

Chapter 1

Radio Occultation Using Earth Satellites

Background and Overview

1.1 Introduction

This monograph is concerned with the phase and amplitude of an electromagnetic wave during a radio occultation episode. This wave has passed through an intervening medium from a distant emitter and arrives at a receiver. The receiver measures the phase and amplitude of the wave over the duration of the occultation episode. These measurement sequences can be used to infer physical properties about the intervening medium.

Radio occultation refers to a sounding technique in which a radio wave from an emitting spacecraft passes through an intervening planetary atmosphere before arriving at the receiver. The words “occultation,” “occulted,” or “occulting” imply that the geometry involving the emitter, the planet and its atmosphere, and the receiver changes with time. Although an occulting or eclipsing planet (or moon) usually is involved, the word has come in recent times to also include non-occulting events, for example, satellite-to-satellite sounding through the ionosphere or receiving a reflected wave from a reflecting surface. From the perspective of the receiver for the strictly occulting case, the emitter is seen to be either rising or setting with respect to the limb of the occulting planet. As the radio wave from the emitter passes through the intervening atmosphere, its velocity and direction of propagation are altered by the refracting medium. The phase and amplitude of the wave at the receiver consequently are altered relative to their values that would hold without the intervening medium or the occulting planet. As time evolves, profiles of the phase variation and the amplitude variation at the receiver are generated and

recorded by the receiver. These profiles provide information about the refractive properties of the intervening medium.

In seismology, an array of seismometers spread over some geographical extent is used to study the various types of seismic waves arriving at each seismometer from a remote earthquake. With the array, one can measure the differential arrival times at the different stations in the array and also the spectral properties of the various waves. The different paths followed by these various waves and certain physical properties of the medium through which they pass can be inferred from these observations. In a radio occultation, it is the kinematics of the emitter/receiver pair over the duration of an occultation episode that provides analogous information.

1.1.1 History of the Occultation Technique

An astrometric observation in the eighteenth century measuring the times of ingress and egress of a lunar occultation of a star was probably among the first scientific applications of the technique. The method of lunar distances, the common seaman's alternative to the relatively expensive chronometer in the eighteenth and nineteenth centuries for keeping accurate Greenwich time for longitude determination at sea, and even for calibrating marine chronometers over very long voyages, depended crucially on an accurate lunar ephemeris. This in turn depended on accurate astrometric observations of the moon relative to the background stars and a good dynamical theory for the lunar orbit. Much later, the limb of the moon has been used as a knife-edge to obtain the microwave diffraction pattern from certain quasars [1]. The fringe spacing and amplitude of this diffraction pattern provide information about the angular distribution of radiant intensity from these very compact radio sources. Also, planetary atmospheres have been studied by analyzing stellar refraction and scintillation effects that occur during the ingress and egress periods when the star is occulted by the planet [2–4].

The radio occultation technique to sound planetary atmospheres using spacecraft began almost at the dawn of the era of planetary exploration. The first spacecraft to Mars in 1964, Mariner 4, flew along a trajectory that passed behind Mars as viewed from Earth [5,6]. The extra carrier phase delay and amplitude variation observed on the radio link between Mariner 4 and the Earth-based radio telescopes as Mariner 4 passed behind Mars and emerged from the other side provided valuable density information about its very tenuous atmosphere and also about its ionosphere [7]. Since then a score of experiments involving planetary missions have been undertaken to study the atmospheres of almost all of the planets in the solar system, including several moons and the rings of Saturn [8–12].

1.1.2 Occultations from Earth Satellites

Sounding of the Earth's atmosphere and ionosphere using the occultation technique theoretically can be performed with any two cooperating satellites. Prior to the Global Positioning System (GPS) becoming operational, a few early radio occultation experiments from a satellite-to-satellite tracking link had been conducted. These included the radio link between the Mir station and a geostationary satellite [13] and between GEOS-3 and ATS-6 [14]. However, this monograph focuses primarily on the carrier phase and amplitude measured by a GPS receiver onboard a low Earth orbiting spacecraft (LEO) while tracking the navigation signals emitted by a GPS satellite during its occultation by the Earth's limb [15]. GPS/MET (Global Positioning System/Meteorology), an occultation experiment that flew on MicroLab-1 and launched in 1995, was the first occultation experiment using the GPS [16–19]. Although experimental, over 11,000 occultations were used from GPS/MET to recover refractivity, density, pressure, temperature, and water vapor profiles [19]. GPS/MET provided a definitive engineering proof-of-concept of the occultation technique, and its data set became an experimental platform for implementing improved tracking and data processing schemes on subsequent Earth satellites with GPS occultation capability. GPS/MET also provided a basis for assessing the scientific and societal value of the technique in such diverse applications as meteorology, boundary layer studies, numerical weather prediction (NWP), and global climate change. Since then the Challenging Minisatellite Payload (CHAMP) (2001) [20], Satellite de Aplicaciones Cientificas-C (SAC-C) (2001), and Gravity Recovery and Climate Experiment (GRACE) (2002) satellites have been launched, from which radio occultation observations are now more or less continually made [21]. These missions alone could return nearly 1000 occultations per day. Future operational missions are planned, such as the Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) configuration of occultation-dedicated LEOs to be launched in 2005 [22–24]. This system will provide near-real-time sounding information from about 4000 globally distributed occultations per day, which will be assimilated into NWP programs. Also, other global navigation satellite systems (GNSSs), such as the Russian Global Navigation Satellite System (GLONASS), and future systems, such as the planned European system, Galileo, will broaden the opportunities for dedicated satellite-to-satellite occultation missions [25,26].

Figure 1-1 depicts in exaggerated form a typical occultation scenario involving a LEO and the GPS satellite constellation. For a setting occultation, the about-to-be occulted GPS satellite will be seen from the LEO to be setting with respect to the limb of the Earth. The duration of a typical analyzed occultation through the neutral atmosphere (from roughly 100-km altitude to sea level) is less than 100 s. A ray from this GPS satellite passes through the

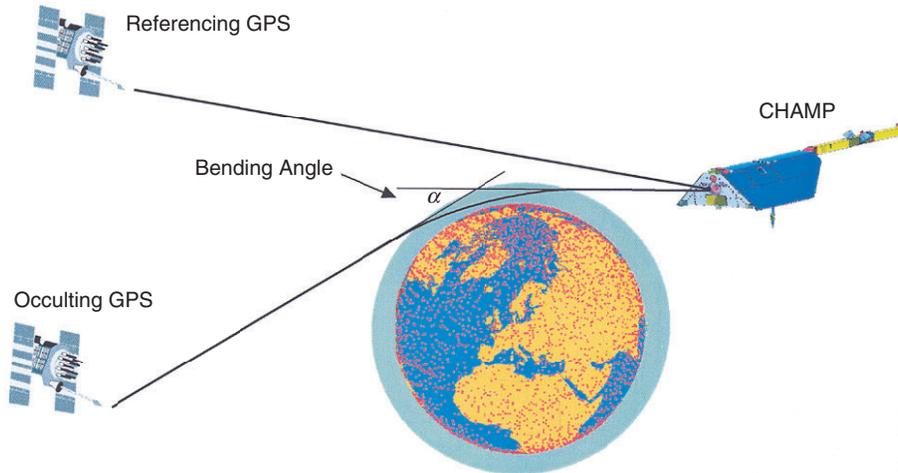


Fig. 1-1. Occultation scenario for a low Earth orbiter. From [91].

upper layers of the Earth's atmosphere at a near-horizontal rate of descent, skimming a deepest layer at its tangency point. A ray in this context may be defined by the normal to the cophasal surface of the carrier wave emitted by the GPS satellite. After the tangency point, the ray begins its near-horizontal rate of ascent through the upper layers, exits the atmosphere, and continues on to the LEO. The maximum total refractive bending of the ray is very small, about 1 deg from dry air. Additional refraction from water vapor, especially for tangency points in the lower troposphere, can double or triple that bending angle. A small bending-angle profile is a hallmark of a "thin" atmosphere. The index of refraction of dry air at microwave frequencies is about the same for light. At sunset the apparent sun for an observer on the ground is refracted through its own diameter, about 1/2 deg. The secular trend in the refractive gradient of dry air with altitude (which is near-exponential) is evident in the oblate shape of the apparent sun as its lower limb touches the horizon. Rays from the bottom limb usually are bent more than rays from the top limb. The fractured shape of the solar disk at some sunsets is caused by abrupt departures of the refractivity profile at low altitudes from the secular trend.

The ray arriving at the LEO from the occulted GPS satellite may not be unique, and indeed may not even exist in a geometric optics context. But for our purpose, we assume in Figure 1-1 that it does exist and that it is unique at the epoch of the observation. The inclination of the ray to the local geopotential surface is very slight; at 100 km from the tangency point it is about 1 deg in dry air, increasing linearly with ray path distance from the tangency point. At 400 km from the tangency point, the ray is about 10 km higher in altitude. Because of the near-exponential decrease in dry air density with height, it follows that most of the information about the atmosphere at a given epoch is

contained in a relatively narrow section of the ray about its tangency point, a few hundred kilometers in length [27].

As time evolves, the tangency point of the ray arriving at the LEO drifts deeper on average into the atmosphere for a setting occultation. Typically for the occultations selected for analysis, the initial cross-track angle of descent of the tangency point relative to the local vertical is less than 30 deg. The excess phase delay observed at the LEO, which is simply the extra phase induced by the refracting medium, will continue to increase on average over the course of the occultation because of the increasing air density with depth. This continues until the tangency point of the ray nears or contacts the planetary surface. The total excess delay can exceed 2 km with refractive bending angles up to 4 deg near the surface when water vapor is in abundance and the vertical gradient of its density is large. Defocusing and multipath, which tend to become strong in the lower troposphere, may reduce the signal amplitude to below a detection threshold before the ray contacts the surface, terminating the occultation episode several seconds or sometimes tens of seconds prematurely. However, sometimes the signal returns several seconds later, and sometimes in very smooth refractivity conditions even a knife-edge diffraction pattern from the limb or interference fringes from an ocean reflection are observed in the phase and amplitude before the direct signal from the GPS satellite is completely eclipsed [28,29]. On average, over the entire globe 80 percent of the CHAMP occultations reach to within 1 km of the surface, and 60 percent reach to within 1/2 km [21]. These encouraging statistics should improve in the future as new signal-tracking algorithms are implemented in the GPS receivers onboard the LEOs.

Figure 1-2 shows results from an early occultation from the GPS/MET experiment [18]. Here the excess phase delay of the L1 carrier in meters and its time derivative, excess Doppler in hertz, are shown over the last 90 seconds of the occultation. The lower abscissa shows coordinated Universal time (UTC), and its scale is linear. The upper abscissa shows the altitude of the ray path tangency point, and its scale is non-linear. The refractive gradient of the atmosphere increases markedly with depth, which effectively slows the average rate of descent of the tangency point because the refractive bending angle of the ray increases on average with depth. Therefore, the LEO must travel increasingly farther along its orbit to intercept these progressively deeper penetrating and more refracted rays. Near the Earth's surface, the average rate of descent of the tangency point typically is an order of magnitude smaller than it is in the upper atmosphere.

Figure 1-3 shows the amplitude of the L1 carrier for the same occultation [18]. The ordinate is the signal-to-noise ratio (SNR) in voltage, SNR_V , that would apply if the individual L1 carrier amplitude measurements were averaged over 1 s. The actual sample rate in this figure is 50 Hz. The averaging

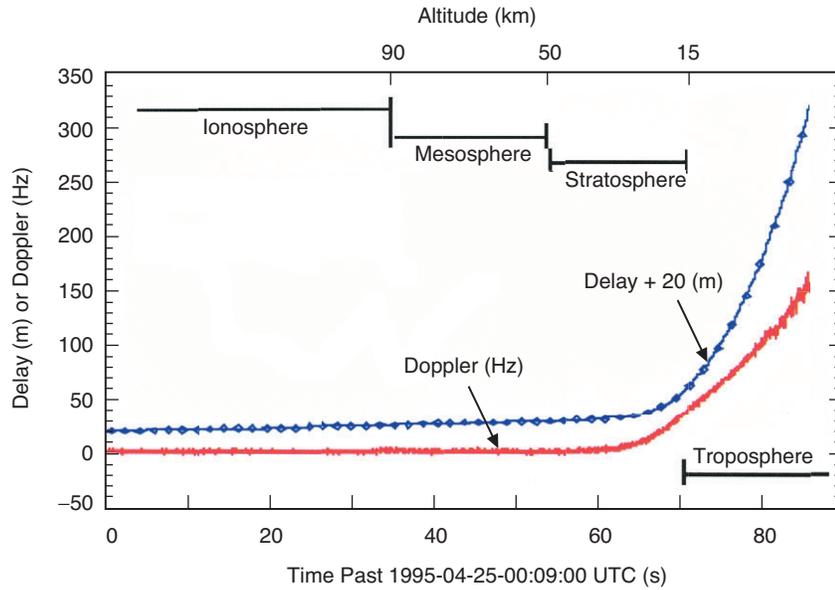


Fig. 1-2. Profile of excess Doppler and phase for a particular occultation of GPS PRN no. 28 observed by the GPS/MET experiment on MicroLab-1 on April 25, 1995, near Pago Pago. Redrawn from [18].

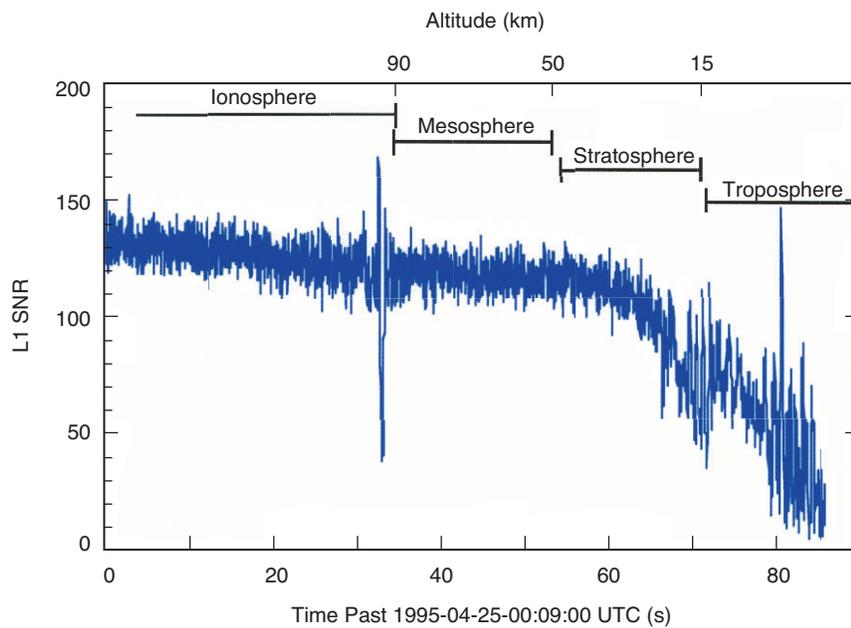


Fig. 1-3. Observed voltage SNR of the L1 carrier versus time for the same occultation shown in Fig. 1-2. Redrawn from [18].

time is 20 ms. Therefore, the thermal noise ($\sim 1/\text{SNR}_V$) for the points in this figure is $\sqrt{50}$ larger than the SNR_V values would imply. This figure clearly shows the secular defocusing caused by the increasing refractive gradient with depth. The refractive gradient disperses the directions of progressively deeper rays after exiting the atmosphere; the gradient effectively “de-collimates” the rays, thereby diluting their power at the LEO by spreading it over a larger area. The ratio of a small “collimated” area proportional to Δa , shown in Fig. 1-4(a), to the resulting de-collimated area proportional to $\Delta\sigma$ at the LEO is the defocusing factor. Appendix A derives a simple form for this factor:

$$\frac{\Delta a}{\Delta\sigma} = \zeta = \left(1 - D \frac{d\alpha}{da}\right)^{-1} \quad (1.1-1)$$

where D is effectively the distance of the LEO from the Earth’s limb and $d\alpha/da$ is the radial gradient of the refractive bending angle α . Figure 1-4(b) shows a relatively mild multipath scenario, including a shadow zone where $d\alpha/da$ is temporarily larger. Both the shadow zone and the interference from multipath waves cause the variability in SNR. The secular trend of SNR_V in Figure 1-3 is effectively given by $\zeta^{1/2}(\text{SNR}_V)_o$, where $\zeta^{1/2}$ is the defocusing from air and $(\text{SNR}_V)_o$ is the voltage SNR of the GPS signal that would be received at the LEO without the planet and its intervening atmosphere, the so-called “free-space” value. In addition to secular defocusing effects, this figure shows the strong transients in signal amplitude as the point of tangency of the ray passes through certain narrowly defined horizontal layers of the lower ionosphere and when it crosses the tropopause. These transients signal the presence of multipath-induced interference between different rays arriving concurrently at the LEO from different levels in the ionosphere and atmosphere. As the tangency point of the main ray cuts further down through successive layers of the atmosphere and into the middle and lower troposphere, the prevailing interference evident in this figure likely is induced by variable water vapor concentrations.

These transients observed in phase and amplitude raise a number of issues, some of which have been better dealt with than others. Significant progress on dealing with multipath has been made using back propagation and spectral techniques. Deep troughs in amplitude, shadow zones, and super-refractivity episodes, as well as caustics, are difficult for ray theory. How well these techniques work when the validity of ray theory itself is being strained is still to be established. The low SNR associated with many of these transients also makes it difficult to maintain connection in the LEO-observed phase measurements across them. We return to these topics later.

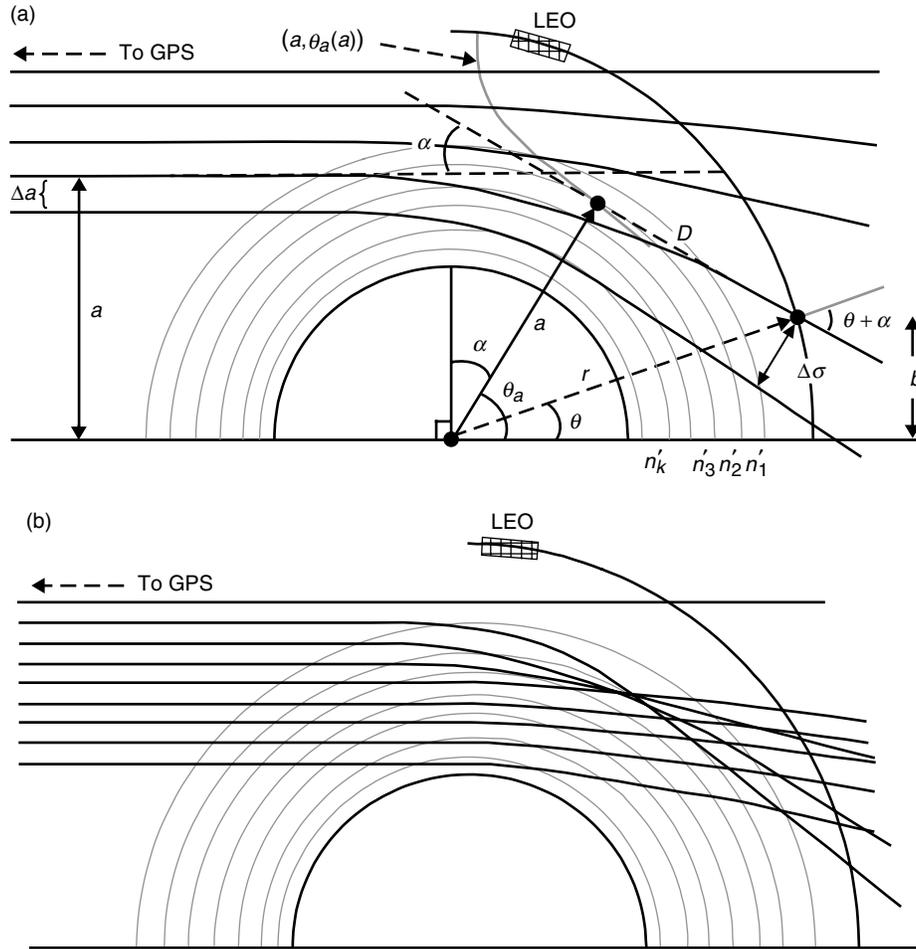


Fig. 1-4. Dispersive bending from the refractive gradient for an occultation observed from a LEO: (a) an ordered set of rays from an increasing refractive gradient with depth; no multipath, and (b) a non-monotonic refractive gradient results in a shadow zone and multipath.

1.1.3 The Global Positioning System

The GPS is operated by the U.S. Air Force (USAF). A system description of the GPS and the signal structure of its broadcast navigation signals are found in [30,31] and in many internet web sites. It suffices here to note a few details.

