|_{t_2} + O(\Delta t_2^2) \qquad (B.18)$$

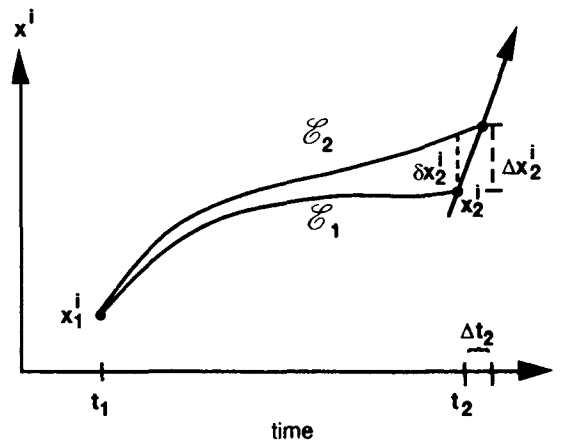


Figure B-2. Transversality condition.

Integrating the second term in the integrand by parts and using eq. (B.17) (since both paths must satisfy Euler's equations), we obtain

$$\Delta I = [(f - \dot{x}^i f_{\dot{x}^i})\Delta t + f_{\dot{x}^i}\Delta x^i]\Big|_{t_2} + O(\Delta t_2^2) \qquad (B.19)$$

However, because of the homogeneity property of f in eq. (B.5), the coefficient of the Δt term in eq. (B.19) is identically zero, which corresponds to the fact that the path delay integral does not depend explicitly on time. Note also that ΔI in eq. (B.19) does not depend to first order on the coordinates along the stationary phase path followed to reach the end point, but only on conditions at the end points, which follows from the stationarity property of $\mathscr{E}$. By dividing eq. (B.19) by Δt_2 and taking the limit, we obtain the time derivative of the stationary phase, i.e., the Doppler, which depends only on current end-point positions and velocities. Using eq. (B.6) for f we obtain

$$\frac{dI}{dt} = n(\mathbf{r})\mathbf{T}\cdot\mathbf{V} \qquad (B.20)$$

Upon applying eq. (B.20) to both ends of $\mathscr{E}$, the Doppler expression in eq. (7.7) is obtained.

As a cultural aside for those determined readers who have toughed it out this far, it should be noted that for a generalized f, the partial differential equations obtained from eq. (B.19) by identifying $\partial I/\partial t_2$ with the coefficient $(f - \dot{x}^i f_{\dot{x}^i})$ and $\partial I/\partial x_2^i$ with $f_{\dot{x}^i}$, form the basis for the Hamilton–Jacobi equation. This equation describes the behavior of the stationary value of I(x, y, z), which is here considered as a function of its right-hand end point only, over an entire region in (x, y, z) space within which a three-parameter family of stationary paths, all originating from the point (x_1, y_1, z_1) is embedded. For the path delay problem, I(x, y, z) becomes the phase

delay associated with a geometric wavefront at the point (x, y, z).

In the Hamiltonian formulation of the calculus of variations, a generalized momentum vector **p**, which is canonically conjugate to **r**, is introduced through the defining relationship

$$p_i = f_{\dot{x}^i} \tag{B.21}$$

from which we can functionally express the $\dot{x}^i$ in terms of the coordinates x^i and the momenta p_i (if the determinant $|f_{\dot{x}^i\dot{x}^j}|$ is non-zero, which holds for most physical problems). For the phase delay integrand given in eq. (B.6), it is easily shown that

$$\mathbf{p} = n\mathbf{T} \tag{B.22}$$

which is the "momentum" vector for the ray. It can be shown that the Hamiltonian function, defined by $H = \dot{x}^i f_{\dot{x}^i} - f$, is given for the phase delay integrand in canonically conjugate position and momentum coordinates by

$$\mathscr{H}(x^i, p_i) = \sqrt{p_x^2 + p_y^2 + p_z^2} - n(x,y,z) \equiv 0 \tag{B.23}$$

and that Hamilton's equations form a double set of first-order differential equations that are equivalent to the second-order Euler equations and describe the positions and momenta of the stationary phase path given by

$$\begin{cases} \dfrac{dx^i}{dt} = \dfrac{\partial \mathscr{H}}{\partial p_i} = \dfrac{p_i}{p} \\ \dfrac{dp_i}{dt} = -\dfrac{\partial \mathscr{H}}{\partial x^i} = \dfrac{\partial n}{\partial x^i} \end{cases} \tag{B.24}$$

Finally, upon noting that $p_i = \partial I/\partial x^i$, the Hamilton–Jacobi equation for this case becomes

$$\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2 + \left(\frac{\partial I}{\partial z}\right)^2 = n^2(x,y,z) \tag{B.25}$$

This is known as the eikonal equation and must be satisfied by the phase of a wavefront passing through a refracting medium.

The Hamilton–Jacobi partial differential equation generally forms a field-theoretic basis for a number of powerful and computationally intensive numerical techniques that with the advent of high-speed computers are used for solving physical problems that can be couched in terms of some stationarity principle. These techniques (e.g., Dynamic Programming) provide not only a single optimal path and the stationary value of its associated integral but an entire field of such paths and integrals corresponding to essentially a continuum of boundary values. It is also a short step (but a long philosophical reach) from the eikonal equation in eq. (B.25) to Schroedinger's equation in quantum mechanics.

Snell's law can be obtained by forcing the stationary phase path to intersect a surface on which the index of refraction is discontinuous. This is the so-called boundary intercept problem in the calculus of variations. In Figure B-3, the stationary phase integral is broken into two segments. The right-end point of the left-hand segment is free to move along the boundary surface S and, therefore, the left-end point of the right-hand segment must do the same. For I to have a stationary phase value, not only must Euler's equations be satisfied along both path segments, but the first variation of I caused by any displacement along the boundary surface must be zero. Let the surface S defining the discontinuity boundary for n be given by

$$S(x, y, z) = 0 \tag{B.26}$$

Then, the normal vector **N** to the surface at the position $\bar{r} = (\bar{x}, \bar{y}, \bar{z})$ is given by

$$\mathbf{N} = \frac{\nabla S}{|\nabla S|} \tag{B.27}$$

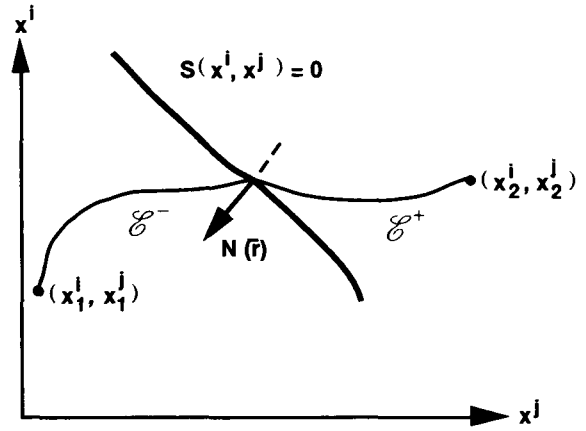


Figure B-3. Boundary intercept problem.

Following the same end-point variation procedure as before but applying it to both $\mathscr{E}^-$ and $\mathscr{E}^+$, the first variation of I is given by

$$\delta I = \left[f_{\dot{x}^i}\right]^- \delta\bar{x}^i - \left[f_{\dot{x}^i}\right]^+ \delta\bar{x}^i = 0 \qquad \text{(B.28)}$$

Using eq. (B.6) for f, eq. (B.28) becomes in vector form

$$[n(\bar{r})^-\mathbf{T}^- - n(\bar{r})^+\mathbf{T}^+] \cdot \delta\bar{r} = 0 \qquad \text{(B.29)}$$

Since the vector $\bar{r}$ is constrained to lie on the surface S, it follows that the vector $\delta\bar{r}$ is constrained to lie in the tangent plane to S at the position $\bar{r}$ (otherwise, the first variation of S would not be zero). Since the vector $\delta\bar{r}$ may assume an arbitrary value within the tangent plane, it follows that the vector $[n(\bar{r})^-\mathbf{T}^- - n(\bar{r})^+\mathbf{T}^+]$ must be perpendicular to the tangent plane, hence the relation

$$\left[n(\bar{r})\mathbf{T}(\bar{r})\right]^- \times \mathbf{N}(\bar{r}) = \left[n(\bar{r})\mathbf{T}(\bar{r})\right]^+ \times \mathbf{N}(\bar{r}) \qquad \text{(B.30)}$$

which is Snell's law.

Snell's law can also be obtained from the continuity conditions on $\mathscr{E}$ as given in eq. (B.11). At the point of discontinuity in n, relax the discontinuity by allowing n to have a very large but finite gradient vector ∇n over a short distance at the position $\bar{r}$ across the surface S, which by the definition of S must be in the direction of **N**. Applying eq. (B.11) to $\mathscr{E}$ on both sides of the jump in n and taking the difference obtain

$$n(\bar{r})^+\mathbf{T}^+ - n(\bar{r})^-\mathbf{T}^- = \int_-^+ \nabla n(\mathbf{r})ds = \mathbf{N}\int_-^+ |\nabla n|ds \qquad \text{(B.31)}$$

from which we again conclude that $[n(\bar{r})^+\mathbf{T}^+ - n(\bar{r})^-\mathbf{T}^-]$ must be perpendicular to the tangent plane of S at $\bar{r}$. Taking the cross product of eq. (B.31) with **N** yields eq. (B.30).

APPENDIX

C COMMENTS ON THE USE OF THE ONION SKIN MODEL

In ray tracing through the layered "onion skin" model, intralayer ray path curvature should be taken into account if an efficient, discrete approximation of the continuous case is to be achieved. The recovered profile of refractivity values will be corrupted by granulation effects unless an accurate method is used. One simple technique is to assume that the gradient of the refractivity remains constant within a layer. In this case, we can invoke the condition of continuity in refractivity across the boundary of each layer, but must allow a stepwise change in its gradient at the boundary. In this scheme, Snell's law need not be applied at all since the bending is obtained from a path integral of the curvature term, eq. (7.3b), within each layer. This approach is in contrast to one in which the refractivity is assumed to be constant within each layer but allowed to change discontinuously at the boundaries of the layers. Here, Snell's law is applied across each boundary, but the intralayer ray path segments are straight lines. This latter approach requires a much finer layer thickness and, therefore, a much larger number of layers.

The difference in efficiency between these two techniques can be seen by applying a spherically symmetric, exponential atmosphere, eq.(7.24), to each. A locally straight-line approximation[1] of the actual ray path is used for the purpose of evaluating the relative rates of convergence of the two techniques to the desired precision. The atmosphere is composed of m layers, all of equal thickness in height, Δh. In the case of a ray that is just grazing the Earth's surface, the total bending angle α of the ray path is given for the constant gradient approach by

$$10^6\alpha(\Delta h) = 2\sum_{k=1}^{m}\int_{s_k}^{s_{k+1}}(T\times\nabla N)ds \qquad (C.1)$$

$$\cong 2\sum_{k=0}^{m-1}\frac{N_o}{H}\exp\left(-\frac{k\Delta h}{H}\right)\left[s_{k+1}-s_k\right]$$

where Δh is the thickness of the layers and where s_k is the arc length to the kth layer, with $s_0 = 0$ corresponding to the midpoint of the ray path. The arc length using the straight-line approximation can be approximated by $s_k = \sqrt{2Rk\Delta h}$, where R is the radius of the Earth. The quantity N_0 is the surface refractivity.

[1] In the actual ray-tracing algorithms the local deflection of the ray path from the direction of its reference straight line, as well as its curvature, should be taken into account to preserve computational accuracy. Here, for the purpose of comparing the rates of convergence of the two formulations to their respective limits, a straight-line approximation is sufficient. Also, small corrections arising from cosine effects are omitted here.

On the other hand, for the constant refractivity approach α is given by

$$10^6\alpha(\Delta h) = 2\times 10^6 \sum_{k=1}^{m}\Delta\alpha_k$$
$$\doteq -2\sum_{k=1}^{m}\Delta N_k \tan\theta_k \qquad (C.2)$$
$$\cong 2\sum_{k=1}^{m}\frac{N_o}{H}\Delta h \exp\left(-\frac{k\Delta h}{H}\right)\left(\frac{R}{s_k}\right)$$

where $\Delta\alpha_k$ is given by Snell's law at the kth boundary, eq. (7.6b), and θ is the angle of incidence of the ray with the normal to the boundary surface.

The rate at which these two forms for α converge to the "true value" of $10^{-6}N_o\sqrt{2\pi R/H}$ (see the first-order expression for α given by eq. (7.25a)), as Δh approaches zero, is shown in Figure C-1. Equation (C.1) achieves a 1% computational precision for a layer thickness of about 1 km, whereas eq. (C.2) requires a layer thickness of about 1 m to achieve the same precision.

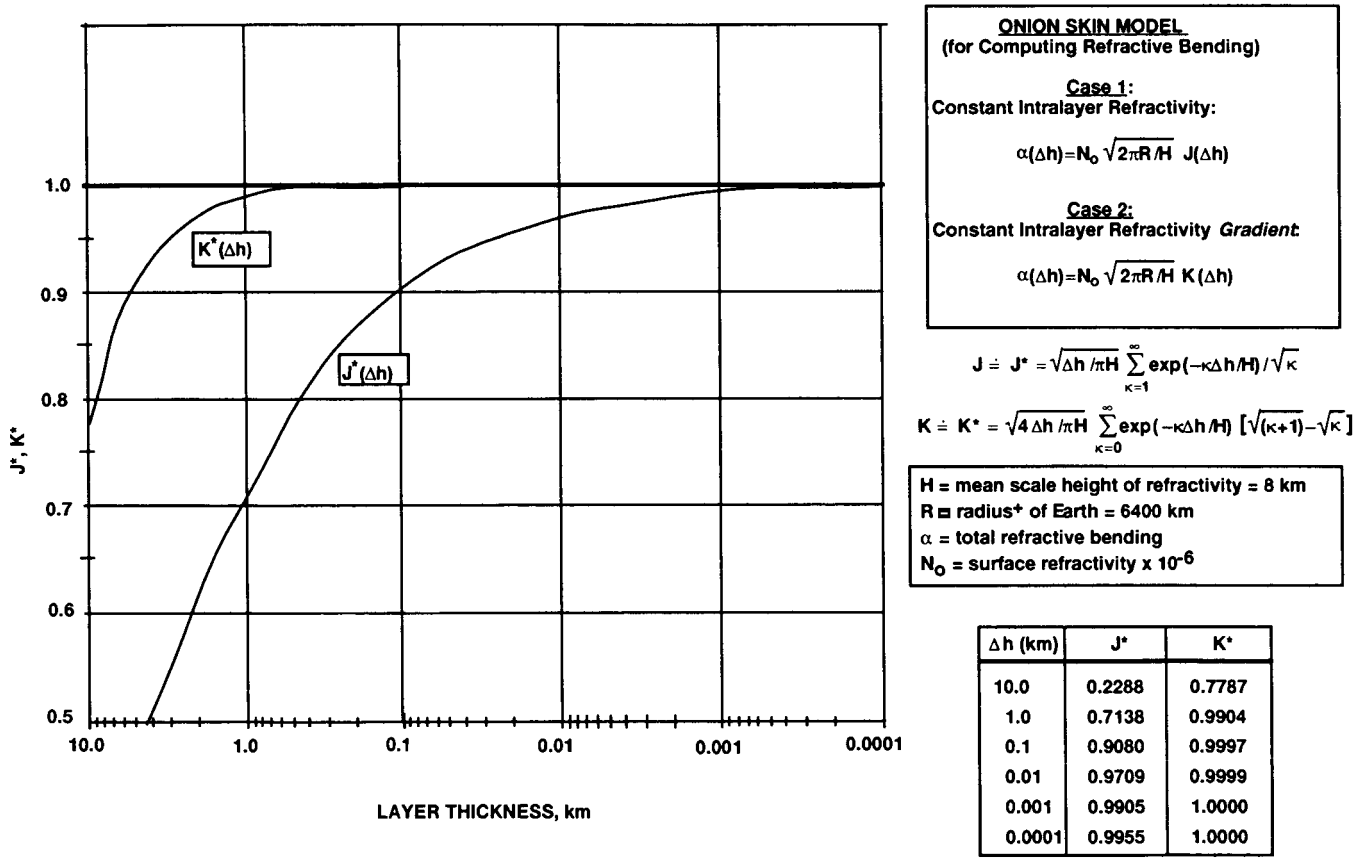


Figure C-1. Required layer thickness to achieve a given precision using Snell's law at layer boundaries (J) or using ray-path curvature within the layers (K).