The GPS constellation is comprised of 24 satellites plus some on-orbit spares, more or less globally distributed up to 55-deg latitude. Their orbits are near-circular with a semi-major axis of about 4.1 Earth radii and with an inclination to the equator of 55 deg. A GPS satellite will be observed from a LEO to rise or set with respect to the Earth's limb on average once per 2 to 3

minutes, or at a rate of several hundred per day. The geographical distribution of the occultation tangency points is more or less global, but the actual distribution depends somewhat on the inclination of the LEO orbit plane and its altitude. For example, a LEO in a polar orbit returns fewer tangency points in the equatorial zone; fewer occulted GPS satellites in polar directions are viewed from that LEO because the GPS orbit inclination is only 55 deg. Figure 1-5 shows the global distribution of tangency points obtained over almost one month in 2001 for analyzed occultations from the CHAMP satellite. The lower density of points in the equatorial zone reveals CHAMP's near-polar orbital inclination. The geographical distribution also reveals vague clustering and striations resulting from commensurabilities between the satellite orbit periods. On roughly one-third of these occultations, a reflected ray from the Earth's surface also was detected [29].

The waveforms of the navigation signals broadcast by a GPS satellite are directional; they bathe the entire Earth with essentially full power. The 3-dB point of their radiant power distribution lobe is about 1400 km above the Earth's surface. Thus, a GPS receiver onboard a LEO, with an orbit radius typically well below this 3-dB threshold altitude, achieves about the same performance as a receiver on the ground.

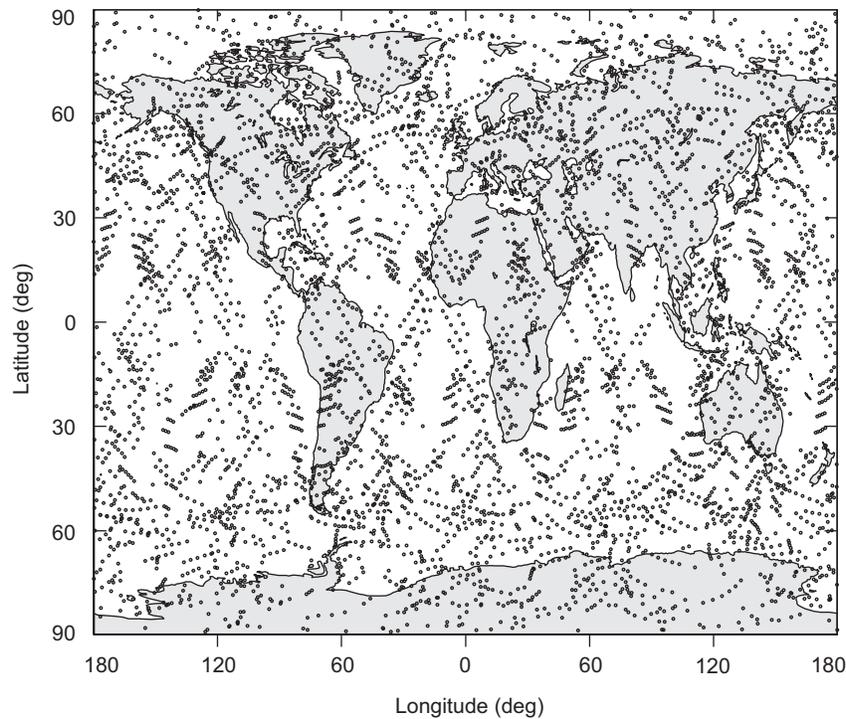


Fig. 1-5. Geographical distribution of nearly 4000 occultations obtained from CHAMP between May 14 and June 10, 2001. Redrawn from [29].

Each GPS satellite continuously broadcasts a set of square-wave codes by coherently modulating the phase of the carrier of the transmitted radio signal. These codes, which are unique to the broadcasting satellite, are pseudorandom and mutually orthogonal. They are used for ranging and for transmitting almanac and timing information. The mutual orthogonality property of the codes enables the receiver by cross-correlation techniques to isolate the received signals broadcast by a given satellite from all others, and to process in parallel the signals from all satellites in view of the receiver. The GPS satellites broadcast ranging codes on a pair of phase coherent L-band carriers, the L1 carrier at a frequency of 1575.4 MHz and L2 at 1227.6 MHz. These include an encrypted precision (P) code with a chip rate of 10.23 MHz on both carriers and the clear access or coarse acquisition (C/A) code at 1.023 MHz on the L1 carrier. The dual carriers are needed primarily to eliminate (or determine) the refraction effect from the ionosphere. For a microwave in the ionosphere, the refractivity is very nearly proportional to the local electron density and inversely proportional to the square of the carrier frequency. Therefore, the range and phase information received separately from the two carriers can be applied in concert to nearly completely decouple the ionospheric refraction effect by using this dispersive property of the ionospheric plasma. Newer versions of the GPS satellites planned for launch this decade will have an additional carrier at 1176.45 MHz (L5) and a C/A-like code also on L2. This will significantly improve receiver tracking operations using clear access ranging codes and increase the accuracy of the ionosphere calibration.

The signal structure of the ranging codes on a GPS signal is designed for near-real-time point positioning. By concurrently tracking four or more GPS satellites in diverse directions, the ranging code measurements can yield within a few seconds absolute point positions with an accuracy of roughly 10 m, and also one's time relative to the GPS clocks. One can obtain near-real-time relative positions (by concurrent tracking with two or more GPS receivers) with sub-meter or even sub-decimeter accuracy.

For occultation applications, however, one needs not the ranging information, but only very accurate measurements of the phase and amplitude of the L1 and L2 carriers, which are by-products of the range code tracking. For occultation applications, we may consider the radio signals arriving at a LEO from a GPS satellite as being a pair of spherical monochromatic waves from a distant point source at frequencies of 1575 MHz and 1228 MHz, respectively, plus Doppler shifts from kinematics and refraction of up to a few tens of kilohertz. Therefore, a high-performance GPS receiver used for space geodetic applications with millimeter-level accuracy requirements is a natural choice for occultation applications because it is designed to measure the phase of each carrier with sub-millimeter accuracy. Chapter 6 discusses certain additional aspects of such a space-rated receiver adapted for operations onboard a LEO.

1.1.4 Timing

A key factor for an accurate phase delay measurement is the epoch of the measurement. The measured phase at the receiver depends on the true phase accumulation between the emitter and the receiver, and the difference between the clock epochs of the emitter and receiver. Knowing the offset in time between the two clocks at a given instant is essential. More precisely, knowing the variability of this offset with time is essential. A constant offset is of no theoretical consequence (but it can be an operational problem) because the refraction information in the phase measurements is contained in their change with time. Each GPS satellite carries onboard up to four very precise cesium and/or rubidium frequency standards for controlling time and time intervals.

In 1999, the U.S. national policy was modified regarding certain operational aspects of the GPS. This change of policy led to the discontinuation of Selective Availability (SA) in May 2000, which had limited the accuracy of near-real-time point positioning to potential adversaries and to civilian users without access to decryption capability. By the late 1990s, SA became increasingly viewed as a cost and productivity issue for many GPS applications, military and civilian. As new technology and alternate means became available to the Department of Defense (DoD) for limiting access to the GPS, it became clear that, even though continuing SA provided a marginal defense benefit, it incurred an economic liability. SA deliberately degrades the near-real-time point-positioning accuracy of the GPS by at least an order of magnitude by causing the clock epoch errors in the GPS satellites to pseudo-randomly wander, nominally by the light-time equivalent of roughly 100 m over several minutes. Although the maximum deviation of the error is bounded and it can be averaged down substantially over 10 minutes or more, the short-term variability of SA poses a significant problem for clock epoch interpolation. SA dithers the onboard GPS master clock oscillator frequency at 10.23 MHz, which is derived from the atomic frequency standards and from which the chip rates of the codes and the frequencies of the L1 (154×10.23) and L2 (120×10.23) carriers are generated. The clock epochs depend on the integral over time of the oscillator frequency. The magnitude of the SA dithering has several possible levels of severity, which were set by certain alert or defense conditions. Over the years prior to 2000, SA had been set at a relatively low level but not at zero. Today SA is set to zero, but it has not been eliminated.

For carrier phase applications requiring high accuracy with SA turned on, multiple ground stations concurrently tracking all the GPS satellites in view at a relatively high sample rate of 1 Hz were needed [32,33]. This high-rate tracking was required to limit interpolation and/or extrapolation errors in rendering measured phases from different GPS satellites to a common epoch. Using these tracking data and applying a “double-differencing” or equivalent scheme among the tracked phase measurements referenced to common transmit epochs,

one can eliminate clock offset errors among the GPS satellites [34]. With these differencing operations, one incurs a penalty in increased effective thermal noise in the phase measurements, which becomes a concern in threshold situations where thermal noise becomes a limiting accuracy factor.

With SA off, the full inherent accuracy of the atomic frequency standards on the GPS satellites (with an effective 10-s frequency stability of a few parts in 10^{12}) can be realized with less costly ground tracking operations [35] and with more leisurely sample rates of 0.1 to 0.03 Hz. The interpolation error in the measured phase change over 10 s from GPS clock instability (with SA off) usually can be kept to well below 1 mm.

In addition to the GPS clock errors, the clock error in the LEO receiver also must be eliminated. However, the LEO usually carries an inferior frequency standard, which requires another strategy for eliminating this error source. This is depicted in Fig. 1-1, which shows the GPS receiver onboard the LEO observing the setting GPS satellite and simultaneously a second GPS satellite clear of any intervening medium, a so-called “clock” or reference satellite. (We assume that the dual-frequency phase measurements eliminate phase effects from the ionosphere.) Differencing the phase measurements from concurrent tracking of these two (now synchronized) GPS satellites at the same reception epoch eliminates the LEO clock error. The cost of this strategy in measurement precision is essentially a $\sqrt{2}$ increase in data noise, usually not a significant limitation except in threshold detection situations.

1.1.5 Ephemerides

For occultation applications, the knowledge error in the relative velocity between satellites must be controlled to a few tenths of a millimeter per second over the roughly 100-s duration of an occultation episode for the neutral atmosphere. This translates into a precision orbit determination (POD) requirement on the LEO of about 30 cm in accuracy, readily achievable these days with a GPS space geodetic receiver onboard. For the GPS constellation, the orbit information for each GPS satellite is included in its navigational signal and is maintained and operated by the USAF, usually with an accuracy of a few meters, or to a few hundred parts per billion. But for many scientific applications of the GPS, particularly space geodetic applications, one needs an accuracy of at most a few parts per billion. Pursuant to this goal, the International GPS Service (IGS) was inaugurated under the auspices of the International Union of Geodesy and Geophysics (IUGG) about a decade ago. The IGS consists of a globally distributed network of over 350 GPS ground tracking stations operated by almost as many different organizations, several communication and data information centers, data analysis centers, and a central bureau for oversight and user interfacing [36,37]. The IGS is comprised of over 200 organizations from over 80 countries collaborating on maintaining

and operating its various system elements. The scientific products from the IGS include very accurate and reliable ephemerides for the GPS satellites accurate to about 1 decimeter, plus a realization of a terrestrial reference frame for ground-based GPS receivers accurate to 1 to 2 parts per billion. These IGS-provided accuracies exceed the requirements for occultation applications.

1.2 Information Content in GPS Occultation Observations

A radio occultation observation profile consists of sequences of amplitude and phase measurements of the L1 and L2 carriers obtained by the LEO receiver over the course of an occultation episode. For the neutral atmosphere, a typical occultation contains a few thousand data points, depending on the sample rate. The measurement error on each point is statistically independent. Clock epoch errors in the transmitter and receiver have been removed using the differential tracking techniques described earlier. Using the POD information about the LEO and the GPS satellites, one can accurately calculate the phase accumulation from satellite kinematics. Subtracting this from the phase measurement leaves as a remainder the excess phase from refraction. Thus, the information content about the refracting medium is contained in the amplitude and excess phase measurement sequences, such as those shown in Figs. 1-2 and 1-3.

1.2.1 Connected Phase

An important aspect of the excess phase measurement sequence is its connectedness. To maintain an accurate phase profile over time, successive phase measurements must (or at least should) be connected, which means that the integer number of cycles accumulated between successive measurement epochs must be exactly accounted for in addition to the fractional cycle determinations made by the receiver at each epoch. In benign signal conditions, the receiver itself can produce connected phase measurements using an internal phase model based on previous phase measurements to extrapolate forward to the next measurement epoch. But in adverse conditions, the raw measurements must be supplemented with the application of more realistic models of the excess phase between epochs. Based on statistical studies from actual occultation profiles, the uncertainty in empirical excess Doppler models for the Earth's atmosphere is about 10 Hz, but with some outliers, which tend to become exacerbated deeper in the moist lower troposphere [34]. Thus, the maximum uncertainty in accumulated cycles between measurement epochs usually is about $10\Delta t$, where Δt is the separation time between successive epochs. For $\Delta t = 0.02$ s, the uncertainty in the integer cycle count usually is 0.2 or less. Fixing cycle breaks can be a significant off-line data-editing task, and sometimes flags are required to denote unfixable cycle breaks.

1.2.2 Sample Rate Versus Vertical Resolution

The canonical sample rate for the phase and amplitude measurements is 50 Hz, which is convenient for the receiver because that is the chip rate of the GPS header code. This is another phase-modulating, very low-rate, square-wave code on the L1 and L2 carriers that carries almanac, timing, health, and other information for the tracked satellite. The individual phase and amplitude samples can be averaged and reported at a coarser rate to reduce thermal noise effects, i.e., to increase the SNR, but at a cost of potentially poorer vertical resolution in the refracting medium. The average vertical velocity of a ray path tangency point is 2 to 3 km/s in the upper atmosphere and an order of magnitude smaller in the lower troposphere, where the refractive gradient is much stronger, and even slower where very large refractive gradients from water vapor occur. Thus, averaging the measurements over a time span broader than where thermal noise is no longer the limiting error source will further limit the vertical resolution with little benefit in measurement accuracy.

Resolution topics, such as the first Fresnel zone, and related topics are discussed further in Chapter 2 and Appendix A. The first Fresnel zone defines the resolution perpendicular to the ray path of a single occultation observation, much like the Airy disk provides a resolution threshold for a circular lens in an optical instrument. As with the Airy disk, the rays passing through this zone are more or less in phase and interfere “constructively”; rays outside this zone interfere destructively. In a refracting medium, the first Fresnel zone is elliptically shaped. The vertical radius is given by

$$\mathcal{F} = \sqrt{\lambda D |\zeta|} \quad (1.2-1)$$

where λ is the wavelength of the GPS carrier, about 20 cm; D is the limb distance of the LEO, roughly 3000 km; and ζ is the defocusing factor. The vertical width of the first Fresnel zone is on average about 1.5 km in the upper atmosphere and roughly 1/2 km in the lower troposphere from dry air alone. But near the Earth’s surface, it can be an order of magnitude smaller when strong refractive gradients from water vapor are present. It suffices here to note that the wave properties of the GPS signal, phase, and amplitude can be used with a sequence of observations to achieve a vertical resolution that is a small fraction of the width of the local first Fresnel zone, perhaps as small as 10 percent. A discussion of data-smoothing strategies involving the trade-off between resolution and measurement accuracy can be found in [34].

1.2.3 Inverting Radio Occultation Data

Although a wave-theoretic approach to recovering the refraction properties of the atmosphere is the principal theme of this monograph, it often is more

convenient and useful to cast propagation processes in terms of geometric optics. The following discussion is set in a ray-theoretic context.

The radio occultation profiles of phase and amplitude are analyzed, broadly speaking, in either a stand-alone mode or in an assimilation mode. In the stand-alone mode, the observation sequences are used to determine a profile for the index of refraction $n(\mathbf{r})$ from some sort of inversion process without preponderant reliance on a priori information. Usually some kind of symmetry or stratification geometry must be invoked for the medium to unambiguously determine $n(\mathbf{r})$. Effectively, the number of degrees of freedom in defining $n(\mathbf{r})$ must be no more than the number of independent observations in an occultation profile in order to achieve an unambiguous determination without using a priori information. For the Earth, local spherical symmetry that accounts for the Earth flattening usually is assumed. Local inhomogeneity in the refractivity along an equipotential surface usually is a small error source for the middle troposphere and higher, but water-vapor-induced horizontal gradients in refractivity near sea level can become a significant error source [38].

Figure 1-4 shows a schematic of one “onion layer” approach. Within each layer, the gradient of the index of refraction is assumed constant. Thus, we have a sequence of unknown parameters, n'_1, n'_2, \dots, n'_M , to be determined from M or more observations. At a specific epoch within the occultation episode, the ray traverses the atmosphere down to a minimum depth, the k th layer, but no deeper. Therefore, the observations for this ray will depend only on the k parameters n'_1, \dots, n'_k . If an unambiguous relationship holds between the observation at a given epoch and the ray—and this is a big if—then it is straightforward to form a linearized system of equations involving a triangular information matrix relating the observation sequence to these refraction parameters. The schematic in Fig. 1-4(a) depicts (with exaggerated bending) an ordered set of rays, i.e., no multipath, with a one-to-one relationship between the ray arriving at the LEO and the value of a , the impact parameter of the ray. With each successive measurement epoch in this figure, the ray at its tangency point passes through a new layer that is lower in altitude and that was not “sounded” before. In Fig. 1-4(b), we have multipath for some periods when more than one ray arrives at the LEO at the same time; the impact parameter of each ray can be moving either up or down with time, depending on the type of ray. Therefore, an analysis scheme must first be applied to the observations to discriminate among concurrently arriving rays.

In the linearized version relating observations to refraction parameters, an observation at a given epoch is the actual observation minus the predicted observation based on an initial estimate for parameter set $\hat{n}'_1, \hat{n}'_2, \dots, \hat{n}'_M$, and based on a provisional ray path passing through the medium that follows Snell’s law. Inversion of this information matrix, or inverting a weighted least-squares matrix version if over-determined, yields a determination of the

corrections $\Delta\hat{n}'_1, \Delta\hat{n}'_2, \dots, \Delta\hat{n}'_M$. One then updates $\hat{n}'_1, \hat{n}'_2, \dots, \hat{n}'_M$ by adding the corrections to these predictions and generates a new ray path, forms new observation differences, and iterates until convergence is achieved. Assuming that $n_0 = 1, n'_0 = 0$, one can integrate this converged sequence $\hat{n}'_1, \hat{n}'_2, \dots, \hat{n}'_M$ downward to obtain a recovered profile $\hat{n}(r)$ for the index of refraction. As a practical matter, at very high altitudes, above 50 to 60 km, the noise in the occultation observable from measurement error or ionospheric calibration error becomes larger than the signal from the atmosphere. In these high altitudes, to initiate the downward integration process, one can statistically combine in a maximum-likelihood sense the noisy occultation observable with an a priori value furnished by a model atmosphere, along with the appropriate covariance matrices [39]. Because refractivity varies approximately exponentially with altitude at these heights, an error in this matching and handing-over process (from strong reliance on a priori information at higher altitudes to strong reliance on the data at lower altitudes) is damped out exponentially with decreasing altitude and usually is nearly gone within 2 scale heights.

When spherical symmetry applies and only a single ray arrives at the LEO over time, or when multiple rays can be discriminated according to their excess Doppler signature, then both the bending-angle and impact parameter values α and a are obtained from the derived excess Doppler information for each ray. Excess Doppler for a given ray is the time derivative of its excess phase. In this case, the Abel integral transform [see Eq. (1.2-5)] can be applied [40,41]. This integral transform directly recovers without iteration the profile $\hat{n}(r)$ from the observed bending-angle and impact parameter sequences for each evolving ray arriving at the LEO. At very high altitudes where the atmospheric signal is weak, the same kind of statistical hand-over using predicted excess Doppler from a model is still used to initiate the integral downward.

The accuracy of the recovery of the index of refraction can be determined theoretically based on an assessment of the error sources within the occultation system. It also can be inferred from statistical comparisons with NWP model values from, for example, the European Centre for Medium-Range Weather Forecasts (ECMWF) [19,42]. Fractional accuracy for the refractivity recovery of 0.1 to 1.0 percent is achievable from about 5 km up to 30 to 40 km. Here the assumption of local spherical symmetry usually is valid and the signal from the atmosphere is relatively strong compared to the measurement noise. But in the lower troposphere, horizontal variations in refractivity can become a significant error source when laminar symmetry is assumed, and tracking becomes more difficult and the measurements are noisier. These factors can limit further improvement with depth of the accuracy of the refractivity recovery.

If the chemical composition of the medium is known, one knows precisely the relationship between the density of the medium and its refractivity. In this case, one obtains the density profile $\rho(r)$ of the refracting medium from the

recovered profile for $n(r) - 1$. Invoking hydrostatic equilibrium for the medium in a gravity field $g(r)$, one has the hydrostatic equation for the pressure $dp/dr = -g\rho$. Integrating this yields the pressure profile $p(r)$. Invoking thermodynamic equilibrium allows one to apply the ideal gas law to the recovered density and pressure profiles to obtain the temperature profile.