APPENDIX

D COMMENTS ON SECOND-ORDER IONOSPHERIC EFFECTS USING A UNIFORM SHELL MODEL

The Uniform Shell Model

Here, we assume a uniform spherical distribution for the electron distribution in the ionosphere, which is given by

$$n_e(h) = \begin{cases} \frac{N_e}{\sigma}, & h_o - \sigma/2 \le h \le h_o + \sigma/2 \\ 0, & \text{otherwise} \end{cases} \quad \text{(D.1)}$$

where N_e is the total integrated electron density along a local vertical with nominal values in the range of 0.1–30 in units of 10^{17} electrons/m^2. Also, h_o is the mean altitude of the shell, and σ is its width; values for the latter two quantities are nominally in the range of 100–400 km. In reality these bulk parameters of the ionosphere depend upon a number of variables: e.g., the diurnal phase, solar cycle phase, sunspot and solar flare activity, and geographical location.

The uniform shell model is a valuable qualitative tool for assessing the magnitude of various ionospheric effects. It is more applicable to vertical soundings than to limb soundings because of inhomogeneities in electron density incurred along the tangential path in the limb-sounding case. This model effectively collapses the transverse gradient feature of a more realistic model (such as the PIM ionosphere shown in Figure 8-15) to the boundary surfaces where Snell's law applies. It follows from Snell's law and Figure 8-15 that the total bending angle at the LEO for this model, ignoring the Earth's magnetic field, is given through second order by ($\delta\alpha$ is positive in a counter-clockwise sense in Figure 8-15)

$$\begin{aligned} \delta\alpha = 2\frac{qN_e}{\sigma f^2}\Big[&\big(\cot(\theta-\phi)-\cot\theta\big) \\ &+\frac{qN_e}{2\sigma f^2}\big(2\cot\theta\,\cot^2(\theta-\phi) \\ &-\cot^3\theta-\cot^3(\theta-\phi)\big)\Big] \end{aligned} \quad \text{(D.2)}$$

where θ is the angle between the incident ray and the local horizontal of the ionospheric surface at the altitude $h_o + \sigma/2$; for an altitude h_o, θ is given to a good approximation by

$$\theta_o \doteq \sqrt{\frac{2h_o}{R_E}} \quad \text{(D.3)}$$

where R_E is the radius of the Earth. The angle θ–ϕ corresponds to the inner boundary of the shell (at an altitude of $h_o - \sigma/2$). From eq. (D.2) it follows that the second-order effect on the bending angle is of the order of a nanoradian times N_e^2, multipled by some potentially significant geometrical factors, particularly so when h_o and σ are small.

From eq. (8.2) the Doppler error at the LEO is given by

$$\delta\dot{\rho} = \dot{b}\delta\alpha = \pm 3\times 10^6\,\delta\alpha \quad \text{(mm/s)} \quad \text{(D.4)}$$

Application of the two-frequency algorithm in eq. (8.74) will eliminate the first-order term in eq. (D.2) but not the second. In fact, if we define ρ_3 as the linear combination of ρ_1 and ρ_2 defined in the LHS of eq. (8.66), the residual error in the ionospherically corrected Doppler, $\dot{\rho}_3$, is given by

$$\delta\dot{\rho}_3 = \pm 3\times 10^6 \frac{q^2 N_e^2}{\sigma^2 f_1^2 f_2^2} \times\left[\left(\cot^3\theta + \cot^3(\theta-\phi) - 2\cot\theta\times\cot^2(\theta-\phi)\right)\right] \quad \text{(D.5)}$$

and corresponds to the bending term (sixth term on the RHS) in eq. (8.74).

It is noted that $\delta\dot{\rho}_3$ is a kind of group delay rate; therefore, it carries inherently the opposite sign of $\dot{b}$. To a good approximation, it is given by

$$\delta\dot{\rho}_3 \doteq \mp 7.5\times 10^5\left[\frac{qN_e}{f_1 f_2}\right]^2 \frac{R_E}{\sigma h_o^2} \quad (\text{mm/s}) \quad \text{(D.6)}$$

For σ = 200 km and h_o = 300 km (the length unit to be used in eq. (D.6) is the meter), eq. (D.6) yields a second-order Doppler error after the two-frequency correction of $\delta\dot{\rho}_3 = 0.002\,N_e^2$ mm/s, which can exceed the nominal SNR-based L3 Doppler error for large N_e. Moreover, for low altitude ionospheric layers, the E layer for example, where typical values of σ and h_o may be ~100 km (which corresponds to a large gradient in refractivity normal to the ray path and a nearly tangential ray path), $\delta\dot{\rho}_3$ can be an order of magnitude larger for the same N_e.

Another second-order effect, which is easier to deal with than the direct ionospheric second-order effects on Doppler, is the deviation Δ in the neutral atmosphere (Δ = z(0) in Figure 8-15) that arises from the L1 and L2 path splitting. This results in different refractive bending angles for L1 and L2 because of the vertical gradient in refractivity in the neutral atmosphere; hence by eq. (8.2), we measure a different Doppler from that obtained from the straight-line path (z = 0). Using the model in eq. (D.1) with eq. (D.2) we can show that the first-order term in Δ is given by

$$\Delta \doteq \frac{qN_e}{f^2}\sqrt{\frac{R_E}{2h_o}}\left[\sqrt{\frac{R_E}{2h_o}} - 1\right] \quad \text{(D.7)}$$

which is typically of the order of $10N_e$ meters. The quantity Δ should be implicitly accounted for in the occultation data analyses; the integral inversion and ray-tracing algorithms should include ionospheric effects on the altitude of the ray paths through the neutral atmosphere. The second-order term in Δ is typically less than 1 m, which need not be further considered here. However, the change in Doppler from the first-order value for Δ given in eq. (D.7) will be proportional to f^{-2} *provided* that the variations in the neutral atmospheric bending caused by a non-zero value of Δ are in the linear regime. The application of eq. (8.74) to the L1 and L2 Doppler measurements effectively eliminates the first-order atmospheric Doppler effects resulting from the deviations Δ_{L1} and Δ_{L2} from the z = 0 line and, therefore, eliminates them in recovering the neutral atmosphere refractivity. The second-order term from the neutral atmosphere using an exponential model is given by

$$\delta^2\alpha_{atmos} = \frac{\Delta^2}{2H^2}\,\alpha_{atmos} \quad \text{(D.8)}$$

where H is the local scale height of the refractivity. This results in a second-order error in the two-frequency corrected Doppler of

$$\delta\dot{\rho}_{3atmos} = \pm 7.5\times 10^5\left[\frac{qN_e}{Hf_1 f_2}\right]^2\left[\frac{R_E}{2h_o}\right] \times\left[\sqrt{\frac{R_E}{2h_o}} - 1\right]^2 \alpha_{atmos} \quad (\text{mm/s}) \quad \text{(D.9)}$$

which for nominal values for H and h_0 yields a second-order Doppler effect of about 100 $N_e^2 \alpha_{atmos}$ mm/s. Thus, even in the upper atmosphere, this second-order term caused by L1 and L2 path splitting in the neutral atmosphere should be accounted for when N_e is large. In this case, apply a differential correction for Δ using the local refractive gradient for the neutral atmosphere.

The ionosphere is the limiting error source for temperature recovery at high stratospheric altitudes where the signature of the neutral atmosphere is weak. The resulting second-order errors in the two-frequency correction translate into a residual Doppler error profile for the occultation event, which can then be inserted into eq. (8.3) and numerically integrated to obtain the effect on $\delta\hat{N}(h)$ and on $\delta\hat{T}(h)$ through eq. (8.14). For a constant ionospheric-induced error on the measured Doppler, it follows from eqs. (8.21) and (D.6) that at an altitude in the neutral atmosphere of 40 km, the resulting error in the recovered temperature $\delta\hat{T}$ due to uncorrected second-order ionospheric effects, is approximately 0.6 K/mm/s Doppler error, or equivalently

$$\delta\hat{T} \approx \pm 0.1 \text{ (K)} \left[\frac{N_e}{10}\right]^2 \left[\frac{200}{\sigma}\right] \left[\frac{300}{h_o}\right]^2 @40 \text{ km} \qquad \text{(D.10)}$$

This can grow to several kelvins when high gradient conditions in electron density prevail at low ionospheric altitudes. Hence, the utility of using strategies that reduce the effect of second-order ionosphere errors on the Doppler measurement to below 1 mm/s is very high.

Clearly, the complexity of the actual ionosphere will result in more complicated expressions for second-order effects than our simple uniform shell model has provided. Simulations are helpful in this regard. Nevertheless, having realistic values for the bulk parameters of the ionosphere as provided by N_e, σ, and h_0, which give a measure of the mean transverse electron density gradient within the ionosphere and its extent along the ray path, provides good qualitative estimates of the second-order effects. They also can provide an estimate of the residual errors after some second-order correction strategy is applied in terms of the remaining uncertainties in the estimates of these bulk parameters.

Mitigation Strategies for Second-Order Ionospheric Effects

A number of techniques for attempting to correct for the second-order effects in GPS applications have been discussed in the literature. (See, for example, Gu and Brunner 1990; Brunner and Gu 1991; Yunck 1991; Bassiri and Hajj 1993; and Chapter 8 of this publication.) The local ionospheric models can be improved by using the two-frequency carrier phase and P-code pseudorange measurements prior to an ingressing occultation (or after an egressing one) to probe the ionosphere. By obtaining from the LEO observations the accurate temporal profiles of N_e in several directions and coupling these with other profiles that are nearby spatially and temporally, including those taken from ground tracking with the GPS global network, a useful estimate of the locally overlying ionospheric structure can be obtained.

The effect of the ionospheric-induced divergence in ray paths also can be partially compensated by sampling the L1 and L2 phases at slightly offset observational epochs to minimize the difference in second-order effects. We can obtain a rough estimate of the required magnitude of the offset by using the uniform shell model. The principal value for the rate of change of the total ionospheric bending angle is given by the first-

order term

$$\dot{\alpha}_{iono} \doteq \frac{2qN_e}{f^2R_E}\dot{\theta}_o\left[\frac{R_E}{2h_o}\right]^{3/2} \qquad (D.11)$$

where $\dot{\theta}_o$ is about ±1 mrad/s during an occultation. The residual second-order error in the bending angle after applying the two-frequency algorithm is

$$\delta\alpha_3 = -\frac{1}{\sigma R_E}\left[\frac{qN_e}{f_1 f_2}\right]^2\left[\frac{R_E}{2h_o}\right]^2 \qquad (D.12)$$

If an offset of $+\delta t/2$ is applied to the L1 phase measurement epoch and $-\delta t/2$ is applied to L2, we obtain the desired offset that nulls the second-order ionospheric effect (for the uniform shell model)

$$\delta t = -\frac{qN_e}{2\sigma f_1 f_2}\sqrt{\frac{R_E}{2h_o}} \Big/ 1.984\dot{\theta}_o \qquad (D.13)$$

which for the nominal values of h_o = 300 km and σ = 200 km, is about $12.5N_e$ ms. Of course, over the $12.5N_e$ ms interval, the L1 and L2 rays will have split in altitude in the neutral atmosphere at closest approach by an amount of $37.5N_e\zeta(a)$ m about a mean altitude; this should be accounted for by using the rate of change of the local refractivity gradient of the neutral atmosphere if the splitting becomes unacceptably large.

APPENDIX

E MISCELLANEOUS DERIVATIONS

Equations (7.17) and (7.20)

In polar coordinates the integral for the residual delay may be written as

$$I = \int n(r)\sqrt{1+r^2\theta'^2}\,dr \qquad (E.1)$$

where I is to be made stationary with respect to $\theta(r)$. The Euler equation from the Calculus of Variations yields a constant of integration

$$\frac{nr^2\theta'}{\sqrt{1+r^2\theta'^2}} = n_0 r_0 = a \qquad (E.2)$$

It follows that

$$I = \int \frac{n^2 r}{\sqrt{n^2r^2 - a^2}}\,dr \qquad (E.3)$$

The bending angle α is given by

$$\alpha = 2\int_{r_0}^{\infty}\theta'\,dr - \pi = 2\int_{r_0}^{\infty}\frac{dr}{r\sqrt{n^2r^2 - a^2}} - \pi \qquad (E.4)$$

Now π is given by

$$\pi = 2a\int_{a}^{\infty}\frac{d(nr)}{nr\sqrt{n^2r^2 - a^2}} \qquad (E.5)$$

Therefore, α is given by

$$\alpha = -2a\int_{r_0}^{\infty}\frac{n'}{n\sqrt{n^2r^2 - a^2}}\,dr \qquad (E.6)$$

which is eq. (7.20). Alternatively, from Snell's law

$$\frac{d\alpha}{dr} = -\frac{n'}{n}\tan\psi = -a\frac{n'}{n\sqrt{n^2r^2 - a^2}} \qquad (E.7)$$

where ψ is the angle of incidence of the ray with the local normal and, therefore, $\tan\psi = r\theta'$.

Equation (7.17) is obtained by rewriting eq. (E.3) as

$$I = \frac{1}{2}\int\frac{d(n^2r^2) - 2r^2nn'\,dr}{\sqrt{n^2r^2 - a^2}} \qquad (E.8)$$

$$= \sqrt{n^2r^2 - a^2} - \int\frac{(n^2r^2 - a^2 + a^2)n'\,dr}{n\sqrt{n^2r^2 - a^2}}$$

$$= \sqrt{n^2r^2 - a^2} - \int\frac{n'}{n}\sqrt{n^2r^2 - a^2} - a^2\int\frac{n'\,dr}{n\sqrt{n^2r^2 - a^2}}$$

and noting that the last integral above is just $a\alpha$ when the appropriate limits are applied.

Equations (7.25a), (7.25b), and (7.26)

To obtain eq. (7.25a) from eq. (7.20) with an exponential refractivity distribution as given by eq. (7.24), we use the fact that the Earth's atmosphere is thin, i.e., $a/\mathscr{R} \leq N_0 a/H << 1$ and, therefore, a straight-line approximation for the ray

should provide moderate accuracy. Moreover, let $\bar{L}$ be defined as the characteristic length of the ray path within the atmosphere, typically a few hundred kilometers for the Earth. Then another approximation uses the fact that $\bar{L}/a << 1$. Also, since $N_0 << 1$, we can replace n with 1 and r_0 with a. With these approximations we obtain

$$r \doteq r_0 + \frac{s^2}{2a} = r_0 + H\frac{x^2}{2}, \quad x = s/\sqrt{aH} \tag{E.9}$$

where s is arc length along the ray path. Replacing r with s in eq. (7.20) we can show that

$$\frac{adr}{\sqrt{n^2r^2 - a^2}} = (1 + O[\beta])ds, \quad \beta = N_0 a/H \tag{E.10}$$

and, therefore, that α is given to zero*th* order in β by

$$\alpha_0 = 2\sqrt{\frac{a}{H}}N(r_0)\int_0^\infty \exp[-x^2/2]dx = \sqrt{\frac{2\pi a}{H}}N(r_0) \tag{E.11}$$

We can improve on this by incorporating ray path bending in the form of a small departure z from the straight-line path. Now

$$\frac{d^2z}{ds^2} \doteq \frac{d\alpha}{ds}, \quad z(0) = z'(0) = 0 \tag{E.12}$$

and it follows that an expression for z to first order in β is given by

$$z_1 = \beta H\left(\frac{x}{2}\int_o^x \exp[-x'^2/2]dx' - (1 - \exp[-x^2/2])\right)$$
$$= \beta H\frac{x^2}{2} + \cdots \tag{E.13}$$