For the Earth, the gaseous constituents of the neutral atmosphere affecting the refractivity at microwave frequencies are dry air and water vapor. For an air/water system in thermodynamic equilibrium, the Clausius–Clapeyron equation gives the partial pressure of water vapor, which varies exponentially with $-1/T$ [43]. Reducing the temperature by 10 percent reduces the saturated water vapor content by an order of magnitude. Below a threshold temperature of about 250 K, it is too cold for any residual water vapor to significantly contribute to the refractivity at L-band frequencies. The altitude corresponding to a temperature of 250 K typically is in the range 6 to 8 km. For altitudes above this threshold and below roughly 40 km, the potential accuracy of the temperature recovery is less than 1 K, probably less than 0.5 K for the 10- to 25-km range [21]. For higher atmospheric temperatures at lower altitudes, water vapor becomes an additional component in the refractivity. We speak of the “dry delay,” which is the phase delay caused by the refractivity from air, and the “wet delay,” which is the delay solely from water vapor. The latter arises from the permanent electric dipole moment of the water molecule, which is a significant contributor to the refractivity at microwave frequencies. The problems with the wet delay generally are that the density of water vapor is highly variable in space and time, its relative abundance or specific humidity is uncertain, and its refractivity per mole at L-band frequencies is much larger, about 16 to 18 times larger, than the refractivity per mole of dry air. Especially in the lower troposphere, this variability in specific humidity results in large refractive gradients that can double or even triple the refractive bending from dry air alone. In the stand-alone mode, therefore, the radio occultation technique requires ancillary information for those regions where the temperature is above 250 K to convert the recovered refractivity profile into unambiguous density, pressure, temperature, and specific humidity profiles. In temperate and tropical regions where the water vapor content is a major uncertainty, the temperature profile usually is relatively better known from NWP models, and it is often used to aid the occultation recovery of water vapor [19,21].

1.2.4 Assimilating Radio Occultation Data

In the assimilation mode, the occultation observable profiles are assimilated through statistical inference processes into a global atmospheric model where the basic meteorological quantities of the refracting medium, density, pressure, temperature, specific humidity, cloud, aerosol and precipitation distributions,

wind fields, and so on are defined. Therefore, the index of refraction $n(\mathbf{r})$ is already constrained to varying degrees in space and time by a usually much larger and more global data set in the atmospheric model. Covariance matrices characterizing the statistical aspects of the global data set and the atmospheric models controlling thermodynamic and dynamic processes are also part of this global system. In contrast to the stand-alone mode, no symmetry or stratification assumptions about the refracting medium are theoretically required here. Moreover, the granularity of the model for the refracting medium for numerical computation or, equivalently, the four-dimensional cell size (three in space and one in time) is already established and controlled by the continuity equation. Usually the number of degrees of freedom, even when constrained by the physical equations in the model, is vastly more than found in a typical stand-alone occultation model.

The four-dimensional variational analysis (4DVAR) technique in NWP programs is a prime example of the assimilation mode for using the occultation data [44–47]. Here the occultation profiles are assimilated competitively in a statistical treatment of the data. The radio signal from the position of the occulted GPS satellite is propagated through this global model atmosphere with its free parameters provisionally fixed. At each observation epoch in the occultation sequence of M observations, the difference between an actual observable of the GPS signal (e.g., excess phase, excess Doppler or bending angle, signal amplitude) and its propagated value based on the provisional parameter set is computed. This difference becomes a component of an additional M -dimensional observation vector in the global data set. One adjusts the free parameters of the global model to minimize a cost function characterizing the goodness of fit of all the data weighted by their covariance matrix inverses, including the occultation data. The efficacy of the occultation information in this environment to modify the values of the free parameters obviously depends on the information content in the global system and on the assigned covariance matrices.

1.2.5 Rays and Stationarity

The Poynting vector, which defines the direction and magnitude of the radiant power in an electromagnetic wave at a given point in terms of the cross product of the electric and magnetic field vectors, is perpendicular to the cophasal surface of the wave at that point. The limiting form that the Poynting vector takes as the wavelength of the wave is shrunk to zero defines the tangent vector of the ray in geometric optics. The second-order differential equation that results from this limiting process gives the curvature of the ray at any point in terms of the gradient of the refractivity there. This limiting process also gives the amplitude of the ray based on the Poynting vector, its reflection and transmission properties that would apply to an electromagnetic wave of very

short wavelength, and so on. What is missing in ray theory, by definition, is diffraction. Also, ray theory (as geometric optics is usually defined) cannot deal adequately with caustic rays or with trapped rays in a super-refracting medium where the curvature of a ray, if it could exist in this environment, would temporarily be greater than the curvature of the local equipotential surface. A caustic ray is one that first comes in contact with the envelope to a family of rays, for example, a spectral component of a rainbow. Born and Wolf provide a comprehensive discussion of the foundations of geometric optics following this limiting approach in [48], which cites many references giving the historical development of geometric optics.

The same differential equation for the ray (the Euler equation) comes from the calculus of variations upon applying Fermat's principle (see Appendix A). The phase accumulation φ along a path through a three-dimensional refracting medium from point A to point B may be written in parametric form as

$$\varphi = k \int_A^B n(\mathbf{r}) ds = k \int_A^B n(\mathbf{r}) (\dot{x}^2 + \dot{y}^2 + \dot{z}^2)^{1/2} dt \quad (1.2-2)$$

where $k = 2\pi/\lambda$ is the wave number of the wave, $n(\mathbf{r})$ is the index of refraction at the position $\mathbf{r} = \hat{x}x + \hat{y}y + \hat{z}z$, $ds = (\dot{x}^2 + \dot{y}^2 + \dot{z}^2)^{1/2} dt$ is the incremental arc length along the path, t is an arbitrary parameter denoting position along the path, $(\dot{}) = d()/dt$, and the triad $(\dot{x}, \dot{y}, \dot{z})$ defines the slope of the path. The path is described by the functions $x(t)$, $y(t)$, and $z(t)$, and we now stipulate that they describe a ray path. Fermat's principle requires that this path integral along a ray be stationary. In other words, the phase accumulation from the same initial point to the same end point along any other physical path neighboring the ray path would differ from the phase accumulation along the ray path in a second-order manner. Let the neighboring path be described by the functions $x(t) + \varepsilon \xi_x(t)$, $y(t) + \varepsilon \xi_y(t)$, and $z(t) + \varepsilon \xi_z(t)$, where ε is a small parameter. The functions $\xi_x(t)$, $\xi_y(t)$, and $\xi_z(t)$ are completely arbitrary other than that they must be physical and they must satisfy the boundary conditions, i.e., $\xi_x(t_A) = \xi_x(t_B) = 0$, etc. Then $d\varphi/d\varepsilon$ must be zero when evaluated along the ray, on which $\varepsilon = 0$. This stationary-phase condition requires that the ray path at every point satisfy the Euler differential equation or its integral equivalent where a discontinuity in $n(\mathbf{r})$ occurs. Usually the stationary value of φ along a ray is a local minimum, i.e., $d^2\varphi/d\varepsilon^2 > 0$, but not always. Anomalous rays provide a local maximum, and caustics involve a breakdown of the assumption of a non-zero second-order variation, which is embedded in the foundations of geometric optics.

The calculus of variations is a powerful technique for developing basic ray theory and many of its attributes. It may be used to develop the eikonal equation for obtaining phase delay along a ray, the transversality condition for

relating the change in end-point phase delay to small changes in end-point position (including relating excess Doppler to refractive bending angle), the Hamilton–Jacobi theory for obtaining the end-point phase delay expressed as a field variable in terms of the end-point position variables for a family of rays, and finally, for obtaining the canonical variable representation of a ray. We discuss these aspects in Chapter 2 and in Appendices A and B.

The stationary-phase property of rays affects one’s strategy for processing occultation data using ray theory, depending on one’s objective. In a sense, stationarity is both a blessing and a curse. Regarding the latter, consider the process of finding the ray path passing through the end points A and B for a given trial index of refraction profile $\hat{n}(\mathbf{r})$. One can forward propagate from A or backward propagate from B using Snell’s law and a ray-tracing method. By adjusting the slope parameters of the ray at these end points, one theoretically can converge on a complete ray path that matches the boundary conditions, if such a ray exists. The problem is that the calculated phase delay obtained by following this provisional path also must match the observed value φ for each observation epoch. This in turn requires adjustments in $\hat{n}(\mathbf{r})$ to reach the observed values for φ ; for the stand-alone mode, this would require an inversion process such as that described earlier. But along the ray, the phase delay is stationary with respect to small variations in the slope parameters or to variations in any other parameter, for example, the impact parameter, that characterizes an alternative nearby path that satisfies the boundary conditions. There is a singularity here. In effect, the “density” of admissible paths around the actual ray with a phase delay within the limits $(\varphi, \varphi + \delta\varphi)$, all satisfying the boundary conditions, becomes infinite along the actual ray where $\delta\varphi = 0$. This singularity in path density forms the basis of Fermat’s principle. The idea is that all physical paths in the vicinity of the actual ray have the same phase delay to first order. Therefore, the phasor, $\exp(i\varphi)$, for each path neighboring the ray is the same, to first order, and when the phasors from all of these possible nearby paths are summed (Huygen’s principle) and averaged to obtain the total field at point B , they sum constructively, reinforcing the amplitude. For paths with larger deviations from the ray, their first-order phase variation with respect to path deviation is not zero. Therefore, their phase delays for different paths in their neighborhood are distributed more or less randomly over 2π radians, and their phasors destructively combine. No average field results from those paths with a non-zero first-order variation, only fluctuations. This concept of constructive reinforcement for nearby paths about a ray and destructive reinforcement for paths away from the ray forms the basis for the idea of a ray, the basis for Fermat’s principle, and the basis for the first Fresnel zone. It also is the central concept in the Feynman sum-over-histories technique used in quantum electrodynamics to calculate the quantum wave function for the probability amplitude of a quantum event [49].

This singularity in path density results in numerical instability when the measured phase sequence per se is used as a criterion to reconstruct the most probable ray path for each observational epoch. On the actual ray, $d\varphi/d\varepsilon = 0$, where ε is a parameter describing the departure of a trial path from the actual ray path. By tinkering with ε , one tries to converge to the actual ray path. One would use an iterative differential correction scheme of the linear form $\delta\varepsilon = -(d\varphi/d\varepsilon)^{-1} \delta\hat{\varphi}$, where $\delta\hat{\varphi}$ is the observed value of φ minus the computed value $\hat{\varphi}$ based on the provisional profile $\hat{n}(\mathbf{r})$ and the provisional path through this medium between the end points. Here one attempts to determine iteratively the maximum-likelihood estimates of $\delta\varepsilon$, $\hat{n}(\mathbf{r})$, and the path. This scheme will run into computational trouble after successive iterations as the iterated path approaches the actual ray path because $d\varphi/d\varepsilon \rightarrow 0$ as $\varepsilon \rightarrow 0$.

1.2.6 Excess Doppler

Stationarity also brings a blessing, the transversality condition. If one slightly changes the end-point values of the ray, a new ray will pass through the new end points. The phase delay along the new ray between the new end points minus the phase delay along the old ray between the old points depends only on local conditions at the end points, not on the path in between. This is rather remarkable considering that the phase delay itself is computed from a path integral along a new path, which explicitly accounts for the variability in $\hat{n}(\mathbf{r})$ along the entire ray path. This local rather than global dependency on change is a unique property of a stationary-phase path.

Using the transversality condition, it is shown in Appendix A that the rate of change in excess phase resulting from the satellite velocities is given by

$$\lambda f_D = n_G \mathbf{T}_G \cdot \mathbf{V}_G - n_L \mathbf{T}_L \cdot \mathbf{V}_L - \dot{r}_{GL} \quad (1.2-3)$$

where $f_D = (d\varphi/dt)/2\pi$ is the rate of change of the excess phase in cycles, or excess Doppler shift of the ray, and λ is the wavelength of the harmonic wave. Here G denotes the position of the end point of the ray at the transmitting GPS satellite, and L denotes the end point position at the LEO. Also, \mathbf{V} is the velocity vector of the satellite at an end point of the ray, \mathbf{T} is the unit tangent vector of the ray at an end point, and \dot{r}_{GL} is the radial velocity between the end points G and L (see Fig. 1-6). POD information provides a determination of \mathbf{r}_G , \mathbf{V}_G , \mathbf{r}_L , and \mathbf{V}_L , and therefore \dot{r}_{GL} , at each observational epoch. It follows that the excess Doppler shift for a given ray, a quantity directly obtained from the observed excess phase profile for that ray, imposes a constraint on the unknown unit ray path tangent vectors \mathbf{T}_G and \mathbf{T}_L . If we assume coplanar propagation, usually a good assumption, then \mathbf{T} is defined by a single deflection angle

$$\log n(a) = \frac{1}{\pi} \int_a^{\infty} \frac{\alpha(\rho)}{\sqrt{\rho^2 - a^2}} d\rho \quad (1.2-5)$$

to recover the refractivity profile in terms of the impact parameter a . The Abel transform is derivable from the basic integral equation for the bending angle in a spherical symmetric medium (see Appendix A):

$$\alpha(a) = -2a \int_a^{\infty} \frac{d \log n}{d\rho} \frac{d\rho}{\sqrt{\rho^2 - a^2}} \quad (1.2-6)$$

By substituting this form for $\alpha(a)$ into the integral operation on the right-hand side (RHS) of Eq. (1.2-5), one indeed recovers $\log n(a)$. If significant multipath occurs within a section of the observation series, then it should be noted that a_k , $k = 1, 2, \dots, M$ will not be time-ordered during that section. But, as long as one can successfully recover a sequence (α_k, a_k) that is ordered, then the Abel transform can be applied.

If spherical symmetry does not apply, then some ray tracing technique combined with the appropriate stratification model for the refractivity can be used to recover the refractivity profile, or one can use the assimilation mode.

Note from Eq. (1.2-6) that the existence of a ray between given end points with an impact parameter value of a requires that $\rho \geq a$ at all points along the ray. In a super-refracting layer, $d\rho / dr = (n + rn') < 0$, and this condition $\rho \geq a$ can be violated for a certain range of tangency points r_* . Within this range, no rays exist. Chapter 6 briefly discusses super-refractivity in a spherical shell.

1.3 Scientific Applications of GPS Occultation Observations

We briefly review the major uses of occultation observations, which fall broadly into the categories of meteorology, weather prediction, and global climate change. Accuracy, resolution, and the global distribution of the GPS occultations are key factors. Regarding accuracy, we have already noted the accuracy of the refractivity recovery from the stand-alone mode, 0.1 to 1 percent fractional accuracy above the lower troposphere, where the assumption of local spherical symmetry is likely to be satisfactory, and below 40 km. This translates into the same fractional accuracy range for the density determination for temperatures below 250 K. It also implies dekameter accuracy for a geopotential height determination of a point on a constant pressure surface, for example, the difference in the potential for the 300-mbar surface minus the potential of the ocean geoid, with the difference divided by the gravity coefficient [50]. It also translates for the same altitude range into sub-kelvin accuracy in temperature recovery.

Regarding these accuracy figures, the caveats about the applicability of spherical symmetry and the temperature being colder than 250 K have more to do with the model used and the analysis approach, and less to do with the inherent accuracy of the measurements. Good, clean phase and amplitude measurements down to the Earth's surface could yield 0.1-K accuracy for temperature recovery if we knew the water vapor content and the stratification. So, these caveats are somewhat misleading and tend to understate the inherent strength of the occultation data if our strategy is instead to assimilate them into a 4DVAR process.

Radio occultations bring new capabilities to the study of processes in meteorological models. With occultation data, one can study the fine structure of various refractive boundaries, such as the tropopause, a marine boundary, and a sporadic E-layer in the ionosphere. Temperature and pressure recovery from the middle troposphere through the middle stratosphere can lead to better understanding of energy transport and exchange processes across the tropopause and of how the atmosphere radiates dynamical energy and heat through acoustic gravity waves. Gravity waves are quite evident in occultation-derived temperature profiles in the stratosphere [51–54]. Phase and amplitude measurements made down to or near the Earth's surface provide powerful constraints on water vapor, which is such an important greenhouse gas and so relevant to energy transfer and balance processes in the atmosphere and to aerosol growth and cloud formation. Occultation-derived specific humidity distributions will provide sharper information about cyclogenesis processes, baroclinic wave and wind field development, and so on. More detail about many promising new capabilities is provided in [22,55].

1.3.1 Weather

The utility of occultation data when assimilated into 4DVAR programs depends greatly on the degree of complementarity of the information content in the occultation data set compared to the 4DVAR data set. The utility will be highest when other data are relatively sparse or when certain physical parameters of the 4DVAR model are less well constrained. This includes both data-sparse geographical locations and certain locations in altitude, even in data-rich geographical areas. Occultation profiles are in their best form for tangency points in the upper troposphere up to the lower stratosphere. Here the technique can provide unprecedented accuracy in density, pressure, and temperature, and it offers very sharp (perhaps dekameter-level) vertical resolution.

The utility of the occultation data also depends on their latency. The ephemeral aspects of the weather force very short latency requirements. That is why planned operational LEO occultation constellations, such as COSMIC,

have a 2- to 3-hour latency goal: occultation data to be completely assimilated within 3 hours of the actual occultation event.

Regarding geographical complementarity, consider again the planned COSMIC constellation, which will return on average about 10 occultations within a 3-hour period over a geographical area the size of the continental U.S. Obviously, 10 occultations from southern ocean or polar locations where weather data are sparse will have a bigger impact in weather models for those areas than will 10 occultations obtained over the continental U.S. and southern Canada in weather models for that area. In North America, the myriad weather sensors and observation programs is constraining.

Nonetheless, the water vapor information provided by the radio occultation data has a global utility. These data are powerful because they highly constrain the water vapor uncertainty. Accurate knowledge of specific humidity distributions, particularly in the lower troposphere, is key to more accurate weather prediction, for example, the ongoing development of wind fields, cloud formation, and precipitation in cyclogenesis processes. It is the water vapor information that makes reaching the Earth's surface with a high percentage of the radio occultations such an important goal for operational programs [56].

Pursuing this complementarity concept further, let us see how the water vapor information content in an occultation might fold into a 4DVAR process. The refractivity at L-band frequencies in the neutral atmosphere is given by [57,58]

$$N = n - 1 = \left(77.6 \frac{p}{T} + 3.73 \times 10^5 \frac{p_w}{T^2} \right) \times 10^{-6} + \text{negligible terms} \quad (1.3-1a)$$

where p is the pressure of the moist air in mbars, p_w is the partial pressure of the water vapor, and T is temperature in kelvins. Here and throughout this monograph, we use an unconventional definition for refractivity, $N = n - 1$, instead of the standard $N = (n - 1) \times 10^6$. Let us rewrite the refractivity equation in terms of the molecular number densities, n_a and n_w , in moles/m³, where n_a and n_w are the dry air and water vapor number densities, respectively. We use the ideal gas law $p_x = n_x RT$, where R is the universal gas constant. Then Eq. (1.3-1a) becomes

$$\left. \begin{aligned} N &= c(n_a + b(T)n_w) + \text{negligible terms} \\ c &= 6.45 \times 10^{-6} \text{ m}^3 \cdot \text{mole}^{-1}, \quad b(T) = 16.0 \left(\frac{300}{T} \right) + 1 \end{aligned} \right\} \quad (1.3-1b)$$

The relatively high sensitivity of the refractivity to the density of water vapor should be noted; the value of $b(T)$ typically is 16 to 18. Because the mean

molecular mass of water is 18 g/mole and dry air is 29 g/mole, the sensitivity of the specific refractivity is even greater.

Suppose now that we have inverted the occultation data to obtain an estimate $\hat{N}(r)$ of the refractivity. We could make a similar case for complementarity using a basic occultation observable, such as excess phase, Doppler, or amplitude. But it is easier to show using the recovered refractivity. Assume that at some epoch during the occultation episode the tangency point of the ray is located at a particular altitude where water vapor is a significant factor. Suppose that the error in the recovered refractivity there, $\delta\hat{N}$, is a random variable that is Gaussian distributed with a mean value of zero and a standard deviation of $\sigma_{\hat{N}}$. Then, from Eq. (1.3-1b), the probability density distribution $P[\delta\hat{N}]$ for $\delta\hat{N}$ written in terms of corresponding Gaussian errors in \hat{n}_a and \hat{n}_w is given by

$$P[\delta\hat{N}] = \frac{c^2 b}{2\pi\sigma_{\hat{N}}^2} \exp\left[-\frac{c^2 b^2}{2\sigma_{\hat{N}}^2} \left(\frac{\delta\hat{n}_a^2}{b^2} + \frac{\delta\hat{n}_w^2}{1^2}\right)\right] \quad (1.3-2)$$

Thus, a contour of constant probability density in $\delta\hat{N}$ is an ellipse with the semi-major axis along (or nearly so) the n_a axis and the semi-minor axis along the n_w axis. The coefficient $b(T)$ gives the ratio of these axes. The left-hand error ellipse, System A, in Figs. 1-7(a) and 1-7(b) gives qualitatively the contour for a constant probability density for $\delta\hat{N}$ from an occultation expressed in terms of the corresponding errors, $\delta\hat{n}_a$ and $\delta\hat{n}_w$, in the statistical estimates of n_a and n_w . For example, the ellipse might be the contour corresponding to the $1-\sigma$ values for $\delta\hat{n}_a$ and $\delta\hat{n}_w$; in this case, the probability of finding $\delta\hat{N}$ within the area inside of this ellipse is 40 percent.