With this expression for z, we can replace r in eq. (E.9) with

$$r = r_0 + H\frac{x^2}{2} - z_1 \tag{E.14}$$

to obtain a more accurate expression for α that is good to the first order in β. Using eq. (E.14), it can be shown that

$$\frac{adr}{\sqrt{n^2r^2 - a^2}} = (1 + O[\beta^2])ds \tag{E.15}$$

Carrying out the integrations we obtain

$$\alpha_1 = \sqrt{\frac{2\pi a}{H}}N(r_0)\left[1 + (\sqrt{2} - 1)\beta\right] \tag{E.16}$$

The expression for z_1 can be used to obtain a second-order expression that is given by

$$z_2 = z_1 + H\beta^2\left(\left[\frac{1}{2}\int_0^x e^{-x'^2/2}dx'\right]^2 - x\int_0^x (2e^{-x'^2} - e^{-x'^2/2})dx' - e^{-x'^2} + e^{-x'^2/2}\right) \tag{E.17}$$

which, in turn, can be used to obtain a second-order expression for α

$$\alpha_2 = \sqrt{\frac{2\pi a}{H}}N(r_0)\left[1 + (\sqrt{2} - 1)\beta + 0.28\beta^2\right] \tag{E.18}$$

Equation (8.66)

Here we assume that the ray path has small excursions from the x axis, and that $y'^2 << 1$. In this case the simplified Euler equation for y is given by eq. (E.12), with x replacing s and with the boundary conditions: $y(0) = 0$, $y(R_{LG}) = 0$. The solution is

$$y(x) = \int_0^x (x - x')\frac{\partial n}{\partial y}dx' - \frac{x}{R_{LG}}\int_0^x (R_{LG} - x')\frac{\partial n}{\partial y}dx' \tag{E.19}$$

and at $x = R_{LG}$, y' is given by

$$y'(R_{LG}) = \frac{1}{R_{LG}} \int_0^{R_{LG}} x' \frac{\partial n}{\partial y}\, dx' \qquad \text{(E.20)}$$

We can think of R_{LG} as a variable that generates a one-parameter family of stationary phase ray paths that all start from x = 0 on the x axis and intersect the x axis at the point (0,x). From the transversality condition in the Calculus of Variations (see eq. (B.19)), the change in the stationary value of the delay integral as a result of moving the right-hand endpoint of the ray path along the x axis from the point (0,0) to (0,x) is given by

$$\Delta I = \int_0^x \frac{n(x',0)}{\sqrt{1+y'^2}}\, dx' \doteq \int_0^x n(x',0)\left(1-\frac{1}{2}y'^2\right) dx' \qquad \text{(E.21)}$$

where y' is the slope of the ray path intersecting the x axis at the point (0, x) and given by differentiating eq. (E.19). The integral in eq. (E.21) is known in its general form as the integral formula of Weierstrass.

Equation (E.21) gives the difference in values of the delay integral obtained from following the actual stationary phase path versus that obtained from following the straight-line path along the x axis. Notice that the difference is second order in y', as it must be if the ray path is a path of stationary phase, and that the delay along the true ray path is less than that obtained in following the path along the x axis. Setting $x = R_{LG}$ in eq. (E.21) and using eq.(E.20) for y', we obtain

$$\Delta I = \int_0^{R_{LG}} n(0,x)dx - \frac{1}{2}\int_0^{R_{LG}} \frac{1}{x^2}\left[\int_0^x \frac{\partial n}{\partial y} x'\, dx'\right]^2 dx \qquad \text{(E.22)}$$

For small excursions of the ray path from the x axis, the term $\partial n/\partial y$ can be evaluated on the x axis; for better accuracy it should be evaluated along the ray path.

Equations (11.39) and (11.40)

The Fresnel transform pair is obtained by multiplying eq. (11.39) by $\exp\left[-(i\pi/2)(u_b - v')^2\right]$ and integrating on u_b to obtain

$$\begin{aligned}
&\frac{1}{1+i}\int_{-\infty}^{\infty} \exp\left[-\frac{i\pi}{2}(u_b - v')^2\right] \int_{-\infty}^{\infty} A(v)\exp\left[i\Psi(v) + \frac{i\pi}{2}(u_b - v)^2\right] dv\, du_b \\
&= \frac{1}{1+i}\int_{-\infty}^{\infty} A(v)\exp\left[i\Psi(v) + \frac{i\pi}{2}(v^2 - v'^2)\right] dv \int_{-\infty}^{\infty} \exp[i\pi(v'-v)u_b]\, du_b \\
&= \frac{2}{1+i} A(v')\exp[i\Psi(v')]
\end{aligned} \qquad \text{(E.23)}$$

The last integral in eq. (E.23) is

$$\int_{-\infty}^{\infty} \exp[i\pi(v'-v)u_b]\, du_b = 2\delta(v'-v) \qquad \text{(E.24)}$$

where $\delta(x)$ is the Dirac delta function. Equation (11.40) follows from this transformation.

Equation (11.46)

The resolution Δv is defined by eq. (11.44) where $R(v,v^*)$ is the recovered impulse function defined by eqs. (11.41) and (11.42). Using the expression for $R(v,v^*)$ given in eq. (11.45), the top integral in eq. (11.44) is given by

$$\begin{aligned}
\int_{-\infty}^{\infty} |R(v,v^*)|^2 dv &= \frac{1}{4}\int_{-\infty}^{\infty} dv \int_{-U_b}^{U_b}\int_{-U_b}^{U_b} \exp[i\frac{\pi}{2}\left(\gamma^2(u_b^2 - u'^2_b)\right. \\
&\quad \left. -2\gamma^2(u_b - u'_b)v - 2(u_b - u'_b)(v^* - v)\right)] du_b du'_b \\
&= \frac{1}{4}\int_{-U_b}^{U_b}\int_{-U_b}^{U_b} \exp[i(\pi/2)\left(\gamma^2(u_b^2 - u'^2_b)\right. \\
&\quad \left. -2(u_b - u'_b)v^*\right)] du_b du'_b \int_{-\infty}^{\infty} \exp\left[i\pi(\gamma^2 - 1)(u'_b - u_b)v\right] dv
\end{aligned} \qquad \text{(E.25)}$$

However, the last integral is given by

$$\int_{-\infty}^{\infty} \exp\left[i\pi(\gamma^2 - 1)(u'_b - u_b)v\right] dv = \frac{2}{\gamma^2 - 1}\delta(u'_b - u_b) \qquad \text{(E.26)}$$

Therefore,

$$\int_{-\infty}^{\infty} |R(v,v^*)|^2 dv = \frac{U_b}{\gamma^2 - 1} \qquad \text{(E.27)}$$

Now $R(v^*, v^*)$ becomes

$$R(v^*,v^*) = \frac{1}{2}\int_{-U_b}^{U_b} \exp\left[i\frac{\pi}{2}\gamma^2\left((u_b - v^*)^2\right)\right] du_b$$
$$\equiv \gamma[\mathscr{C}(U_b) + i\mathscr{S}(U_b)], \quad v^* << U_b \qquad \text{(E.28)}$$

By multiplying eq. (E.28) by its complex conjugate and combining with eq. (E.27), we obtain the result given in eq. (11.46).

Equation (11.87)

The thin-screen model mimics the propagation effects of the actual atmosphere on the signal by assigning a profile to the phase terms $\psi(h)$ and $\varepsilon(h)$, embedded in the thin screen, that attempts to achieve the correct geometric optics phase delay and doppler observed by the LEO. The reference phase term $\psi(h)$ is derived from the nominal bending angle profile through the stationary phase condition. The perturbation term $\varepsilon(h)$ is similarly derived through eq. (11.86), where $\delta\alpha(r)$ is the departure from the nominal bending angle profile due to some sort of refraction anomaly in the atmosphere. For a spherically symmetric atmosphere with a single boundary across which a discontinuity in the refractivity or in one of its derivatives occurs, eq. (11.87) provides an integral relation that yields $\delta\alpha(r)$. For the case of a discontinuous lapse rate, we need to develop an explicit form for $\delta\alpha(r)$. Let $\alpha[r,\gamma]$ describe the bending angle that would be observed if γ were constant throughout the atmosphere. Then from eq. (11.87)

$$\alpha^{\pm}(r) = \alpha[r, \gamma^{\pm}] \qquad \text{(E.29)}$$

We need an expression for $\alpha[r,\gamma]$ and for the integral differences in eq. (11.87).

From the assumption of hydrostatic equilibrium and the use of the perfect gas law, we have

$$-\frac{1}{\rho}\frac{d\rho}{dr} = H_\rho^{-1} = H_p^{-1} + \gamma, \quad H_p^{-1} = \mu g/kT \qquad \text{(E.30)}$$

Let the temperature vary linearly with altitude and be given by

$$T = T_0[1 + \gamma(h - h_0)] \qquad \text{(E.31)}$$

Inserting eq. (E.31) into eq. (E.30) and integrating, we obtain for the density

$$\ln\left[\frac{\rho}{\rho_0}\right] = -\frac{1}{\gamma H_{\rho_0}}\ln[1 + \gamma(h - h_0)], \quad H_{\rho_0}^{-1} = H_{p_0}^{-1} + \gamma \qquad \text{(E.32)}$$

where ρ_0 is the density at h_0. If we assume that $N = N_0\rho/\rho_0$, eq. (E.32) can be used in eq. (E.6) to obtain a modified expression for α that explicitly contains terms in γH_{p_0}. Using similar approximations that were used to derive eqs. (E.11), (E.16), and (E.18), we can expand in a power series in β and in γH_{p_0} to obtain

$$\alpha = \sqrt{\frac{2\pi a}{H_{p_0}}}\, N(r_0)\left[1 + 0.41\beta + \frac{3}{8}(\gamma H_{p_0}) + 0.28\beta^2 + 0.37\beta(\gamma H_{p_0}) - \frac{7}{128}(\gamma H_{p_0})^2 + \cdots\right] \qquad \text{(E.33)}$$

Defining K by

$$K = \frac{1}{\alpha H_{p_0}}\frac{\partial\alpha}{\partial\gamma} \qquad \text{(E.34)}$$

one obtains

$$K = \frac{3}{8} + \left(\frac{31\sqrt{2} - 32}{32}\right)\left(\frac{aN_0}{H_{p_0}}\right) - \frac{1}{4}(\gamma H_{p_0}) + \cdots \qquad \text{(E.35)}$$

For the tropopause, K has a value of about 0.39 for $\gamma = 0$. To obtain an explicit expression for $\delta\alpha(r)$ we linearize $\alpha[r,\gamma^+] - \alpha[r,\gamma^-]$ using the form

$$\alpha[r,\gamma^+] - \alpha[r,\gamma^-] \doteq \alpha K(\Delta\gamma H_{p_0}) \qquad \text{(E.36)}$$

and we use similar approximation techniques to evaluate the integrals in eq. (11.87) that were used to obtain eq. (E.11). It can be shown that to zero*th* order in β one obtains

$$\delta\alpha(h) \doteq (\Delta\gamma H_{p_0})\alpha^-(r_0)\left[\left(K + \frac{3\sigma^2}{2} + \frac{\sigma^4}{2}\right)F(\sigma) - \frac{1}{\sqrt{\pi}}\left(\frac{5\sigma}{4} + \frac{\sigma^3}{2}\right) - K\exp[\sigma^2]\right], \quad h \le h_0 \qquad \text{(E.37)}$$

where $F(\sigma)$ is given by

$$F(\sigma) = \exp\left[\sigma^2\right][1 - \text{erf}\,[\sigma]] \qquad \text{(E.38)}$$

and σ is given by

$$\sigma = \sqrt{\frac{h_0 - h}{H_{p_0}}} \qquad \text{(E.39)}$$

The phase perturbation $\varepsilon(h)$ follows from eq. (11.86), which can be expanded to yield

$$\varepsilon(h) \doteq -k(\Delta\gamma H_{p_0})\alpha^-(r_0)H_{p_0}\left[\frac{4\sigma^3}{3\sqrt{\pi}} - \frac{3\sigma^4}{4} + \frac{8\sigma^5}{5\sqrt{\pi}} - \frac{2\sigma^6}{3} + \cdots\right], \quad h \le h_0 \qquad \text{(E.40)}$$

The method stationary phase (Born and Wolf 1980) is an asymptotic approximation technique for evaluating integrals of complex exponential functions such as that given by eq. (11.39). It uses the fact that the contributions to the integral from the rapidly oscillating integrand largely cancel out except in neighborhoods where the argument of the exponential function is stationary, or at the endpoint of the integration where the variation of the argument is minimal. One expands the argument in a Taylor series about the stationary point or the endpoint, retaining only terms through a quadratic; the resulting integrals are then expressable, in the case of eq. (11.39), in terms of Fresnel integrals.

When the gradient of $\delta\alpha(r_0)$ is sufficiently positive, it creates a caustic surface below r_0, which results in multiple stationary phase altitudes in the thin screen. For the case of discontinuous lapse rate, the gradient of $\delta\alpha$ becomes infinite at r_0 and a caustic or, equivalently, a multipath situation will arise if $\Delta\gamma > 0$. A condition for the existence of a caustic surface in the neighborhood about a point u_b^* is $d\bar{u}(u_b)/du_b \to \infty$ at that point. Figure E-1 shows that point and the relationship between the altitudes of the stationary phase points in the thin screen and h_b, which is defined through eqs. (11.90) and (E.37).

It is noted that $\varepsilon''(u)$ has a $\sigma^{-1/2}$ singularity at $\sigma = 0$, which enters through $\zeta^-(u_0)$ in the evaluation of the first Fresnel integral in eq. (11.93). However, $\zeta^-(u_0)$ arises from an integral average using the stationary phase approximation; therefore, an average value of ζ^- over a small neighborhood $u < u_0$ should suffice.

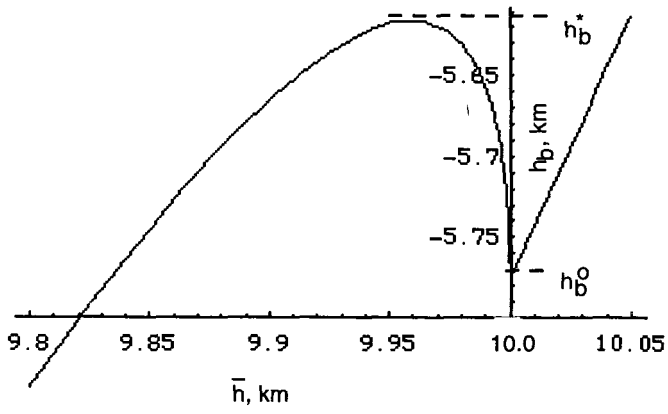


Figure E-1. Multipath geometry for $\Delta\gamma > 0$.

Figure E-2 shows the behavior of $\varepsilon(h)$ for typical values of the relevant atmospheric parameters at the tropopause. In the vicinity of u_0 the average value of $\varepsilon''(u)$ (in Fresnel units) is about $-34(\Delta\gamma H_{p_0})$ rad.