Several comments can be made about Figs. 1-7(a) and 1-7(b). At this altitude of the tangency point and at this observation epoch, the error ellipse from the occultation information, System A, highly constrains the range of probable values for \hat{n}_w compared to the range of probable values for \hat{n}_a . System B in Fig. 1-7(a) shows another error ellipse for $\delta\hat{n}_a$ and $\delta\hat{n}_w$ from another information system. It is derived from a presumably higher-dimensional covariance matrix describing the uncertainties in the recovered values for the global parameter set in System B mapped to the same epoch and altitude and projected into our two-dimensional $(\delta\hat{n}_a, \delta\hat{n}_w)$ space. The information in System B is statistically independent from the information in System A. In Fig. 1-7(a), the information in System B is more or less “parallel” with the information in System A, nearly the same strength and the same weakness. In Fig. 1-7(b), System B is complementary to System A, more or less orthogonal, as characterized by the crossed error ellipses; the strength in

one system is a weakness in the other. The error ellipses for the two systems, A and B, may or may not be of comparable scale or eccentricity, but one can make an argument that they are often comparable. With the 2×2 covariance matrices for System A and System B, one can statistically combine the information in these two systems to obtain minimum variance estimates for n_a and n_w . The combined covariance matrix yields the System C minimum variance error ellipse. In Fig. 1-7(a), the combined System C, when the information in the two systems is comparable in accuracy, shows a roughly $1/\sqrt{2}$ improvement in probable errors all around, but it inherits the same strengths and weaknesses from the two systems. In the orthogonal case in Fig. 1-7(b), System C inherits the strengths from each system and none of the weaknesses; its error ellipse is dramatically improved. Water vapor content in NWP models is very difficult to know accurately in place and time, and it is generally less well-known in these models than are the total pressure and temperature. The near-orthogonal error ellipse in Fig. 1-7(b) for System B probably is more typical for NWP models.

In summary, in weather applications, complementarity from radio occultations, if they are timely, comes mainly from locations with relatively sparse meteorological data or from regions with more loosely constrained parameters. The other major contribution to NWP from the radio occultations is their highly constrained water vapor information, as just discussed. The long-perceived *bête noir* of the radio occultation technique for the Earth—water vapor—is in fact its strength.

1.3.2 Climate

In global climate change studies, detecting weak signatures in a noisy Earth system environment over decadal periods and longer is a major challenge. Here,

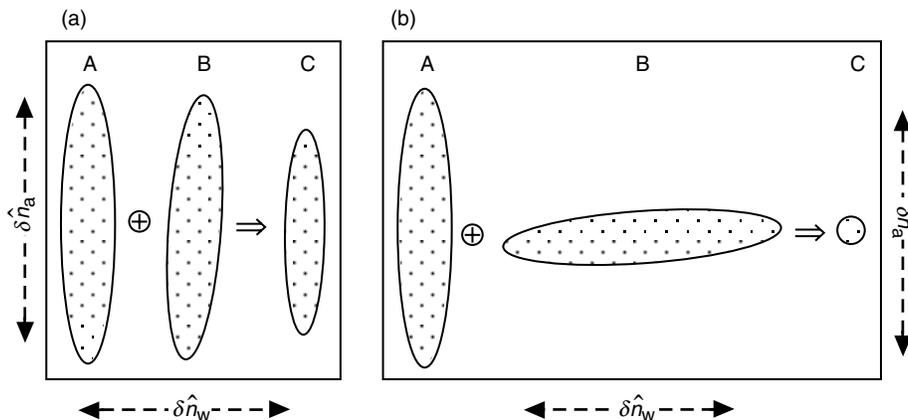


Fig. 1-7. Error ellipses for independent information systems A and B, and their minimum variance combination C: (a) parallel information and (b) complementary information.

the stability, high accuracy, and fine vertical resolution of the GPS occultations are key assets for use in climate studies. Except for the ionosphere, the GPS observations are essentially calibration-free, with no instrumental biases or calibrations that vary slowly. Even with the ionosphere, it would be a decadal variation in the usually small refraction effects from third- and higher-order frequency terms that could possibly “leak into” long-term observations in the stratosphere. This seems an unlikely scenario because modeling the ionosphere to aid the dual-frequency corrections significantly reduces errors from higher-order terms on the recovered refractivity in the stratosphere [59,60].

In monitoring possible secular trends, one can spatially average the recovered temperatures at a given altitude, or geopotential heights at a given pressure, to reduce by perhaps an order of magnitude the effect of random errors and variations. Thus, all the points lying within a specified geographical area, say a region 500×500 km in size, could be lumped and then averaged over time, weeks or months, enhancing the probability of detecting signatures with spectral power at very long periods.

Secular variations in the geopotential height of the 300-mbar surface, for example, can be a sensitive indicator of secular changes in the average temperature, \bar{T} , below. The sensitivity coefficient, $\partial h / \partial \bar{T}$, is about 30 m/K. A geopotential height change of 30 m at 300 mbar almost surely would be detectable over a long series of occultation observations, especially if it were regionally averaged, but ascertaining that it was a temperature signal rather than some other possible physical signal would remain a challenge.

A predicted secular effect associated with the global warming scenario is a warming troposphere but a cooling stratosphere. Therefore, monitoring averaged changes in recovered temperature profiles across the tropopause could be a very sensitive indicator of this warming trend. Other signatures associated with global warming to which averaged radio occultations might be applied are increased water vapor content and variability in the tropics, changes in the residual water vapor in the upper troposphere, and warming in the lower troposphere in polar regions. More examples and details on climate applications are given in [22,26].

1.4 Problems from Multipath and Some Remedies

The main problem from multipath, other than stressing the signal-tracking operations by the receiver, is in the ambiguity that it introduces regarding the appropriate excess Doppler value to use for each arriving ray. We have already noted in Fig. 1-3 the interference in amplitude from multipath rays for a particular occultation from GPS/MET. Figures 1-8(a) through 1-8(d) from CHAMP show similar interference in amplitude for a selected group of occultation profiles. The de-trended phase measurements for these occultations show similar interference. One should note from Fig. 1-8 the generally

improved SNR on CHAMP compared with GPS/MET. The GPS antenna on CHAMP has about 4 dB more gain, which on average yields about a factor of two increase in free space SNR_v . These profiles reveal striking and highly transient features in signal amplitude as the ray path tangency point descends through the atmosphere. Occultations passing through narrowly defined layers of the ionosphere also reveal similar transients in SNR and phase as a result of sharp gradients in electron density. Examples of passing through a sporadic E-layer in the lower ionosphere are seen in Fig. 1-8(b) and will be found later in Chapter 2 in Figs. 2-15 and 2-16 from GPS/MET.

A characteristic of many of these transients is their brevity, which, when the downward velocity of the ray is considered, translates into an altitude range that is usually narrow compared to the vertical width of the first Fresnel zone. Moreover, deep troughs in the SNR of the received signal combined with contiguous and abrupt flaring (and concomitant transients in phase) are often observed. These are almost sure signs of multipath and/or shadow zones, and even of ducting on occasion through a super-refracting layer [Fig. 1-8(a)]. In some instances, caustic rays are evident and also diffraction fringes when the physical refractivity feature is sharp enough and the neighboring environment is smooth enough.

Multipath is essentially the rule in occultations, not an exception. There are benign periods, of course, but even those usually show some small interference or scintillation that exceeds the thermal noise; it is a matter of degree. For example, the SNR in Fig. 1-8(c) at $t=0$ predicts a 1 percent thermal noise level ($\sqrt{50} / \text{SNR}_v$), but the actual noise is 2 to 3 times larger. One can average the samples down from a 50-Hz rate to a smaller rate, for example 1 Hz. Because the typical frequency difference, Δf , between interfering rays is several hertz, averaging dramatically improves the multipath “noise” compared to thermal noise. Over an averaging time $\Delta t \gg (\Delta f)^{-1}$, the amplitude of the fringes from multipath averages down as $1 / \Delta t$, whereas thermal noise averages down as $(1 / \Delta t)^{1/2}$. On the other hand, averaging can degrade the vertical resolution potential. For a LEO, the vertical motion of the ray path tangency point spans the vertical diameter of the local first Fresnel zone in a time interval of $\Delta T \approx 0.6\zeta^{-1/2}$ s, where ζ is the defocusing factor, unity at the top of the atmosphere and usually 0.1 to 0.01 at sea level. Therefore, averaging observations over a time interval longer than ΔT exceeds the Fresnel limit, which is the resolution from only a single observation. Sub-Fresnel vertical resolution, obtained from an ensemble of observations by means analogous to the resolution obtained from synthetic aperture radar systems, and a holy grail for boundary studies, can be considerably hampered by time-averaging the observations. Ultimately good SNR is the key.

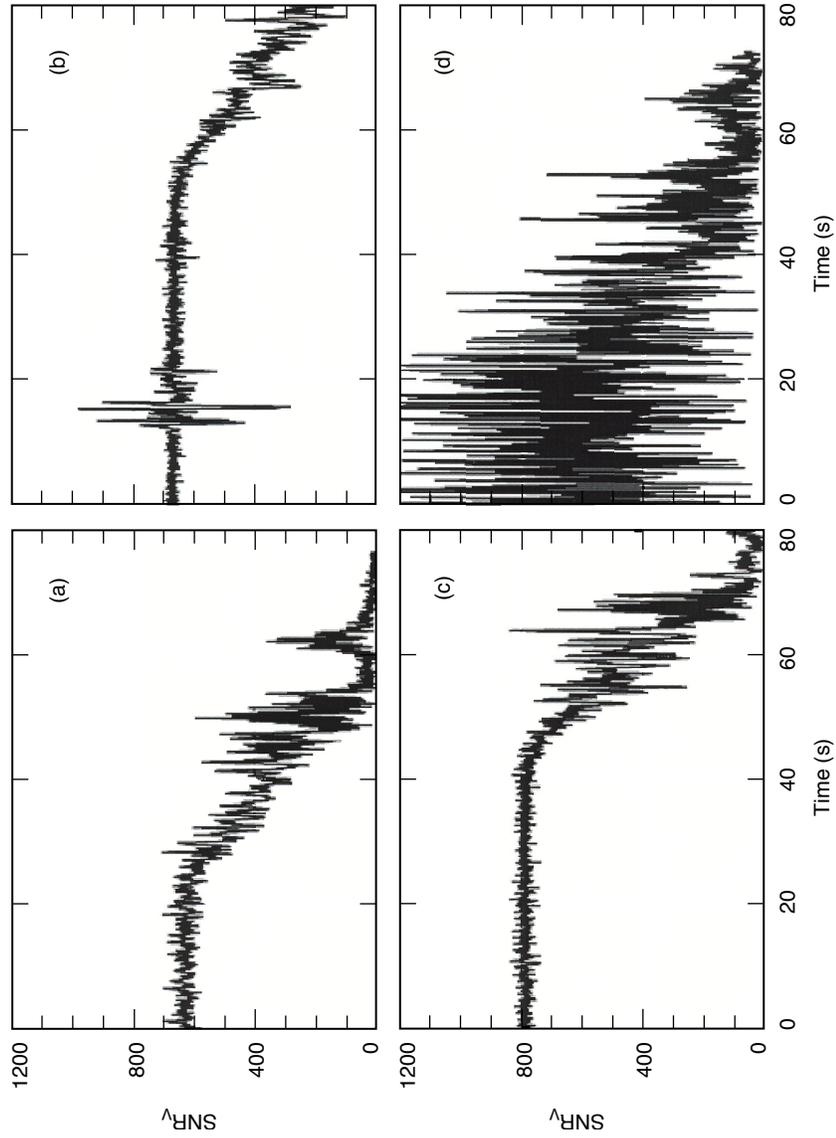


Fig. 1-8. Voltage SNR profiles from four CHAMP occultations in October 2001: (a) probable ducting in the lower troposphere (at ~54 to 62 s), (b) sporadic E-layer (at ~15 s), (c) strong, clean signal above the tropopause (at ~55 s), and (d) troublesome.

Figure 1-9 shows a phasor diagram in the complex plane containing a snapshot at a given instant of the phasors for three interfering rays, the “main” ray m , and the additional rays a and b . This is one of the simplest multipath scenarios, which usually comes in odd-numbered packets, 3, 5, 7, ..., except at caustic contact points, or when a reflected ray is present. The vector sum

$$\hat{E}(t) = E_m(t) + E_a(t) + E_b(t) \quad (1.4-1)$$

gives the observed amplitude and excess phase. The excess phase and amplitude of the individual components are not directly observable. The amplitudes of these rays, E_m , E_a , and E_b , usually differ and they also vary with time, but usually at a much slower rate than $\hat{E}(t)$ can vary because of interference. Their amplitudes are determined by their respective defocusing factors, which usually vary relatively slowly except near a caustic contact point. The individual excess phase of each contributing ray, φ_m , φ_a , or φ_b , depends on its individual ray path through the medium, which is changing with time as the ray path tangency point migrates downward or upward through the medium as a result of the orbital motions of the satellites. Because there are about 10^8 wavelengths along the ray path, the phases of these rays change rapidly with time. The relative phases, $\varphi_a - \varphi_m$ and $\varphi_b - \varphi_m$, vary at a rate that depends on the size of the refractivity perturbation that caused the multipath and on the local defocusing by smooth dry air. Typically this difference in rate is below 10 to 15 Hz. The relative values of the amplitudes E_a and E_b , compared to the amplitude of the main ray E_m , determine whether we have deep scintillation or light interference from these particular rays. Each of these three rays has its own excess Doppler signature because the directions of their ray path tangent vectors leaving the transmitting GPS satellite and arriving at the LEO are different. The excess Doppler relationship given in Eq. (1.2-3) is valid for each of these contributing rays, but not to the Doppler for $\hat{E}(t)$. In

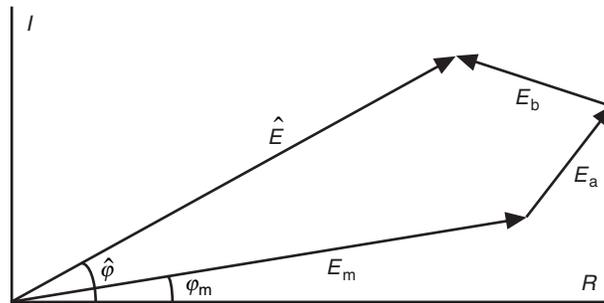


Fig. 1-9. Complex diagram showing composite amplitude \hat{E} and phase $\hat{\varphi}$ from three multipath rays.

other words, $d\hat{\varphi}/dt$ does not yield per se a bending angle from Eq. (1.2-3) because $\hat{E}(t)$ does not represent an actual ray, only a composite.

The interesting point about these multiple rays is that, although they arrive at the LEO at the same time, they all have different excess Doppler values because their tangent vectors at the LEO differ. Thus, from Eq. (1.2-3) it follows that when we have multiple rays arriving concurrently at the LEO, they arrive with distinct bending angles.

Figure 1-10 shows a schematic diagram for a hypothetical multipath scenario in which up to 5 rays concurrently arrive at the LEO. The ordinate is the excess Doppler shift for individual rays, and the abscissa is the epoch of the observation at the LEO. For this scenario, the phase and amplitude contributions from possibly 1, 3, or 5 rays will simultaneously be registered in the total amplitude and phase measured at the LEO. The actual number of contributing rays depends on the observation epoch t_k , $k = 1, 2, \dots$, as shown in Fig. 1-10(a). The measured composite phase will be an unknown combination of the phases of the individual rays weighted by their respective amplitudes per the vector diagram in Fig. 1-9.

1.4.1 Spectral/Holographic Techniques

What this multipath problem, described in Fig. 1-10, needs is a transformation that converts the multi-valued time series of observations (in Doppler or bending angle) into a single-valued series. There are several ways of accomplishing this. One is to convert the multi-valued time series into a single-valued spectral series, as suggested in Fig. 1-10(b). Here ω is a spectral variable, for example, a Fourier variable from a fast Fourier transform (FFT)

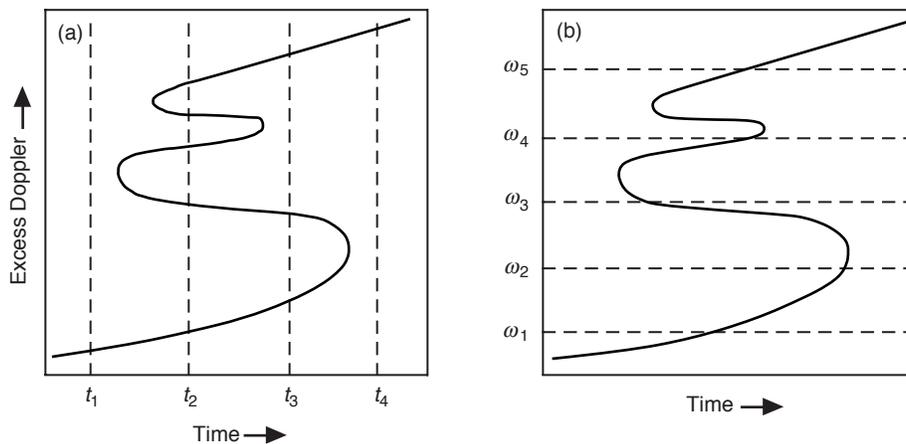


Fig. 1-10. Hypothetical multipath scenario for a setting occultation: (a) the phase at the LEO is measured at the epochs t_k , $k = 1, 2, \dots$, resulting in multiple Doppler contributions at some epochs, and (b) the time series of observations is transformed into a single-valued spectral series.

that can be made proportional to or at least related in a one-to-one way to the excess Doppler. In Fig. 1-10(b), we have converted through a Fourier transform the time series of phase and amplitude measurements of the field, in which the bending angle may not be a unique function of time, into a spectral series in which the bending angle is a unique function of the spectral variable. The peak power in the Fourier spectrum locates the excess Doppler from an individual ray and, therefore, the bending angle per Eq. (1.2-3). This approach was applied to the open-loop phase and amplitude measurements observed from the Earth from the occultation of Voyager 2 by Uranus in 1986 [10]. Figure 1-11 shows an open-loop power spectrum over time, composed of contiguous strips of 10-s temporal width and an excess Doppler breadth of about 15 Hz.

The word “radio-holography” is often used to describe this class of “wave/optic” analysis techniques. Radio-holographic techniques first “stop” the phase rate of the LEO observations by subtracting a time-dependent phase predicted from a realistic model that includes both geometric delay from satellite kinematics and the refractive delay [62–69]. This narrows the effective bandwidth around the spectral peak of the reference ray used in the reference model. The variability of the Doppler shift in the model typically ranges from several hertz to at most a few tens of hertz. Then a complex spectral algorithm,

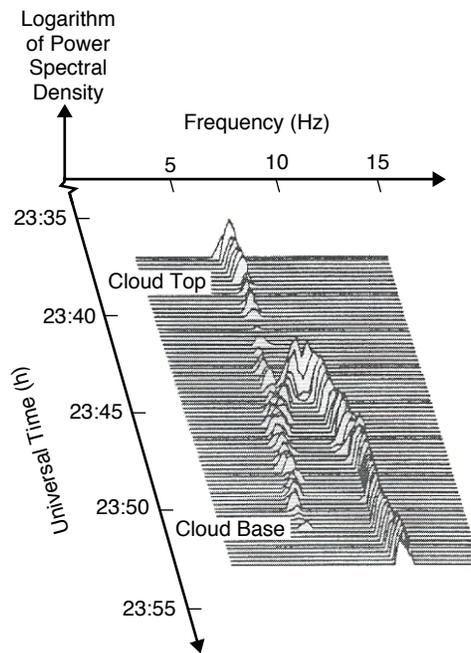


Fig. 1-11. The power spectrum of the radio signal from Voyager 2 while occulting behind Uranus' lower troposphere in 1986. Redrawn from [10].

for example, an FFT, is applied to obtain the power spectrum over the relatively narrow remaining bandwidth of the stopped phase profile within a prescribed time window or time width. Maximum power yields the Doppler tone, or tones when atmospheric multipath is present. From Eq. (1.2-3), one then obtains the bending angle for each tone. The temporal evolution of the tones is obtained by sliding the sample time window used in the FFT. Under the assumption of spherical symmetry, the one-to-one relationship between the recovered excess Doppler/bending angle and the impact parameter of the ray obtained from Bouguer's law yields the value of the impact parameter. From this approach, one recovers a unique bending-angle profile versus impact parameter for a given ray member, but corrupted as always by measurement noise and modeling errors. From this point, the Abel transform yields the refractivity profile versus impact parameter, and then from the relationship $a = r_* n(r_*)$ at a tangency point, one obtains $n(r)$.

Figure 1-12 shows two snapshots of the bending-angle spectra recovered from a GPS/MET occultation over the Sea of Okhotsk in the Russian Far East using a holographic technique [29,54,67,68]. Figure 1-12(a) shows a narrow (half-power width $\sim 20 \mu\text{rad}$) single tone in the mesosphere at 56-km altitude. This translates into a vertical resolution of about 60 m. Figure 1-12(b) shows the complex tone structure about 1/2 minute later for the same occultation. Here the tangency point of the direct ray lies deep in the lower troposphere between 1 and 2 km above sea level. This figure also includes a weaker tone corresponding to a near-specular reflection from the ocean.

1.4.2 Back Propagation

A second, fundamentally different approach to achieving an equivalent single-valued time series and to improving resolution is to map the field measurements recorded by the LEO onto another surface much closer to the refracting medium. This technique is based on the Helmholtz–Kirchoff integral theorem from classical electrodynamics. This integral theorem expresses the amplitude and phase of an electromagnetic wave at a given point in terms of an integral involving the distribution of the amplitude and phase of the wave over an enclosing surface. It explicitly accounts for the retardation time between any point on the radiating surface and the interior point, that is, the travel time between these points resulting from the finite speed of light. The theorem is valid when the scale of the radiating surface is very much larger than the wavelength of the wave. Both the Rayleigh–Sommerfeld and Fresnel–Kirchoff scalar diffraction theories follow directly from this asymptotic theorem [48,70]. These scalar diffraction theories enable both the forward-propagation techniques using a phase screen model to mimic the observations [71–74] and the backward-propagation techniques to map the observations to a back plane [75–77].

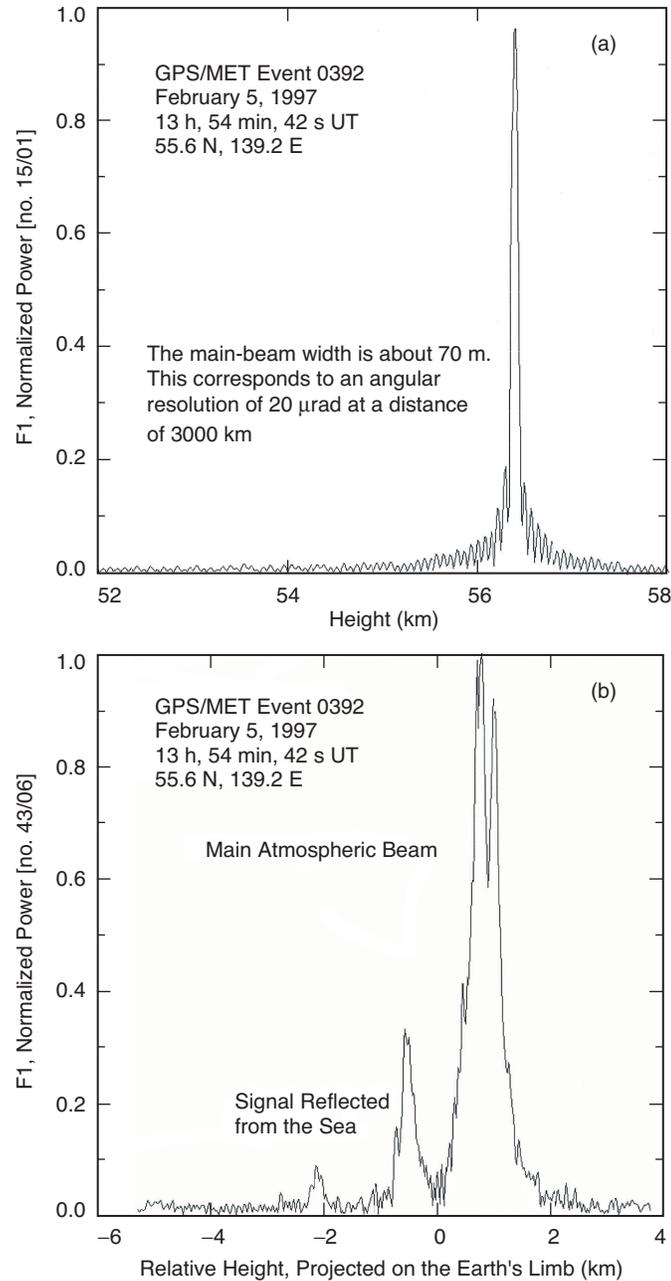


Fig. 1-12. Bending-angle spectra from a GPS/MET occultation using the radio holographic technique [67]: (a) narrow spectral distribution from a single ray in the mesosphere and (b) complex ray structure low in the troposphere, including an ocean surface reflection.

The integral theorem also can be used to propagate a wave through a succession of surfaces. For example, one form of the multiple phase screen technique involves placing successive parallel planes spaced along the paraxial direction of a wave through an inhomogeneous medium. The planes are mounted perpendicular to the paraxial direction, essentially the preferred direction of propagation. The medium between the screens is taken as homogeneous. Therefore, to compensate for the actual inhomogeneous medium between any two adjacent screens, a phase offset is added to the wave in the subsequent screen. This technique can be more tractable for propagating a wave than solving Maxwell's equations. In two-dimensional problems, the multiple phase screen technique involves a one-dimensional scalar diffraction integral applied successively, whereas Maxwell's equations form a second-order system that in general requires simultaneous integration over two dimensions [78–82].

In the multipath problem, the mapped or back-propagated field (toward the emitter) can provide an equivalent virtual time series of phase and amplitude values in another surface that has the favorable property of no or at least substantially fewer multi-valued virtual Doppler points. The one-dimensional back-propagation diffraction integral in a vacuum (see Eq. (A-22) in Appendix A) is given by

$$E(\mathbf{r}_1) = \sqrt{\frac{i}{\lambda}} \int_C \left(\frac{E(\mathbf{r}_2)}{\sqrt{r_{12}}} \exp(-ikr_{12}) (\hat{\mathbf{r}}_{12} \cdot \hat{\mathbf{n}}(\mathbf{r}_2)) \right) ds_2 \quad (1.4-2)$$

where $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$; \mathbf{r}_2 denotes a point at the LEO; ds_2 denotes an incremental arc length along the path C , defined by the trajectory of the LEO over which observations were made during the occultation; and $\hat{\mathbf{n}}(\mathbf{r}_2)$ is the outward unit normal vector to C . The intervening medium is a vacuum. We apply this path integral using the LEO observations of phase and amplitude to obtain the mapped field at the position \mathbf{r}_1 . Because maintaining phase coherence in this diffraction integral is so important, the three-dimensional relative motion of the transmitting GPS satellite and the LEO over the integration span of the occultation observations must be accurately modeled. Also, phase connection in the mapped phase along the back surface must be maintained. In addition, the stationary-phase technique can be used to set practical integration limits in the diffraction integral for the LEO phase and amplitude measurements as a function of the position \mathbf{r}_1 . These and other details are found in [76,77,83].

Figure 1-13 provides a one-dimensional schematic for the concept. Here the LEO travels vertically downward in the LEO plane and vertical distance in this plane is proportional to elapsed time, 2 to 3 km/s. Multiple rays from different altitudes in the atmosphere arrive concurrently at the LEO plane. We use the scalar diffraction integral in Eq. (1.4-2) to propagate the field backward from

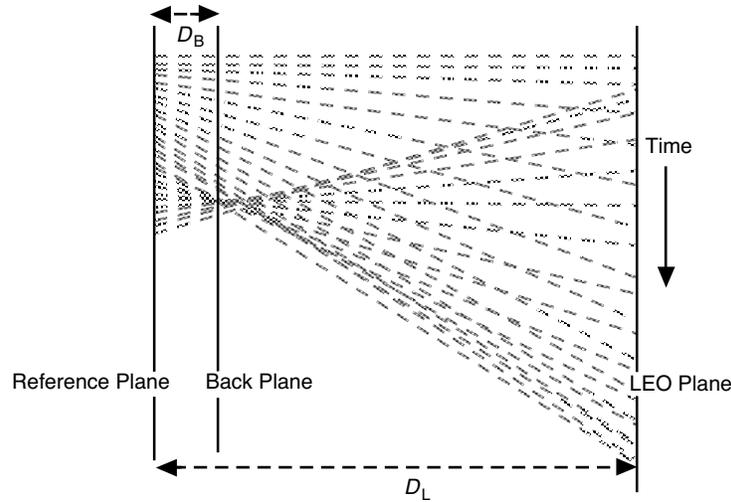


Fig. 1-13. Back propagation geometry. The measured field in the LEO plane is back propagated to the back plane to reduce multipath.

the LEO plane toward the emitter to the position of the back plane mounted perpendicular to the LEO/GPS line. Figure 1-13 suggests that a significant reduction in multiple rays can be achieved from transforming the observations to an equivalent “more focused” set in a back plane [76,77,83], particularly when the actual observation distance is large compared to the vertical scale in the atmosphere over which multipath occurs. The limb distance of a LEO is about 3000 km, but the altitude range in the troposphere from which most multiple rays arrive at the LEO is 10 km or less. Reducing this at least 300 to 1 ratio to 30 to 1 or even 10 to 1 can improve the multipath problem and reduce the Fresnel zone.

Here is a heuristic way of looking at the back-propagation concept, which is further discussed in Chapter 2. To simplify the math, let us place the emitting GPS satellite at infinity, as indicated in Fig. 1-4, and we assume a circular orbit for the LEO. Consider a point (r, θ) well out of the atmosphere ($n \equiv 1$) through which a ray passes after traversing the atmosphere. Suppose we displace that point by a small angular increment to $(r, \theta + \Delta\theta)$, holding the radius r fixed. A new ray arrives at that new point with a new impact parameter value, $a + \Delta a$. Then the difference between the excess phase on the new ray and the excess phase on the old one through the point (r, θ) is given by $\Delta\phi = (\partial\phi / \partial\theta)\Delta\theta$. From Eq. (1.2-3), we have

$$\frac{\partial\phi}{\partial\theta} = k \left(T \cdot \frac{\partial r}{\partial\theta} - \frac{\partial r_{GL}}{\partial\theta} \right) = -k(a - b) = -k \left(D\alpha + a \frac{\alpha^2}{2} + \dots \right) \quad (1.4-3)$$

where $b = |\mathbf{r} \times \mathbf{r}_{GL}| / r_{GL} \rightarrow r \sin \theta$, the impact parameter of the straight line between points G (now placed at infinity along the direction $\theta_G = \pi$ in Fig. 1-6) and L. Also, from Eq. (1.2-4) we have Bouguer's law for the impact parameter of the ray, $a = |\mathbf{T} \times \mathbf{r}| = r \sin(\theta + \alpha)$. For collimated incident rays, the angle γ between the ray and the radius vector is simply $\gamma = \theta + \alpha$. These quantities are indicated in Fig. 1-4. The distance D is given by

$$D = r \cos(\theta + \alpha) \quad (1.4-4)$$

D is essentially the limb distance minus the small increment $a\alpha$. It is the distance between the point (r, θ) and the point (a, θ_a) with $\theta_a = \pi/2 - \alpha(a)$, as shown in Fig. 1-4. The latter point is on the impact parameter space curve associated with the rays after their encounter with the atmosphere. The excess Doppler shown in Fig. 1-10 is proportional to $\partial\varphi/\partial\theta$ in Eq. (1.4-3). Multiplying Eq. (1.4-3) by $\lambda\dot{\theta}_L/2\pi$ yields the relationship between excess Doppler (assuming that $\dot{r}_L = 0$) and the bending angle

$$\lambda f_D = -D\dot{\theta}_L\alpha + O[\alpha^2] \quad (1.4-5)$$

where $\dot{\theta}_L$ is the projection of the LEO orbital rate in the plane of propagation, typically in the range 0.7 to 1.0 mrad/s.

We now construct $\partial^2\varphi/\partial\theta^2$. When θ is varied and r is held fixed, $\Delta a = r \cos(\theta + \alpha)(\Delta\theta + \alpha'\Delta a)$, where $\alpha' = da/da$. Therefore,¹

$$\left. \begin{aligned} \frac{\partial a}{\partial \theta} &= \frac{r \cos(\theta + \alpha)}{1 - \alpha' r \cos(\theta + \alpha)} \\ \frac{\partial b}{\partial \theta} &= r \cos \theta \end{aligned} \right\} \quad (1.4-6)$$

It follows from Eqs. (1.4-3) and (1.4-6) that

$$\frac{\partial^2 \varphi}{\partial \theta^2} = -k \left(\frac{D^2 \alpha'}{1 - D \alpha'} + a \alpha + O[\alpha^2] \right), \quad D = r \cos(\theta + \alpha) \quad (1.4-7)$$

¹ Incidentally, Eq. (1.4-6) provides an expression for estimating the duration of an occultation in the neutral atmosphere. Noting that $\Delta\theta = \int (\partial a / \partial \theta)^{-1} da$, we have

$\Delta T = \dot{\theta}_L^{-1} \int_{a_1}^{a_2} (\partial a / \partial \theta)^{-1} da \doteq (a_2 - a_1 + D(\alpha_1 - \alpha_2)) (\dot{\theta}_L D)^{-1} \approx 50 - 100$ s. The duration strongly depends on the magnitudes of $\dot{\theta}_L$ and α near sea level.

The first point to notice is that, on a highly defocused ray with $|D\alpha'| \gg 1$, it follows that $\partial^2\varphi/\partial\theta^2 \rightarrow kr \cos\theta$. Therefore, the LEO excess Doppler rate approaches essentially a constant in strong defocusing, that is, $df_D/dt \doteq (\partial^2\varphi/\partial\theta^2)\dot{\theta}_L^2/2\pi \rightarrow \dot{\theta}_L^2 r \cos\theta_L/\lambda$, or 10 to 15 Hz/s, depending primarily on the value of $\dot{\theta}_L$. This is an important coherence issue for holographic techniques applied to the lower troposphere. Holographic techniques subtract a model phase from the observations to “stop” their phase variability, i.e., to greatly narrow their spectral bandwidth, making an FFT or a similar complex spectral analysis practicable. The near-constancy of the excess phase acceleration in highly defocused areas is an important factor in the accuracy of the modeled phase profile.

The second point about Eq. (1.4-7) concerns choosing a value for the back-plane distance D_B . Referring to the excess Doppler profile shown in Fig. 1-10 evaluated along the LEO orbit, the undulations in the phase rate curve in the back plane will be greatly diminished if we choose $D_B \ll D$. Basically we want $\partial\varphi/\partial\theta$ to be monotonic, no turning points with respect to θ , or with respect to time, or with respect to any other displacement metric, for example, arc length along the path defined by the intersection of the back plane with the plane of propagation. It follows from Eq. (1.4-7) that this is equivalent to requiring that the defocusing on the back plane be such that $\zeta_B^{-1} = 1 - D_B(d\alpha/da) > 0$ for all values of a spanned by the occultation episode. This condition on the value of back-plane distance D_B , i.e., that $D_B(d\alpha/da) < 1$ for any point along the back plane, ensures monotonicity in excess phase rate versus displacement along the back plane if we can place the back plane close enough to the impact parameter space curve defined by (a, θ_a) . This may not be achievable with a fixed back plane.

Why not set $D_B = 0$? No multipath and no defocusing there. It turns out that that is exactly the right choice, but the problem is that one knows a priori the value of neither α nor a . From Eq. (1.4-4), it follows that setting $D_B = 0$ is equivalent to placing the back plane at the point (a, θ_a) with $\theta_a = \pi/2 - \alpha(a)$. Setting $D_B = 0$ is impossible to uniformly accomplish for all values of a with a fixed back plane because the impact parameter space curve as a function of a is in general not straight, and it is markedly non-linear in multipath zones.

So, with the back-plane methodology, some compromise in choosing D_B must be made. The difference in phase between the point (a, θ_a) on the impact parameter space curve and a nearby point (r_B, θ_B) on the back plane is given to first order by

$$\Delta\varphi = -a\Delta\theta + 0\Delta r \quad (1.4-8)$$

where $\Delta\theta = \theta_B - \theta_a$ and $\Delta r = r_B - a$. The first-order phase change in the radial direction is zero because the ray through the point (a, θ_a) is orthogonal to r_a at that point. We select $r = r_B$ and $\theta = \theta_B = \pi/2 - \alpha_B$ as the fixed point to position the back plane perpendicular to the LEO/GPS line, with α_B given a fixed value. Then the phase difference $\Delta\varphi$ for a given value of a between the point (a, θ_a) on the impact parameter space curve and a nearby point on the back plane is given to first order by $\Delta\varphi = -a(\theta_B - \theta_a) + 0(r_B - r_a) = -a(\alpha(a) - \alpha_B)$. We now have a new caustic possibility in the back plane, $d\varphi/da = \alpha_B - \alpha - a d\alpha/da = 0$. This arises on the back plane whenever it is away from the impact parameter space curve $r = a$, $\theta_a = \pi/2 - \alpha(a)$, generated by varying a . Since most caustics occur where refractivity gradients are large, and therefore where α is large, a good rule of thumb to reduce the probability of caustics occurring on the back plane has been to set $\alpha_B = \alpha_{\text{Max}}$, where α_{Max} is the largest typical bending angle encountered in the lower troposphere, where major multipath is likely to occur, 30 to 40 mrad. This choice places the back plane close to the impact parameter space curve in that vicinity [65,83]. Even though this plane is in the Earth's atmosphere, it is treated as though it were in a vacuum. It serves only as a platform for recovery of the refractivity profile.

We designate the mapped phase using the diffraction integral in Eq. (1.4-2) from the LEO to the back plane as $\varphi_B(s_B)$, where s_B is path length along the back plane perpendicular to the LEO/GPS line. From Eq. (1.2-3) or Eq. (1.4-5), it follows that, when $d\varphi_B/da$ is monotonic in the back plane, then the bending angle of a ray intersecting the back plane can be inferred unambiguously from the directional derivative of $\varphi_B(s_B)$. For collimated incident waves, it follows that

$$\frac{d\varphi_B}{ds_B} = -k \sin \alpha \quad (1.4-9)$$

By assuming spherical symmetry, one obtains the impact parameter for each bending angle using Bouguer's law. From the profile of bending-angle and impact parameter pairs, one can use the standard geometric optics Abel transform technique to recover the refractivity profile. In severe multipath situations, the caustic condition on the back plane can be violated, but experience with actual observations has established that the technique significantly mitigates multipath problems [83].

The back-plane technique gains two benefits, reduced multipath and also enhanced vertical resolution. The Fresnel zone is smaller at the back plane by roughly the factor $\sqrt{|D_B \zeta_B / D_L \zeta_L|}$.

1.4.3 The Canonical Transform Technique

We have noted that rays are stationary-phase paths. Therefore, their space coordinates can be transformed with Hamilton–Jacobi theory into canonical coordinates or ray coordinates. The canonical transform technique [84] essentially uses Hamilton–Jacobi theory to transform the two-dimensional space coordinates (z, x) and their slopes dx/dz for a ray system, that is, for a family of rays generated by varying, for example, the impact parameter a , into three-dimensional phase–space coordinates (z, x, a) . Here x is vertical directed in Fig. 1-4 and z is horizontal directed. In these transformed coordinates, a becomes the canonical conjugate momentum to the variable x . One then applies a Fourier integral transform similar to the scalar diffraction integral in Eq. (1.4-2) to map the field from the phase and amplitude measurements along the LEO path in space coordinates onto the a -plane in phase–space coordinates, from which the phase $\varphi(a)$ is obtained. On this a -plane, $k\alpha = -d\varphi(a)/da$. The technique yields an unambiguous determination of α and a , provided that spherical symmetry holds and super-refractive zones are avoided, and subject, of course, to the quality of the LEO observations and the modeling. The technique fails in super-refractivity zones because it is in part ray theory. The technique avoids the potential back-plane caustic and multipath problems of the general back-propagation technique, and its vertical resolution potential is comparable to wave theory.

1.4.4 The Impact Parameter Space Curve

Another less elegant but conceptually straightforward method for avoiding caustic and multipath problems maps the phase and amplitude measurements made by the LEO along its path to the impact parameter space curve associated with the rays after their encounter with the atmosphere. As Fig. 1-4 indicates, the rays are ordered through a transverse plane located perpendicular to the incoming rays at a position prior to their entering the atmosphere. This pre-encounter impact parameter space curve is a simple straight line perpendicular to the incoming collimated rays, or a circular arc of radius $\rho_G/2$ if the incoming rays are from a spherical wave. This is equivalent to the proposition that the rays are single-valued as a function of impact parameter a before encountering the atmosphere; a one-to-one relationship holds between the value of the impact parameter of the ray intersecting the transverse plane (or arc) and the position of that intersection point on the plane (or arc). If spherical symmetry holds, the rays will continue to be single-valued after passing through the atmosphere, that is, each ray will continue to have a unique value of the impact parameter. Therefore, after transecting a spherical symmetric atmosphere, there will remain a one-to-one relationship between the impact parameter of a ray and the angle through which it was refracted. The point

$r_a = a$, $\theta_a = \pi/2 - \alpha(a)$, located on an outgoing ray, also preserves this one-to-one relationship with that ray, if spherical symmetry holds. Bouguer's law in Eq. (1.2-4) requires it. The curve traced out by the point (r_a, θ_a) as a is varied is the space curve for the impact parameter *after* atmospheric encounter. As mentioned earlier, there will be no caustics on this post-encounter curve, but it is no longer a simple curve because of the dispersive refractive gradient.