Figures E-3 and E-4 show the LEO-observed Fresnel perturbations in phase and amplitude for a range of $\Delta\gamma$ values, based on eqs. (11.88) and (11.93). Using the asymptotic forms for U(x) and V(x) given in Chapter 11, footnote 3, one can show that the asymptotic form for eq. (11.88) is

$$\tilde{E}\exp[i\delta\varphi] \to \sqrt{\frac{\zeta^-(\bar{u}^-)}{\zeta^+(\bar{u}^+)}}\exp[i\Delta^-(\bar{u}^-)] \times\left\{1-\frac{1+i}{2\pi}\left[\frac{\exp[i\pi(w^-)^2]}{w^-}\right]\right\}-\frac{1+i}{2\pi}\frac{\exp[i\pi(w^+)^2]}{w^+}+O[w^{-3}] \tag{E.41}$$

and that eq. (11.93) becomes

$$\tilde{E}\exp[i\delta\varphi] \to 1+\frac{1-i}{2\pi^2}\frac{\exp[i\pi(w^+)^2/2]}{(w^+)^3}\left[1-\frac{\zeta_o^-}{\zeta^+}\right]+\cdots \tag{E.42}$$

Except within the central region, a distance of one or two Fresnel zones about the altitude of the discontinuity in lapse rate, these asymptotic expressions provide an excellent representation of the Fresnel processes, including many of the fine structure features and trends exhibited in Figures E-3 and E-4. We note that the Δ^- term, which has two branches within the caustic zone, scales essentially linearly with $\Delta\gamma$, but the altitudes over which multipath is present scale essentially quadratically with $\Delta\gamma$.

The problem of recovery of the values of $\Delta\gamma$ and r_0 from the phase and amplitude measurement sequences is related to the problem of how to isolate the slope (Doppler) of Δ^- from the background Fresnel noise, as well as from additional atmospheric noise arising from turbulence and

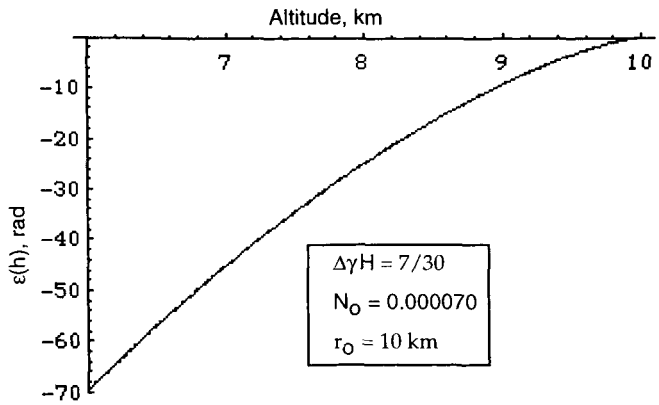


Figure E-2. Thin screen phase perturbation.

other spatial inhomogeneities in refractivity; also there is thermal noise from the receiver and a likely undersampling problem. The Fresnel phase signature from a discontinuous lapse rate is very small compared to the nominal phase profile and can easily be masked by these noise sources. Figure E-5 shows the Fresnel perturbation in LEO-observed phase over a short vertical distance (~200 m) about the discontinuity. On the other hand, Figure E-6 shows the total change in LEO-observed phase over a similar range. The ratio of the perturbed to total phase is roughly proportional to $\Delta\gamma H$.

The Fresnel patterns in Figures E-3 and E-4 can be used to construct Fresnel filters which, when applied to the measurement streams in phase and amplitude, could aid in detecting boundary layer transitions. To obtain a corresponding altitude in the real atmosphere, the altitude h in the thin screen should be reduced by eq. (11.36) to correct for ray path bending. Finally, it should be noted again that the thin screen model is a planar approximation to a three dimensional propagation process involving diffraction effects from a boundary surface. A study of the differences between thin screen and exact results should be undertaken to assess the limits of validity of the thin screen model.

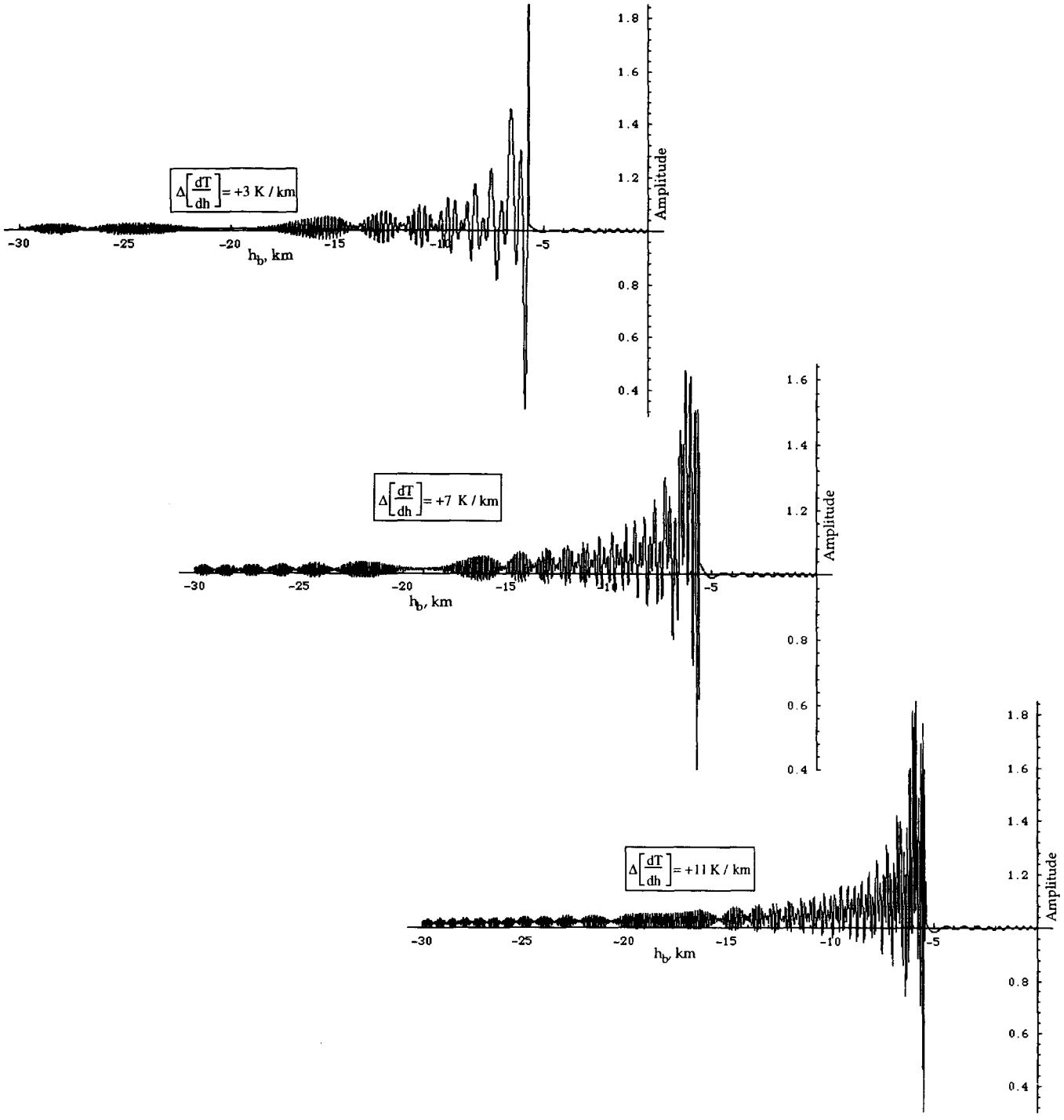


Figure E-3. Fresnel perturbation in received signal amplitude due to a discontinuity in lapse rate at an altitude of h_o =10 km, based on the thin-screen model using the method of stationary phase. At a 10-km altitude, a nominal refractive bending angle of 5.4 mrad, a pressure scale height of 7 km, a temperature of 210 K, and a dry exponential atmospheric model were used as a reference. The variable h_b is the altitude of the straight-line path between the LEO (at a 700-km altitude) and the GPS satellite at its point of tangency with the Earth's limb. At h = 10 km, the nominal value of h_b is -5.78 km. The ray path altitudes covered by these plots are ~4.5 to 13 km. To convert the abscissa scale into a time interval in seconds, divide by ~3 for a vertically descending ray path. Equation (11.36) should be applied to the altitude h in Figures E-3 and E-4 to correct for ray path bending.