It is important to keep in mind that we are not so much interested in ray reconstruction or mapping the electromagnetic field through space as we are in recovering the refractivity profile. Any construction, however non-physical, that facilitates refractivity recovery becomes a valid candidate for this application. In this regard, a fictitious curve placed after a ray's atmospheric encounter that denumerates the impact parameter uniquely along its path should suffice for spherical symmetric geometry because the rays through points in this curve will be unique [77]. This holds even though the impact parameter space curve actually lies within the atmosphere.

On the post-encounter impact parameter space curve, the phase of the intersecting ray with an impact value a can be obtained from the defining integral for phase delay given in Eq. (1.2-2) using Bouguer's law. Let this phase be defined by $\varphi_a(a, \theta_a)$ [with $\theta_a = \pi/2 - \alpha(a)$]. It can be shown from Eq. (1.2-2) (with the GPS transmitting satellite at infinity and the phase referred to the line $\theta = \pi/2$) that

$$\varphi_a(a, \theta_a) = k \left(a\alpha(a) + \int_a^\infty \alpha(\omega) d\omega \right) \quad (1.4-10)$$

The space curve itself is generated by the tip of the vector $\mathbf{r}_a(a)$, with $r_a = a$ and $\mathbf{r}_a \cdot \mathbf{T}_a = 0$, where \mathbf{T}_a is the unit tangent vector of the ray passing through the tip of \mathbf{r}_a . The angle ψ_a between the tangent vector ds_a to the impact parameter space curve and \mathbf{r}_a is given by

$$\left. \begin{aligned} \cos \psi_a &= \frac{\mathbf{T}_a \cdot ds_a}{ds_a} = \frac{1}{\sqrt{1 + (a\alpha')^2}} \\ \sin \psi_a &= \frac{-a\alpha'}{\sqrt{1 + (a\alpha')^2}} \end{aligned} \right\} \quad (1.4-11)$$

with

$$\frac{ds_a}{da} = \sqrt{1 + (a\alpha')^2} \quad (1.4-12)$$

From Eq. (1.4-10), it follows that

$$ka \frac{d\alpha}{da} = \frac{d\varphi_a}{da} \quad (1.4-13a)$$

or

$$\alpha(a) = -\frac{1}{k} \int_a^\infty \frac{1}{a} \frac{d\varphi_a}{da} da \quad (1.4-13b)$$

By replacing φ_a in this integral with the mapped phase $\varphi_a(\mathbf{r}_a)$ obtained from the back-propagating diffraction integral of the LEO observations given in Eq. (1.4-2), we recover the profile for $\alpha(a)$. Note that, from Eq. (1.4-10), $\varphi_a \rightarrow 0$ as $a \rightarrow \infty$ if $\alpha \rightarrow 0$.

Chapter 2 discusses one method for recovering the impact parameter space curve $(a, \theta_a(a))$ starting from an initial known point $(a_1, \theta_a(a_1))$ on the curve. It uses the back-propagation diffraction integral to map by iteration the LEO observations to a converged point $(a_1 + \Delta a, \theta_a(a_1) + \Delta \theta_a)$ near the initial point. Each successive converged point on that curve yields the recovered values for $\varphi_a(a_k)$ and $\alpha(a_k)$, $k=1, 2, \dots$. For a spherical symmetric atmosphere, the recovered impact parameter space curve can be used with the Abel transform to recover $n(a)$. This method also fails in super-refractive layers. Also, Eq. (1.4-13) shows that the actual shape of the (a, θ_a) curve, however complicated it might be, is of no theoretical consequence in recovering $\alpha(a)$, although a complicated shape could slow the iteration process.

The difference between the phase $\varphi(a)$ on the a -plane in phase space, which uses the canonical transform technique, and $\varphi_a(a)$ on the post-encounter curve in space coordinates (a, θ_a) is simply

$$\varphi(a) = \varphi_a(a) - ka\alpha(a) = k \int_a^\infty \alpha da \quad (1.4-14)$$

1.4.5 A Full-Spectrum Wave Theory

The various spectral or holographic approaches and the scalar diffraction approaches, both the forward- and backward-propagation techniques, involve a hybrid of wave and ray theory concepts. In the end, profiles of bending angle versus impact parameter are obtained from which the refractivity is recovered using the Abel integral transform.

In the quest to deal with multipath problems and related resolution issues, there is yet another approach that seems not to have been fully considered in radio occultations. This involves a full-spectrum wave-theoretic treatment of the electromagnetic propagation through a refracting medium. The technique presented here uses a modified Mie scattering approach adapted to an

inhomogeneous medium. It is applicable to the more general case of a scattering surface embedded in a transparent, spherical refracting medium. An interesting feature of this approach is its full-spectral series representation of the wave.

The spectral series is a solution to the Helmholtz equation for the time-independent part of the wave. Each spectral component of integer number l consists of a variable spectral coefficient multiplied by the basis functions, which in their native form are the spherical Hankel functions of integer order l and the spherical harmonic functions. The effect of the inhomogeneity in the propagation medium is contained in the spectral coefficients, leaving the basis functions unchanged from their functional form in a homogeneous medium.

To derive the variation of the spectral coefficients across a spherical surface, one invokes the continuity conditions from electrodynamics that apply to the radial and tangential components of the incident, reflected, and transmitted electromagnetic field vectors, just as is done in Mie scattering theory. A limiting process yields the differential equations describing the variation of the spectral coefficients in a stratified medium. The accuracy of the technique deteriorates when the truncation assumption about the smallness of certain curvature terms in the field equations becomes invalid, for example, near a turning point. The related Wentzel–Kramers–Brillouin (WKB) method has similar problems at a turning point. The technique is not applicable to a medium with significant back scattering, but it is adaptable to an absorbing medium.

In this wave theory, each spectral coefficient contains a spectral density function for the phase delay induced by the refractivity gradient of the atmosphere. Basically the spectral density function gives the cumulative phase delay induced by the atmospheric refractivity gradient on the l th spectral component of a wave that has traveled from outside the atmosphere down to a given radial distance. This spectral density for phase delay, which is a function of the radial coordinate and the spectral number, can be formally written in terms of an integral over the radial coordinate that explicitly contains the refractivity gradient. Without a refractivity gradient, there is no phase delay in the spectral coefficient.

This series representation of the wave can be directly linked to the Fourier transform of the time series of LEO-observed amplitude and phase. The spectral density function can be recovered by taking the Fourier transform of the stopped version of the formal spectral series from wave theory and the Fourier transform of the stopped phase and amplitude observations made by the LEO over time. Equating these two transforms enables recovery of the spectral density function for the phase delay. The refractivity gradient follows from an inversion of the defining integral for the derivative of the spectral density function.

In a wave theory, the fundamental observations are the phase and amplitude measurements of the field, not Doppler observations or the resolving of possible ambiguities contained therein from multipath. There is no need to propagate the field or map it from one surface to another. The spectral series gives the field at an arbitrary point, even inside of the refracting medium, provided turning points are avoided and the curvature truncation assumption is valid. Multipath and caustic rays become almost irrelevant in wave theories. This applies to parabolic equation techniques with multiple phase screens and to the full-spectrum technique described here. A wave theory calculates the phase and amplitude of a caustic ray as accurately as any other ray. The predicted amplitude of the field at the LEO when it is located on a caustic ray is proportional to $|d^2a/d\theta^2|^{1/3}$ in wave theory, unlike the infinite value predicted in (second-order) geometric optics.

In a full-spectrum wave theory, there will be stationary-phase points in spectral number from where the principal contributions to the summation of the spectral series originate. These stationary-phase points in spectral number closely correspond to the impact parameter values (in phase units) of the multiple rays, including caustic contact points. But in the full-spectrum technique, all points in spectral number are swept up together into a spectral density function that holds for all relevant spectral numbers, not just stationary-phase points.

As mentioned above, we are more interested in the refractivity recovery than in calculating the electromagnetic field vector at the LEO. In this regard, the main emphasis in this full-spectrum technique is on computing the spectral density function for the phase delay, a somewhat simpler task than computing the electromagnetic wave from the spectral series itself. Moreover, being fundamentally a result from wave theory, the recovered refractivity profile has the potential for a much finer radial resolution than dictated by the local first Fresnel zone. This also applies to certain holographic and scalar diffraction techniques mentioned above. Theoretical vertical resolution for wave techniques is proportional to the carrier wavelength divided by the angular synthetic aperture of the observations $\dot{\theta}\Delta T$, or roughly $150(\Delta T)^{-1}$ m for a LEO. Here ΔT is in seconds; it is the minimum of either the time span of the selected observation sequence or the coherence limit of the observations. This would suggest a dekameter-level resolution potential, but the actual vertical resolution will be limited by atmospheric vicissitude, such as horizontal variations in refractivity caused by advection and turbulence, and also by other GPS signal structure and data processing limitations [72]. Perhaps 10 percent of the vertical axis of the local first Fresnel zone is a practicable goal, although the evidence for this conjecture is still inchoate at this time.

Full-spectrum wave-theoretic approaches are powerful for a certain class of geometries, but they are notoriously cumbersome. That is why so many

alternate schemes for propagating waves through complex media have evolved over the years since Mie scattering theory was first formulated in the early 20th century. The use of the Helmholtz–Kirkchoff integral theorem for multiple phase screen propagation is one example. Other parabolic wave equation techniques for propagating waves through a medium can be found in [81]. Spherical symmetry or some similar symmetry is almost de rigueur by definition in a full-spectrum wave-theoretic approach. When the wavelength is small compared to the scale of the refracting medium, the spectral series are slow to converge. The basis functions in the spectral representation of the wave, the spherical Bessel functions, and the spherical harmonic functions become difficult to evaluate in their native forms for large values of spectral number.

Nevertheless, accurate asymptotic forms for the Bessel functions in terms of the Airy functions exist that are quite comprehensive and easy to use [85,86]. Most computer mathematics programs carry a full library of the Airy functions of the first and second kind and their derivatives. Also, the asymptotic forms for the spherical harmonic functions for large spectral number are given by sinusoid functions. While asymptotic functions are used throughout the series, the full-spectral character of the wave equation solution is retained. On the issue of convergence, the stationary-phase technique can be used to identify neighborhoods in spectral number that contribute to the summation of the series representation of the wave, thereby eliminating enormous tracts of non-contributing spectral numbers. Moreover, the stationary-phase technique can be used to establish a duality between certain wave-theoretic quantities evaluated at the stationary-phase points in spectral number and the corresponding quantities in geometric optics. For example, when a stationary-phase point exists, and when super-refractivity situations are avoided, then there is a close correspondence between the spectral density function for phase delay evaluated at a stationary-phase point in spectral number and the eikonal function in geometric optics for the corresponding ray. This is the ray with an impact parameter (in phase units) equal to the stationary-phase point in spectral number.

The essential difference between a full-spectrum wave theory and most holographic techniques [62–69] is in the treatment of the wave itself. Loosely speaking, a full-spectrum wave theory involves a pre-stationary-phase process; holographic techniques involve a post-stationary-phase process. In the full-spectrum technique, the time-independent part of a harmonic wave is described by a spectral distribution within an infinite series summation over spectral number, here an equivalent spectral integral. Holographic treatments describe the wave as a finite sum of the complex amplitudes from the multiple rays involved at an observation epoch, as the schematic in Fig. 1-10 suggests. In the recovery process, both techniques apply a spectral analysis to the phase and amplitude measurements, for example, a fast Fourier transform. The full-spectrum technique recovers the spectral distribution for the phase delay

induced by the refractivity gradient, and it does so with a limit in spectral resolution that is imposed by the uncertainty principle in the Fourier analysis, $\sim 1/\Delta T$, where ΔT is the time interval used in the observations. From the spectral derivative of the recovered spectral distribution for the phase delay, the gradient of the refractivity profile is recovered from inverting the formal integral equation that expresses this spectral derivative in terms of the refractivity gradient. From this one obtains the profile of the change in refractivity over the span of impact parameter values set by the time interval ΔT . One could place, if needed, the recovered spectral density for phase delay into the spectral integral for the wave to compute the stationary-phase points of this integral for successive increments in observation epochs. This gives profiles of the impact parameter values of the multipath rays, including caustic rays if they occur there, the bending angles of the rays, and their angular power spectra.

In comparison, a holographic technique obtains the maximum-likelihood estimate for the bending angle and impact parameter associated with each of the finite number of rays detected in the Fourier analysis of the observations within a given time increment, as well as their angular power spectra. And it then applies the Abel integral transform from ray theory to the recovered bending-angle and impact parameter sequences for each ray to obtain the profile of the change in refractivity over the integration limits. The integration limits in impact parameter for each recovered ray are set to obtain the refractivity profile over the impact parameter range applicable to that ray. Appropriate hand-over at each caustic contact point of the integration limit in the Abel integral from the one nascent ray to the other provides the connections between integral segments, and it yields the entire refractivity profile.

If for certain positions of the LEO there are no stationary-phase values in spectral number in the full-spectrum technique, then in ray theory there will be no rays or eikonal functions to give phase delays. But in a wave theory, there still will be a field predicted for such locations, diminished in amplitude probably, but not zero.

The overviews of Chapters 5 and 6 in Section 1.5 give additional details on this wave-theory approach.

1.5 Overview

1.5.1 Chapter 2

Chapter 2 is largely a revision of an earlier Jet Propulsion Laboratory (JPL) publication [87]. Chapter 2 discusses the changes in the phase and signal amplitude profiles that are observed by a LEO during an occultation of a GPS satellite as a result of a sharp change across a spherical surface in a refraction-related property. This would include (1) the refractivity itself, (2) its scale

height, and (3) the lapse rate of the temperature profile. These sharp changes are embedded in an otherwise smoothly varying refractivity profile. Across the tropopause the lapse rate can change abruptly. Strong refractivity gradients often are associated with a water vapor boundary in the lower troposphere or an electron density change in a sporadic E-layer of the ionosphere. A discontinuous refractivity profile could be a useful limiting case.

Chapter 2 uses a combination of geometric optics and scalar diffraction theory in a thin-screen model to interpret these transients in terms of multipath interference, shadow zones, diffraction, and caustics that are likely to be observed by the LEO. In other words, we primarily address the forward-propagation effects of certain precipitous changes in the vertical profile of the atmospheric or ionospheric refractivity, such as a discontinuous change in refractivity or in one of its derivatives. The applicability of the thin-screen approach in a medium with strong refractive gradients is discussed. The stationary-phase technique for interpretation and for aiding the computation of diffraction integrals is discussed.

This hybrid thin-screen/scalar diffraction approach is used to calculate the phase and amplitude perturbations that would be observed by the LEO from these perturbations in refractivity. Although one customarily thinks of diffraction as producing high-frequency fringe effects, which is true, it also can soften or mitigate the harsh effects predicted by geometric optics, partially filling in the troughs in shadow zones and rounding the peaks in flaring regions where caustics are a factor.

Use of an inverse transform technique to sharpen the resolution of localized features is briefly covered, and it is applied to a specific occultation where the ray tangency point crosses a sporadic E-layer.

1.5.2 Chapter 3

Chapter 3 also is a revision of an earlier JPL publication [88]. To prepare the development of a wave theory in a refracting medium, we review basic Mie scattering theory. Chapter 3 discusses the scattered electromagnetic field from a spherical scattering surface in a homogeneous medium. The field is expressed as series solutions to the Helmholtz equation involving spherical Bessel functions and spherical harmonic functions summed over integer spectral number l . In Mie theory, the scattering coefficients in the spectral series solutions are determined by applying the continuity conditions from electrodynamics that must hold across a boundary bearing a discontinuity ΔN in refractivity. The total scattering from the sphere includes all scattering modes, which arise from external reflection, refraction, and internal reflections.

It is well-known that these spectral series converge slowly with spectral number when the radius r_o of the refracting sphere is very large compared to the wavelength λ of the incident wave. Asymptotic forms for the Bessel and

Legendre polynomial basis functions in terms of the Airy functions and the complex exponential functions, which are applicable when $r_o / \lambda \gg 1$, can be used to greatly facilitate the computation of the scattered field. The asymptotic scattering series is evaluated by converting it into an integration of a phasor multiplied by a slowly varying function of l . Neighborhoods around spectral number points yielding stationary-phase values are the principal contributors to the scattering integrals.

Certain attributes of the scattered field, such as phase, amplitude, reflection, bending angle, defocusing, and caustics, are evaluated at the stationary points in spectral number. A close correspondence is established between these attributes in electrodynamics and their analogs in geometric optics when the observation point is sufficiently away from the shadow boundary.

Rainbow caustics arising from one or more internal reflections within the refracting sphere are evaluated and compared to the predictions from the thin-screen model. For small discontinuities in refractivity, rainbow effects are shown to be prominent only within relatively narrow directions. Third-order stationary-phase theory is used to determine their characteristics in terms of LEO position, the value of ΔN , the ratio r_o / λ , and the number of internal reflections prior to being observed.

Special limiting cases, such as scattering by a perfectly reflecting sphere, by a strongly absorbing sphere, and by a sphere with infinite radius also are discussed in terms of stationary-phase concepts with numerical examples.

When the magnitude of the discontinuity in refractivity at the scattering surface is sufficiently small, $|\Delta N| \sim (2\pi\lambda / r_o)^{2/3}$, about 30×10^{-6} for the Earth's radius or equivalently about 20 mrad of refractive bending at the surface, and when the observation point is sufficiently far from the limb, e.g., at the LEO, then one obtains good agreement in amplitude between the computations based on Mie scattering theory and those based on the thin-screen model combined with scalar diffraction theory.

1.5.3 Chapter 4

In Chapter 4, we review a technique that uses the unitary state transition matrix for the system of first-order electromagnetic wave equations in a refracting medium that is stratified and thin, and whose electromagnetic properties are linear [48,89,90]. This approach has been useful for calculating the propagation of an electromagnetic wave through a thin film. Although not essential to the development of the modified Mie scattering approach, several key propagation concepts are introduced here that are used in the Mie formulation. These include the representation of incoming and outgoing standing waves, the use of osculating parameters, their asymptotic forms, asymptotic matching of incoming and outgoing solutions, and evaluation of the accuracy of the osculating parameter technique.

1.5.4 Chapter 5

In wave theory, we are concerned with solutions to Maxwell's equations. For a harmonic wave in a homogeneous medium, the time-independent parts of the electric and magnetic field vectors are solutions to the Helmholtz equation, which become the components in a spectral series representation of the wave. Each component consists of a constant spectral coefficient multiplied by the basis functions used in the spectral series, which in a spherical medium are the spherical Hankel functions and spherical harmonic functions. Chapter 5 develops a spectral representation for a harmonic electromagnetic wave in a spherical refracting medium. The resulting spectral series for the field in the refracting medium also are expressed in terms of the spherical Hankel and spherical harmonic functions as basis functions. The spectral coefficients for these basis functions are derived for a wave propagating through this refracting medium. The spectral coefficients, which in a homogeneous medium are functions only of spectral number, also vary with radial distance in a refracting medium. Through a modified Mie scattering theory, these spectral coefficients also account for, if applicable, scattering from a spherical surface embedded in the refracting medium. The scattering surface could be defined by a discontinuity in refractivity, or in its gradient or a higher derivative, or by a reflecting surface, or by an absorbing surface.

Chapter 5 uses an osculating parameter technique to calculate the spectral coefficients describing wave propagation in a refracting medium. The method for obtaining the spectral coefficients is similar to a certain parabolic equation technique for wave propagation [81]. In a typical parabolic equation technique, one forms the reduced solution $u(\mathbf{r}) = \exp(-i\mathbf{k} \cdot \mathbf{r})\psi(kr)$ and then eliminates less relevant terms. Here $\psi(kr)$ is a solution to the Helmholtz equation and \mathbf{k} is the wave number vector in the paraxial direction. In our formulation, the form is, for a given spectral number l , $u(r) = \exp(-iG[knr])\psi(knr)$, where $-G[knr]$ is the phase delay incurred by the l th spectral component of the wave *only* from the refractive gradient. The "optical" delay is retained in the original Helmholtz solution $\psi(knr)$. In the continuous medium overlying the scattering surface, the solution is applicable when the medium induces negligible back scattering of the wave, i.e., the medium is transparent. The technique can be adapted to an absorbing medium. It is shown for the limiting case, where the spherical stratification approaches a Cartesian stratification, that the osculating parameter solution becomes the WKB solution for a variable index of refraction. The accuracy of this technique and its range of applicability are addressed. The technique fails at or below a "turning point" (which is spectral-number-dependent in wave theory) because the truncation assumption regarding the smallness of certain curvature terms in the defining differential equations for the electromagnetic field is not valid there. The WKB method also fails for the same reason without appropriate connection formulas between

regimes above a turning point and below, the tunneling regime. On the other hand, one must match the incoming and outgoing spectral coefficients at a turning point to ensure the absence of singularities at the origin from the Hankel functions, and to obtain the correct form for the outgoing spectral coefficients. The outgoing coefficients are the only coefficients applicable at the LEO. Individual Hankel functions are singular at the origin, but the sum of the first and second kinds equals the spherical Bessel function of the first kind, which is well-behaved at the origin. This matching of incoming and outgoing coefficients is accomplished by an asymptotic technique, even though the individual forms for the spectral coefficients fail at and below the turning point.

Chapter 5 is relevant to the applicability of the single thin phase screen/scalar diffraction model, discussed in Chapter 2, to a spherical atmosphere with a significant gradient in refractivity. The thin-screen model serves as a surrogate for the actual atmosphere. Both the thin-screen/scalar diffraction approach and the wave-theory approach lead to predictions of the observed phase and amplitude of the wave having passed through an intervening atmosphere and perhaps having encountered an embedded scattering surface. The wave-theory approach explicitly accounts for scattering and for the effects of the refractive gradient in the surrounding medium on the phase and amplitude of the electromagnetic wave propagating through the medium. In comparison, the thin-screen model lumps these effects into a phase delay profile embedded in the thin screen. This profile mimics the extra phase delay experienced by the wave due to the atmospheric refractivity profile. In accounting for the scattering effects from and through the spherical surface of discontinuity, wave theory also includes the possibility that the scattered wave has undergone one or more reflections inside this surface. Interference, shadowing, diffraction, and caustics can be evaluated using both the thin phase screen/scalar diffraction approach and the full-wave theory approach. The level of agreement between these two approaches and how that level depends on the adversity of the wave propagation conditions in the atmosphere are addressed here. Good agreement with thin-screen results is obtained when “thin atmosphere” conditions hold in a large sphere, i.e., no caustic points in the screen, when rainbow caustic directions are avoided, and when the LEO is some distance from the refracting medium.

Stationary-phase concepts are applied to key observables, such as phase delay, propagation direction, and wave amplitude, the stationary-phase values of which are established in spectral number using a phasor representation in the wave-theory approach. Stationary-phase values for these observables also are obtained in impact parameter space in geometric optics. Chapter 5 establishes a duality between stationary-phase variables in spectral number, when they exist, and their counterparts in geometric optics.

Chapter 5 introduces a general spectral density function $G[\rho, \nu]$ for the phase delay in the spectral coefficient induced by the refractive gradient in the

medium. This function of radial coordinate $\rho = krn(r)$ and spectral number ν accounts for the extra phase delay at the radial position ρ on each spectral component induced by the refractive gradient on an incident wave that asymptotically is either planar or spherical at large approaching distances relative to the refracting sphere. Correspondence between this spectral density function for phase delay and key concepts from geometric optics, cumulative bending angle and the eikonal function giving path delay, is established.

Chapter 5 also describes the properties of turning points, caustics, shadow zones, and multipath from a wave-theoretic point of view in a spectral framework. It also discusses these quantities in a second-order geometric optics framework, including its shortcomings near caustics or in dealing with nascent ray pairs with nearly merged impact parameters. Third-order stationary-phase theory is introduced to develop a more accurate ray-theoretic approach near caustics.

Chapter 5 develops a phasor-based approach for evaluating the spectral series using numerical integration combined with the stationary-phase technique. The numerical integration of the spectral representation has been aided by the stationary-phase technique to identify contributing neighborhoods in spectral number, greatly improving their efficiency.

Special topics are addressed—for example, calculating the field when a reflecting surface is embedded in the refracting medium and dealing with turning point computational difficulties where the osculating parameter technique degrades in accuracy. Numerical solutions for the spectral representation of the field at the LEO are presented for various refractivity profiles and scattering surfaces. Numerical examples also include the field at the LEO from refractivity profiles giving multipath, shadow zones, super-refractivity, and caustics, and the field at the LEO from a perfectly reflecting sphere embedded in an overlying refracting medium.

1.5.5 Chapter 6

Here our goal is use the wave-theory concepts in Chapter 5 to develop a technique to recover the refractivity profile from occultation observations of phase and amplitude. The chapter begins with a discussion of GPS receiver operations and ends with a discussion of a recovery technique using the spectral representation developed in Chapter 5. We provide a brief summary here of the recovery sections.

In spherical coordinates, the coplanar components of the time-independent solutions to the Helmholtz equation for the electric field evaluated at the point (ρ, θ) in the plane of propagation may be expressed in spectral series form:

$$\left. \begin{aligned} E_r(r, \theta) &= \frac{E_o}{2\rho^2} \sum_{l=1}^{\infty} l(l+1) \left(a_l^-(\rho) \xi_l^-(\rho) + a_l^+(\rho) \xi_l^+(\rho) \right) P_l^1(\cos \theta) \\ E_\theta(r, \theta) &= \frac{E_o}{2\rho} \sum_{l=1}^{\infty} \left(a_l^-(\rho) \xi_l'^-(\rho) + a_l^+(\rho) \xi_l'^+(\rho) \right) \frac{\partial}{\partial \theta} P_l^1(\cos \theta) \end{aligned} \right\} (1.5-1)$$

where $a_l^\pm(\rho)$ are the spectral coefficients, $\xi_l^\pm(\rho)$ are the spherical Hankel functions of the first (+) and second (-) kinds, $\rho = krn(r)$, and $P_l^m(\cos \theta)$ are the associated Legendre polynomials with $m=1$. In this coordinate frame, the emitting source is located in the direction $\theta = \pi$. Here E_o is an arbitrary constant when the emitting source is infinitely far, but in the finite case it also accounts for space loss from the wave-front curvature. Wave theory allows for the possibility of two polarization modes for the field vectors, the so-called transverse electric (TE) and transverse magnetic (TM) modes. Here we assume a TM wave, i.e., the magnetic field vector is oriented perpendicular to the plane of propagation. The electric field calculated from Eq. (1.5-1) includes both incoming (-) and outgoing (+) waves, and it takes into account the possibility that radiant energy has come to the point (r, θ) via any sector in the refracting sphere, for example, from the bottom limb and from the top limb. But for an occultation geometry from a distant emitting source and when the LEO is located at a point (r_L, θ_L) well into the top and outgoing sectors, $\pi/2 - \alpha_L \gg \theta_L \gg 0$, only the terms with plus signs contribute to the evaluation of these series at the LEO when $r_L / \lambda \gg 1$.

These spectral series converge only after the value of the spectral number becomes comparable to ρ_o , the radial phase distance of the refracting source, about 2×10^8 for GPS wavelengths. To compute the actual electromagnetic field from these series, one uses special techniques, for example, contour integration in the complex spectral number plane, saddle point integration, and numerical integration aided by stationary-phase theory. One also needs a number of asymptotic forms for the Hankel functions and the Legendre polynomials generally valid for large values of the spectral number l and the radial coordinate ρ . For $l, \rho \gg 1$, the Bessel functions, or spherical Hankel functions in this case, can be written accurately in terms of the Airy functions. Similarly, the spherical harmonic functions can be written in terms of sinusoids.

Hardly any of that computational firepower is necessary when one is interested only in recovery of the refractivity profile. The important aspect for refractivity recovery is the variation in phase of each spectral coefficient $a_l^+(\rho)$, which varies with radial coordinate because of the refractive gradient in the medium. For a refracting medium with negligible back scattering and absorption, these coefficients evaluated at the LEO can be written in the form

$$a_l^\dagger(\rho_L) = f(l) \exp(-i2G^\dagger(v)), \quad v = l + \frac{1}{2} \quad (1.5-2)$$

where $f(l)$ is a function of spectral number. The specific functional form of $f(l)$ depends on the asymptotic boundary conditions for the incoming wave from the emitting GPS satellite. Chapter 5 mainly uses a collimated incident wave for the spectral coefficients, but the adjustments to account for a spherical incident wave are noted there, and they are used in Chapter 6. The quantity $-2G^\dagger(v)$ is the spectral density function of the phase delay for the l th spectral component evaluated at the position of the LEO for a wave that has originated from a distant source; both the LEO and the emitter are assumed to be out of the refracting medium. For a homogeneous medium, $G^\dagger(v) \equiv 0$. Otherwise, $G^\dagger(v)$ is obtained from the general spectral density function $G[\rho, v]$, which is given to sufficient accuracy by the integral

$$\left. \begin{aligned} G[\rho, v] &= \int_{\rho}^{\infty} \left(\frac{d \log n}{d \rho} \right) g(\hat{y}) d\rho, \quad \rho \geq \rho^\dagger(v) = v - \hat{y}^\dagger K_v, \quad g(\hat{y}^\dagger) = 0 \\ g(\hat{y}) &\doteq \pi K_v^2 \left(\text{Ai}'[\hat{y}]^2 + \text{Bi}'[\hat{y}]^2 - \hat{y} \left(\text{Ai}[\hat{y}]^2 + \text{Bi}[\hat{y}]^2 \right) \right), \\ \hat{y} &= \hat{y} \left[\frac{v}{\rho} \right] \doteq \frac{1}{4K_v^4} (v^2 - \rho^2), \quad K_v = \left(\frac{v}{2} \right)^{1/3}, \quad v = l + \frac{1}{2} \end{aligned} \right\} \quad (1.5-3)$$

Here $\text{Ai}[\hat{y}]$ and $\text{Bi}[\hat{y}]$ are the Airy functions of the first and second kind, and \hat{y} is their argument; its exact functional form is given in terms of the ratio v/ρ [85]. The quantity $\hat{y}^\dagger = 0.441 \dots$ is the zero point of $g(\hat{y})$, which is unique. The spatial difference between v and ρ^\dagger is $k^{-1} \hat{y}^\dagger K_{\rho_o} \approx 7$ m, almost negligible. For $\rho < \rho^\dagger(v)$, the form for $g(\hat{y})$ in Eq. (1.5-3) does not apply; the correct form rapidly approaches zero for increasing $\hat{y} > 0$.

The interpretation of the general form $G[\rho, v]$ is as follows: $-G[\rho, v]$ is the phase delay accumulated by the spectral coefficient $a_l^-(\rho)$ for a wave that has traveled from infinity (where $dn/d\rho \rightarrow 0$) down through a refracting medium to a radial distance r with $\rho = krn(r)$. For a given spectral number v , the rate of phase accumulation by the corresponding spectral coefficient $a_l^-(\rho)$ rapidly approaches zero for decreasing ρ values less than v . In this regime, these spectral coefficients are unaffected by the refracting medium overhead or by a scattering surface contained therein. Therefore, $G[\rho, v] \rightarrow \text{constant}$ for decreasing $\rho < v$.

The integration variable $\rho = krn(r)$ is used for convenience. Crossing a super-refractivity boundary where $d\rho/dr$ reverses sign requires special treatment, which is discussed in Chapter 6.

$G^\dagger(\nu)$ applies to an outgoing wave evaluated at the LEO outside of the refracting medium. The general form for $G^\dagger(\nu)$ when $d\rho/dr > 0$ is given by

$$G^\dagger(\nu) = G[\rho^\dagger, \nu], \quad \rho^\dagger \doteq \nu - \hat{y}^\dagger K_\nu, \quad K_\nu = \left(\frac{\nu}{2}\right)^{1/3}, \quad \hat{y}^\dagger = 0.441\dots \quad (1.5-4)$$

This connection between the outgoing and general forms for the spectral density functions is accomplished by asymptotically matching the incoming and outgoing spectral coefficients at a turning point to satisfy the general boundary condition that the spectral series shall have no singularity at the origin. It uses the property that $G[\rho, \nu] \rightarrow \text{constant}$ for decreasing $\rho < \rho^\dagger(\nu)$ to match the incoming and outgoing coefficients. The accuracy of this asymptotic matching technique, which involves an approximation, is discussed in Chapters 5 and 6.

For $\rho > \rho^\dagger(\nu)$ and provided that super-refractivity situations are avoided, it can be shown that $\partial G[\rho, \nu]/\partial \nu \doteq \tilde{\alpha}(\rho, \nu)$. Here $\tilde{\alpha}(\rho, \nu)$ is the cumulative one-way refractive bending angle evaluated at ρ for an incoming ray with an impact parameter equal to ν . Thus, $2\tilde{\alpha}(\nu, \nu) = \alpha_L(\nu)$, a virtual bending angle of a ray with an impact parameter value of ν that would be observed at the LEO. If we can recover the profile of $2G^\dagger(\nu)$ versus spectral number, then it is clear from Eqs. (1.5.3) and (1.5-4), and from the close association of $2dG^\dagger(\nu)/d\nu$ with the total virtual bending angle $\alpha_L(\nu)$ under the conditions just cited, that we can recover the profile of $d \log n/d\rho$ versus ρ , and thence $N(\rho)$. Note that scattering and reflecting surfaces that are embedded in the refracting medium, including a discontinuity in $n(r)$, can be calculated using the defining integral for $G[\rho, \nu]$ in Eq. (1.5-3). For the case of a reflecting surface, the complete spectral series breaks conveniently into two series, one for the direct field and one for the reflected field.

One technique for accomplishing the recovery of $2G^\dagger(\nu)$ is to convert the spectral series in Eq. (1.5-1) into an integral representation in phasor form using the asymptotic forms for the Hankel functions and the spherical harmonic functions. This technique also is applicable for evaluating the electromagnetic field at the LEO using numerical integration aided by the stationary-phase technique. This converted series in scalar form is given to sufficient accuracy (for $\nu < \rho_L$) by

$$\left. \begin{aligned}
 E(r_L, \theta_L) &= \frac{E_o}{\sqrt{2\pi\rho_L \sin\theta_L}} \int_0^\infty \left(\frac{\sin\theta_v^L}{\cos\theta_v^G \cos\theta_v^L} \right)^{1/2} \exp(i\Psi) dv, \\
 \Psi &= \sqrt{\rho_G^2 - v^2} + \sqrt{\rho_L^2 - v^2} + v(\theta_v^G + \theta_v^L - \theta_L) - 2G^\dagger(v) - \frac{\pi}{4}, \\
 \theta_v^G &= \sin^{-1}\left(\frac{v}{\rho_G}\right), \quad \theta_v^L = \sin^{-1}\left(\frac{v}{\rho_L}\right), \quad v < \rho_L < \rho_G
 \end{aligned} \right\} (1.5-5)$$

Here $\exp(i\Psi)$ is the phasor. $\Psi = \Psi(\rho_G, \rho_L, \theta_L, v)$ gives the spectral density of the complete phase delay at the LEO position (r_L, θ_L) relative to the emitting GPS satellite, located at (r_G, θ_G) with $\theta_G = \pi$. $\Psi(\rho_G, \rho_L, \theta_L, v)$ includes the geometric delay terms and the term $-2G^\dagger(v)$ for the delay from the refractive gradient. The diagram in Fig. 1-14 shows the total geometric delay from the emitter to the receiver expressed in spectral number space. The term, $(\rho_L^2 - v^2)^{1/2}$, gives the geometric delay in phase along a straight line between the LEO and the tangency point of the line on a sphere of spectral number radius $v < \rho_L$ centered at the origin. The term $v(\theta_L - \theta_v^L)$ is an arc length along this sphere of radius v , and it is subtracted from $(\rho_L^2 - v^2)^{1/2}$ to correct it to the intersection of the sphere with the line $\theta = \pi/2$, which is the fixed reference line for computing phase delays at the LEO. Similarly, the term $(\rho_G^2 - v^2)^{1/2} + v\theta_v^G$ is the geometric phase delay from the emitting GPS satellite along the straight line to the tangency point on the sphere of radius v and thence along the sphere to the line $\theta = \pi/2$.

Although we have set the upper bound in the spectral integral to ∞ , as a practical matter the stationary-phase contributions to the integral come from

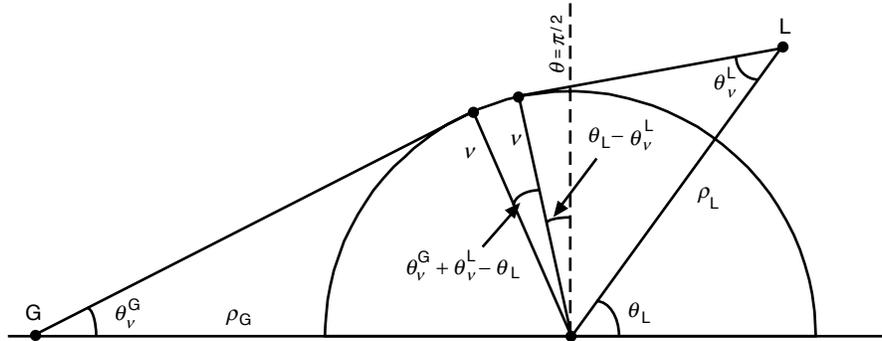


Fig. 1-14. Geometric phase delay diagram in spectral number space.

spectral numbers that are smaller than ρ_L , about 10 percent smaller for a LEO. Therefore, the asymptotic forms given in Eq. (1.5-5) for $\nu < \rho_L$ are valid.

For given satellite positions $(\rho_G, \rho_L, \theta_L)$, the stationary values of Ψ with respect to ν define the neighborhoods in spectral number that contribute to the spectral integral. When a stationary value of Ψ exists, the spectral number providing that stationary value closely equals the impact parameter (in phase units) of the corresponding ray, if near-super-refractivity conditions are avoided. In this case, the stationary value of Ψ essentially equals the value of the eikonal function for the corresponding ray, which is the phase delay φ given in Eq. (1.2-2) for the path integral along the ray (see Eq. (1.5-9) and also Appendix A for its derivation). The stationary values of Ψ with respect to spectral number involve a trade-off between geometric delay and atmospheric delay.

Next, one counter-rotates or “stops” the phase rate of $E(r_L, \theta_L)$ using a realistic model $\varphi_m(\rho_G, \rho_L, \theta_L)$ for the phase, which includes the Doppler shift from the LEO and GPS satellite velocities and the rate of change of the atmospheric phase delay based on geometric optics. Then the phasor in the spectral integral for $E(r_L, \theta_L)$ in Eq. (1.5-5) becomes $\exp[i(\Psi - \varphi_m)]$. This general phase-stopping operation is sometimes referred to as forming a hologram, the coherent mixing of two or more waves. The same holographic operation is performed on the LEO observations, and then a fast Fourier transform (or a similar discrete spectral transform) is performed. Equating these two series, the Fourier transform of the stopped LEO observations to the Fourier transform of the stopped version for $E(r_L, \theta_L)$ in Eq. (1.5-5), leads to an explicit evaluation of $2G^\dagger(\nu)$ versus spectral number. The resolution of this recovery is limited by the uncertainty principle in the discrete spectral analysis technique used on the data. The refractivity profile can be obtained by inverting the integral equation for $2G^\dagger(\nu)$ defined in Eqs. (1.5-3) and (1.5-4).

It is more efficient to recover the refractivity from $2dG^\dagger(\nu)/d\nu$.² It can be shown from Eq. (1.5-5) that

$$2 \frac{dG^\dagger(\nu)}{d\nu} = (\theta_V^G + \theta_V^L - \theta_{L0}) + i\dot{\theta}_L \left. \frac{d \log \hat{E}[\omega]}{d\omega} \right|_{\omega=\omega_\nu}, \quad \omega_\nu = \omega_m + \dot{\theta}_L \nu \quad (1.5-6)$$

where $d\hat{E}[\omega]/d\omega$ is the spectral derivative of the Fourier transform of the stopped $E(r_L, \theta_L)$, θ_{L0} is the orbit angle of the LEO in the plane of propagation

² In wave theory, the reason for using $dG^\dagger(\nu)/d\nu$ instead of $G^\dagger(\nu)$ to recover the refractivity is similar to the reason in geometric optics for using excess Doppler instead of phase. It turns out that $\partial G[\rho, \nu]/\partial \rho = 0$ at $\rho = \rho^\dagger$, but not $\partial G[\rho, \nu]/\partial \nu$.

at the temporal center of the data interval, and $\dot{\theta}_L$ is the component of the LEO orbital angular velocity relative to the occulted GPS satellite in the plane of propagation. Note in Eq. (1.5-6) that $\omega_m \approx \pm 2 \times 10^5$ rad/s, or $\pm(30$ to $35)$ kHz; the sign depends on whether the occultation is rising or setting. On the other hand, within the time interval ΔT over which the Fourier transform is applied, $\omega_v / 2\pi$ varies over only a few tens of hertz at most. Equating $d\hat{E}[\omega_v] / d\omega_v$ in Eq. (1.5-6) to the first derivative of the Fourier transform of the LEO amplitude and stopped phase data yields a determination of the profile for $2dG^\dagger(v) / dv$. This recovered profile for $2dG^\dagger(v) / dv$ is expressed in terms of the refractivity profile through an integral equation derivable from Eqs. (1.5-3) and (1.5-4). This is given by

$$\left. \begin{aligned} 2 \frac{dG^\dagger}{dv} &\doteq 2\pi K_v \int_{\rho^\dagger}^{\infty} \left(\frac{d \log n}{d\rho} \right) \left(\text{Ai}[\hat{y}]^2 + \text{Bi}[\hat{y}]^2 \right) d\rho, \\ \hat{y} &\doteq \frac{1}{4K_v^4} (v^2 - \rho^2), \quad K_v = \left(\frac{v}{2} \right)^{1/3}, \quad \rho^\dagger = v - \hat{y}^\dagger K_v \end{aligned} \right\} \quad (1.5-7)$$

The left-hand side (LHS) of this integral equation is evaluated from the spectral derivative of the Fourier transform of the stopped LEO observations. The RHS is a convolution integral that must be inverted to obtain $\log n(\rho)$ in terms of the recovered sequence of values for $2dG^\dagger / dv$. By replacing the Airy functions with their asymptotic forms applicable for negative values of \hat{y} , $\text{Ai}[\hat{y}]^2 + \text{Bi}[\hat{y}]^2 \rightarrow \pi^{-1}(-\hat{y})^{-1/2}$, one can show that the asymptotic version of the integral in Eq. (1.5-7) equals $\alpha_L(v)$, which is given in Eq. (1.2-6).

One can apply a wave-theory analog of the Abel integral transform to recover $\log n(\rho)$, but in wave theory the kernel in the integral equation giving the weighting distribution of the contributions to $\log n(\rho)$ is not a Dirac delta function, as it is with the Abel transform, but rather it is spread over a characteristic width ΔL . This characteristic width corresponds to $\sim -2 \leq \hat{y} \leq 0$. Over this interval, the Airy functions make the transition from their tunneling forms to their negative argument asymptotic forms (sinusoids) corresponding to ray theory. ΔL is given by

$$\Delta L = 2 \frac{\lambda K_{\rho_o}}{2\pi} = \left(\frac{n_o r_o \lambda^2}{\pi^2} \right)^{1/3} \quad (1.5-8)$$

where r_o is the radius of curvature for the refracting surface. For GPS wavelengths at sea level, $\Delta L \approx 30$ m. Away from super-refractive areas, shadow zones, and caustics, the principal differences in results from a full-

spectrum wave theory versus a wave/optics approach originate from within this relatively narrow altitude band.

Chapter 5 establishes the correspondence between a stationary value with respect to spectral number for the spectral density function of the total delay, $\Psi^* = \Psi(\rho_G, \rho_L, \theta_L, v^*)$, when a stationary value exists, and the phase delay or eikonal function $\mathcal{S}(\rho_G, \rho_L, \theta_L, \rho_*)$ for the corresponding ray. This is the ray with an impact parameter value $\rho_* = ka = kr_*n(r_*) \doteq v^*$, where v^* is the spectral number providing a stationary value for $\Psi(\rho_G, \rho_L, \theta_L, v)$. From geometric optics, one can show [see Appendix A, Eqs. (A-55) and (A-56)] for an emitting GPS satellite located at the point (r_G, θ_G) with $\theta_G = \pi$, that the phase delay at the LEO position (r_L, θ_L) along a ray with an impact parameter value of ρ_* is given by

$$\mathcal{S} = \rho_G \cos(\chi_G + \delta_G) + \rho_L \cos(\chi_L + \delta_L) + \rho_* \alpha(\rho_*) + \int_{\rho_*}^{\infty} \alpha(\omega) d\omega \quad (1.5-9)$$

where δ_G and δ_L are the ray path deflection angles with $\alpha = \delta_G + \delta_L$, and χ_G and χ_L are orbit-related internal angles of the triangle OLG shown in Fig. 1-6. It follows from Eq. (1.5-5) that

$$\Psi(\rho_G, \rho_L, \theta_L, v^*) \Leftrightarrow \mathcal{S}(\rho_G, \rho_L, \theta_L, \rho_*), \quad v^* \doteq \rho_* \quad (1.5-10)$$

At a stationary point in spectral number, $v^* \doteq \rho_*$, and it follows from Fig. 1-14 that $\theta_{v^*}^G + \theta_{v^*}^L - \theta_L \doteq \alpha_L(\rho_*)$. But if for certain positions $(\rho_G, \rho_L, \theta_L)$ there are no stationary-phase values in spectral number for $\Psi(\rho_G, \rho_L, \theta_L, v)$, then in geometric optics there are no rays, no bending angle, and no eikonal function. This would apply to super-refractivity conditions and to strict shadow zones resulting from a discontinuity in the gradient of the refractivity (not necessarily super-refracting) or from an eclipsing limb. But even in these severe situations, $\Psi(\rho_G, \rho_L, \theta_L, v)$ still exists as a spectral density function, and it still provides a value for the field at (r_L, θ_L) according to Eq. (1.5-5). The field won't be zero there, just greatly diminished with damping fluctuations. Chapter 2 discusses wave-theory predictions based on scalar diffraction theory of the amplitude and phase at a LEO from various perturbations in the refractivity profile that compromise ray-theory accuracy in the transition regions (see Figs. 2-10 and 2-11). These include a strict shadow zone in amplitude caused by a discontinuity in the gradient of the refractivity, which is not super-refracting, and also by a discontinuity in refractivity, which is super-refracting.

As already mentioned, the potential radial resolution of this wave-theory technique is essentially proportional to λ / L , where L is the component of the distance traveled by the LEO in the propagation plane perpendicular to the limb

direction over ΔT , the time interval. ΔT may be limited by coherence issues; an error in φ_m used to stop the phase rate results in a loss of coherence in the Fourier integral transform if ΔT is too large. Horizontal variations in refractivity probably pose a larger limitation to resolution in the lower troposphere. A horizontal error δD in the assumed value of the limb distance (effectively from an unknown mesoscale horizontal refractivity gradient) translates into a vertical resolution limit that is equal to at least $500(\delta D/D)^{1/2}$ m [72,73]. A 1 percent error in D translates into a 50-m vertical resolution limit.

The formal integral for $G[\rho, \nu]$ in Eq. (1.5-3) enables one to consider a variety of features embedded in the refracting medium. In addition to a discontinuity in $n(\rho)$ or one of its derivatives, or a specular reflecting surface, or an absorbing medium by adding a complex term to $n(r)$, there are other possibilities. Spectra of recovered bending-angle profiles show broadening to varying degrees, which arise from measurement noise and atmospheric “noise” from turbulence, scintillation, and so on [29,68]. With regard to atmospheric noise, one could treat $n(\rho)$ as a stochastic variable, for example, a first-order Markoff process, with a mean value at any altitude equal to the recovered value, but with statistical parameters that are also recovered from the observations.

1.5.6 Appendices

Miscellaneous topics in geometric optics and wave theory are covered by a number of appendices. Appendix A deals principally with concepts from geometric optics and scalar diffraction theory, rays and refractive bending angle, defocusing, the Fresnel zone, the bending-angle perturbation profile from a perturbation in the refractivity profile, phase delay, and so on. Appendices A through F are intended to accompany Chapter 2. Appendix B briefly discusses a caustic surface for a ray family in terms of envelope theory. Appendix C estimates the separation altitudes of the tangency points of multipath rays as a function of the refractivity perturbation across the boundary. This includes a discussion of the effects of a discontinuity in refractivity or in one of its gradients, and the resulting caustics that can follow. Appendix D deals with third-order stationary-phase theory, an important adjunct when dealing with caustic rays in a ray-theoretic context. An index is developed for setting the accuracy of second-order geometric optics (using third-order theory) in the vicinity of a caustic contact point. Appendix E gives the bending-angle profile versus altitude for a Gaussian refractivity distribution, useful in discussing multipath, quasi-shadow zones, and caustic rays. Appendix F comments on the effect of cycle slips from either receiver operations or recovery analyses. Appendix G gives a short summary of the contour integration technique using the complex spectral number plane for summing spectral series over real integers. Appendix H develops the characteristic matrix for a stack of Airy

layers, which is discussed in Chapter 4. Appendix I summarizes the electromagnetic field equations in a stratified medium and the use of modified scalar potentials to derive the field equations. Appendix J addresses the conditions for near-equivalence between certain wave-theory phase delay quantities and their counterparts in geometric optics.

1.6 Limitations and Simplifications

This book is surely not oriented toward working with actual data. Numerous simplifying assumptions have been made to minimize the impact on an already arcane mathematical framework. Regarding the treatment of actual data from the GPS observations, regretfully we have ignored several aspects essential to use of the occultation technique for science. Not the least of our omissions is a discussion of rendering raw data streams into validated data with minimal cycle breaks, obtaining corrections to clock epoch offsets in transmitters and receivers, dealing with the attendant light-time problem to ensure that the differencing of the phase measurements to eliminate clock offsets occurs at common transmitter and/or receiver epochs, strategies for smoothing and sampling noisy data, applying estimation theory, using precision orbit determination, and, very importantly, using data information systems. The references [34,51,91,92] address many of these aspects. In addition, we have simplified the model in which GPS occultations are assumed to occur, again to minimize the impact on the mathematics. Notable examples are the following.

1.6.1 Ionosphere

Except in Chapter 2, where the thin-screen/scalar diffraction technique has been used on the sporadic E-layer, phase effects from the ionosphere have been ignored. For a comprehensive overview of ionospheric applications, see [93]. Linearly combining the dual-frequency signals from the GPS eliminates most of the ionospheric effect [93–95], but small terms involving f^{-3} and higher degrees remain that can be significant, particularly for tangency points in the upper stratosphere. Modeling the ionosphere to capture the higher-degree terms has had some success. Also, extrapolation of the dual-frequency correction from ray tangency points in the mid-troposphere to the lower troposphere (where the increased noise, particularly on the L2 phase from defocusing and interference, can result in errors in the local dual-frequency correction that exceed the correction) has been useful [59,76].

1.6.2 Placing the Occulted GPS Satellite at Infinity

Placing the emitting GPS satellite at infinity simplifies certain equations, mostly in wave theory. In geometric optics and scalar diffraction theory, this approximation can be compensated for to a certain extent by also replacing the

limb distance of the LEO, D_L , with the reduced limb distance, $D^{-1} = D_L^{-1} + D_G^{-1}$, where D_G is the actual limb distance of the occulted GPS satellite; D is about 10 percent smaller than D_L . With this simplification, the fundamental relationship between excess Doppler f_D and bending angle α of a ray is given by Eq. (1.4-5). $D\dot{\theta}_L$ is essentially the component of the orbital velocity of the LEO in the plane of propagation perpendicular to \mathbf{r}_{LG} , the direction of the GPS/LEO straight line, or, equivalently, $D\dot{\theta}_L$ is the vertical rate of descent or ascent of the GPS/LEO straight line through the atmosphere, 2 to 3 km/s. For the Earth's atmosphere, $\alpha < \sim 0.05$ rad for dry air and water vapor combined, and above the lower troposphere α is less than 0.01 rad. Thus, we have this near-linear relationship between excess Doppler and bending angle given in Eq. (1.4-5).

In contrast, for the case where the GPS satellite is at a finite distance, the appropriate form when the satellites are outside of the refracting medium is given by Eq. (1.2-3). This form can be reduced to geometric quantities in the form

$$\lambda f_D = \mathbf{T}_\perp \cdot (\mathbf{V}_L \delta_L - \mathbf{V}_G \delta_G) + O[\delta_L^2, \delta_G^2] \quad (1.6-1)$$

Here, \mathbf{T}_\perp is a unit vector in the propagation plane perpendicular to \mathbf{r}_{LG} , and δ_L and δ_G are the (small) deflection angles between the vector \mathbf{r}_{LG} and \mathbf{T}_L and \mathbf{T}_G , respectively (see Fig. 1-6). Thus, $\alpha = \delta_L + \delta_G$, and δ_L is about 8 to 9 times larger than δ_G . Nevertheless, dealing with two ray path deflection angles instead of one (even though mutually constrained by Snell's law) and accounting for curvature effects in the approaching wave front from the occulted GPS satellite would further complicate the matter.

As described above, in wave theory a spectral integral representation is used to describe the electromagnetic wave. The main impact of adopting a finite value for \mathbf{r}_G would fall on the asymptotic form that the l th incoming spectral coefficient $a_l^-(\rho)$ would assume for large values of $\rho = knr$ out of the atmosphere. Here we have assumed that the asymptotic form for $a_l^-(\rho)$ corresponds to a collimated wave, a very simple form involving only the spectral number, $i^{l-1}(2l+1)/l(l+1)$. But with \mathbf{r}_G finite, the asymptotic form for $a_l^-(\rho)$ would correspond to a spherical wave $\exp(i\rho_{LG})/\rho_{LG}$ with its center at (r_G, θ_G) with $\theta_G = \pi$ (see Fig. 1-6). Therefore, it is an explicit function of spectral number and ρ_G through a spherical Hankel function $\xi_l^+(\rho_G)$ —not intractable, but certainly more complicated [96]. The asymptotic form of $a_l^-(\rho)$ for the case of a spherical incident wave is briefly discussed in Chapter 5, Eq. (5.5-3b). This asymptotic form involves an additional phase offset,

denoting a shift of the reference point for phase calculation from the line $\theta = \pi/2$ used in the collimated case to the position of the emitter. Also, there is an amplitude term to account for the space loss from a spherical wave. But almost all of the subsequent discussion in Chapter 5 assumes a collimated incident wave. The adjustments to account for a finite GPS satellite distance are briefly presented without derivation in Section 5.10, where ray-theory correspondence is addressed, and in Section 6.4 in connection with refractivity recovery.

1.6.3 Time

We have assumed Newtonian time throughout this monograph. But with actual data when the predicted phase must be computed for a wave moving at the speed of light, several different times and time intervals are involved: proper and coordinate times from special and general relativity, asynchronous timekeeping among all of the receiver and transmitter clocks in the system, transmit time, receive time, topocentric time, atomic time, and coordinated Universal time [34,97]. Keeping track of all these aspects of representing and synchronizing time, essential in actual data processing, unnecessarily complicates our presentation.

1.6.4 Spherical Symmetry

For the terrestrial planets with mild oblateness, the assumption of spherical symmetry is moderately accurate, but for the Jovian planets it was demonstrated about 25 years ago that oblateness must be carefully accounted for [98–100]. The usual approach for the Earth, which has a 20-km difference between its equatorial and polar radii, or 0.3 percent, is to apply the oblateness correction by adjusting the location of the geocenter using a local geoid. For a given geographical location of the tangency point of the occultation, the local radius of curvature vector for the sphere can be adjusted to fit the radius of curvature vector for the local geoid. Thus, a local spherical symmetry is assumed that more or less fits the curvature of the geopotential surface at that geographical location. The radius vector of the local sphere is equated to and aligned with the radius of curvature vector for the geopotential surface. We have assumed spherical symmetry here.

Other aspects of departures from spherical symmetry, for example, horizontal variability in the refractivity, are not considered here [38].

1.6.5 Coplanarity

The assumption of coplanarity between the propagation plane and the satellite orbit planes runs into difficulty mainly when kinematics arise. In Chapter 6, we briefly address this by identifying two different orbit angles that

locate the position of the LEO relative to the occulted GPS satellite, θ in the LEO orbit plane and $\tilde{\theta}$ in the plane of propagation. The obliquity between these two planes is readily expressed in terms of orbit elements, and it is nearly constant over an occultation episode for the neutral atmosphere. Depending on the actual orbits, the obliquity factor $d\tilde{\theta}/d\theta$ ranges between 0.6 and 1.0 for most occultations used for analysis. As discussed above, the term $d\tilde{\theta}/dt$ appears in the expressions for the excess Doppler, but it can be expressed in a near-linear form in terms of $d\theta/dt$ through spherical trigonometry when the obliquity factor is given. Also, departures from planarity from cross-track horizontal variations in refractivity are not considered here.

1.6.6 Circularity

The assumption that the LEO orbit is circular only arises here in connection with the representation of the excess Doppler in terms of the atmospheric bending and the satellite velocity. It is a fairly valid assumption, but with actual data no such approximation is made. The orbits of the LEO and the GPS satellite constellation and their clock epoch offsets are determined independently from POD information based on continuous GPS tracking data from the LEO and from a network of ground stations including the IGS.

1.6.7 Treating the GPS Signal as a Harmonic Wave

The C/A and P ranging codes on the GPS carriers are pseudorandom, phase-modulating square waves that fully suppress the carrier tone. Each transition of a code, occurring at a frequency or chip rate of 1.023 MHz for the C/A code and at 10.23 MHz for the P code, involves a change in phase of the carrier of either zero or 180 deg in accordance with the pseudorandom recipe for that particular code. Therefore, the radio frequency (RF) power spectra of the GPS navigation signals have somewhat complicated shapes. The L1 signal carries both the C/A and P codes, but on the present GPS satellites the L2 carries only the P code; however, all are phase coherent at the transmitter. Both the C/A and P codes produce $(\sin x/x)^2$ -like spread spectra with single-side bandwidths of about 1 and 10 MHz, respectively. The side lobes of the P-code spectrum at the higher harmonics of 10 MHz are attenuated at a higher rate than the $(\sin x/x)^2$ envelope decays to minimize spillover power in nearby user bands. The power in the C/A code is 3 dB greater than the power in the P code. The GPS receiver applies an appropriately time-delayed and Doppler-shifted replica of the transmitted C/A code to the received signal, and carries out a number of cross-correlation and digital signal processing tasks. Through these operations, the receiver isolates the navigation signal emitted by a specific GPS satellite from all others using the orthogonality property of the ranging codes. The receiver in essence collapses the spread-spectrum signal for each carrier

from the selected GPS satellite into a single tone, which is the equivalent of a harmonic wave with a frequency equal to that of the received carrier. Then the receiver measures the phase and amplitude of this equivalent harmonic wave in the presence of noise. Some of the operational aspects of extracting the phase and amplitude measurements from noisy digital data streams are discussed briefly in Chapter 6. For details on digital signal processing by a particular space geodetic GPS receiver, the TurboRogue, and on dealing with the encrypted P code—the so-called anti-spoofing (AS)—for recovery of the L2 phase, see [101,102]. We assume here that the received signal is indeed from harmonic waves with the L1 and L2 frequencies plus the Doppler shifts from satellite motion and from refraction. We also assume that the receiver is reporting at a suitable sample rate, nominally at 50 Hz, perhaps higher if needed, the measured phase and amplitude for this harmonic wave. Also, we assume that phase connection between successive measurements has been successful—a task not always completely achievable with actual data in adverse signal conditions.

1.7 Recommendations for the Next Chapters

Which chapters to pursue depends on one's interest. Chapter 2 and Appendices A through E have a great deal of basic material pertaining to geometric optics, scalar diffraction theory, multipath, shadow zones, caustics, stationary-phase theory, and third-order stationary-phase theory for dealing with caustics.

Chapter 3 discusses basic Mie scattering theory, but with an emphasis on a phasor-based spectral representation for the scattered wave and associated stationary-phase concepts. It can be used as a reference as needed in the later chapters for asymptotic forms, phasor representations, correspondence with ray theory concepts, and so on.

Chapter 4 stands alone. It addresses thin-film theory, which uses a unitary state transition matrix to describe electromagnetic wave propagation in a laminar medium. Several useful concepts found also in Chapters 3 and 5 are developed in Chapter 4, which offers an easier Cartesian framework for their introduction: incoming and outgoing standing waves, osculating parameters, asymptotic matching methodology, turning points, and problems therefrom. It also introduces the Airy layer wherein the refractivity gradient is constant. It is called an Airy layer because the wave equation solutions in this medium are Airy functions. It has important uses in Chapter 5 for asymptotic matching of incoming and outgoing spectral coefficients and for dealing with a turning point.

Chapter 5 provides the basis for the modified Mie scattering theory applicable to a refracting medium with or without an embedded scattering surface.

However, if one's interest is primarily in recovery issues using the full-spectrum wave-theory approach, then the main effort should be on Chapter 6 with occasional reference to Chapter 5 for the basic wave-theory fundamentals regarding the full-spectrum treatment in a spherical symmetric refracting medium.

References

- [1] C. Hazard, "The Method of Lunar Occultations and Its Application to a Survey of the Radio Source 3C212," *Monthly Notices of the Royal Astronomical Society*, vol. 124, pp. 343–357, 1962.
- [2] S. Chandrasekhar, "A Statistical Basis for the Theory of Stellar Scintillation," *Monthly Notices of the Royal Astronomical Society*, vol. 112, pp. 475–483, 1952.
- [3] W. Hubbard, J. Jokopii, and B. Wilking, "Stellar Occultations by Turbulent Planetary Atmospheres: A Wave-Optical Theory Including a Finite Scale Height," *Icarus*, vol. 34, pp. 374–395, 1978.
- [4] E. Saltpeter, "Interplanetary Scintillations. I. Theory," *The Astrophysical Journal*, vol. 147, pp. 433–448, 1967.
- [5] A. Kliore, T. Hamilton, and D. Cain, *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, 1964.
- [6] G. Fjeldbo, *Bistatic Radar Methods for Studying Planetary Ionospheres and Surfaces*, Science Report No. 2, NsG-377, SU-SEL-64-025, Stanford Electronics Laboratories, Stanford, California, 1964.
- [7] A. Kliore, D. Cain, G. Levy, V. Eshleman, G. Fjeldbo, and F. Drake, "Occultation Experiment: Results of the First Direct Measurement of Mars' Atmosphere," *Science*, vol. 149, pp. 1243–1248, 1965.
- [8] G. Fjeldbo, A. J. Kliore, and V. R. Eshleman, "The Neutral Atmosphere of Venus as Studied by Mariner V Radio Occultation Experiments," *The Astronomical Journal*, vol. 76, no. 2, pp. 123–140, 1971.
- [9] E. Marouf, G. Tyler, and P. Rosen, "Profiling Saturn's