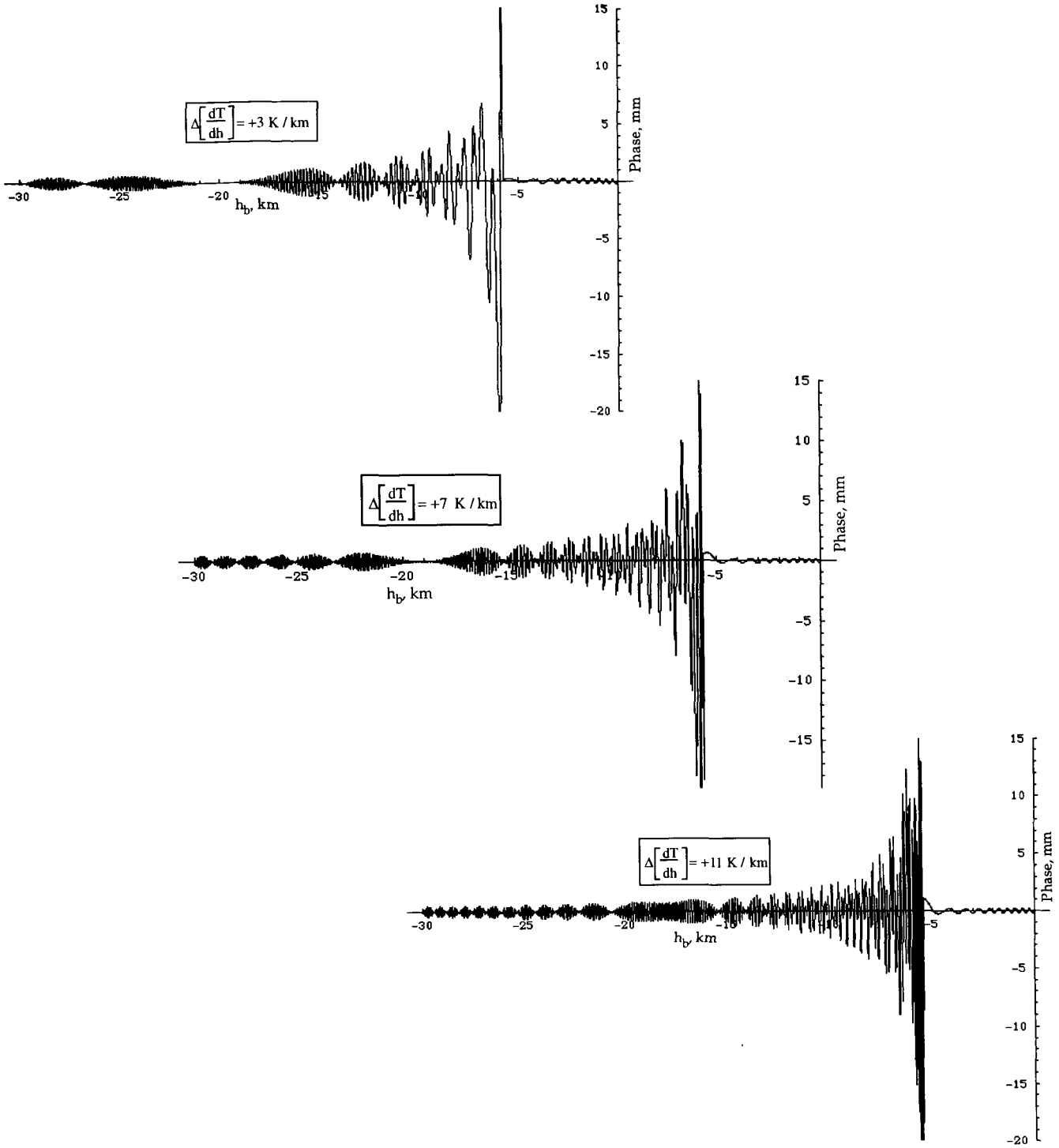


Figure E-4. Fresnel perturbation in a received signal phase (L1) due a discontinutiy in lapse rate at an altitude of h_o = 10 km, based on the thin screen model using the method of stationary phase to evaluate the Fresnel integrations. Geometric optics phase term $(\doteq \varepsilon(\bar{h}^-) + \pi(D/\lambda)\zeta^+(\bar{h}^-)\delta\alpha^2(\bar{h}^-)$ if $\bar{h}^- \le h_o$, else $0)$ ***has been removed. At a 10-km altitude, a nominal refractive bending angle of 5.4 mrad, a density scale height of 7 km, a temperature of 210 K, and a dry exponential atmospheric model were used as a reference. The variable h_b is the altitude of the straight-line path between the LEO (at a 700-km altitude) and the GPS satellite at its point of tangency with the Earth's limb. At h = 10 km, the nominal value of h_b is -5.78 km. To convert the abscissa scale into a time interval in seconds, divide by ~3 for a vertically descending ray path.***

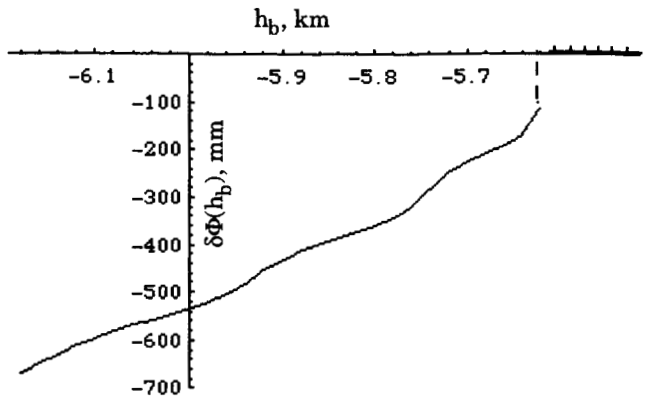


Figure E-5. Perturbation in LEO-Observed Phase.

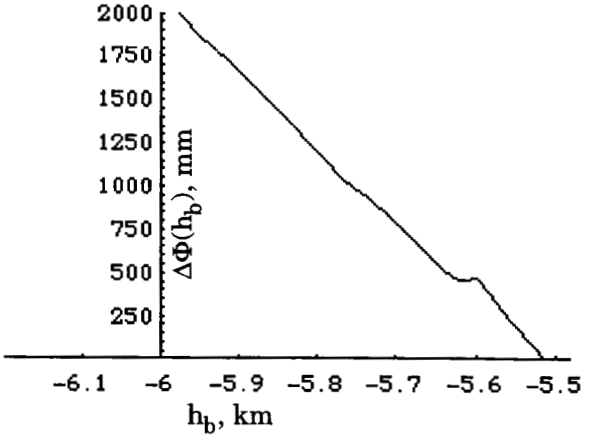


Figure E-6. Total LEO-Observed Phase

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 16, 1994	External

4. TITLE AND SUBTITLE

The Application of Spaceborne GPS to Atmosphere Limb Sounding and Global Change

5. FUNDING NUMBERS

B 05510010997

6. AUTHOR(S)

W.G. Melbourne, E.S. Davis, C.B. Duncan, G.A. Hajj, K.R. Hardy, E.R. Kursinski, T.K. Meehan, L.E. Young, T.P. Yunck

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109

8. PERFORMING ORGANIZATION REPORT NUMBER

JPL Publication 94-18

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NAS7-1260

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This monograph is intended for readers with minimal background in radio science who seek a relatively comprehensive treatment of the mission and technical aspects of an Earth-orbiting radio occultation satellite. Part I (Chapters 1-6) describes mission concepts and programmatic information: Part II (Chapters 7-12) deals with the theoretical aspects of anlyzing and interpreting radio occultation measurements. In this mission concept the navigation signals from a Global Positioning System (GPS) satellite that is being occulted by the Earth's limb are observed by a GPS flight receiver on board a low Earth orbiter (LEO) satellite. As the received signal traverses deeper layers of the atmosphere, its amplitude and phase are progressively altered through atmospheric refraction and possibly interference arising from spatial irregularities in refractivity. The signal delay and amplitude profiles recorded by the LEO during the occultation, which typically lasts about 1 min, are used to recover profiles of atmospheric refractivity, pressure, and temperature from the surface up to the upper stratosphere at heights of 50-60 km. Below the tropopause, water vapor sharply limits the accuracy of temperature and pressure recovery. Nominal profiles allow recovery of valuable information.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Global Positioning System; low Earth orbit; radio science; meteorology; radio occultation	142
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified			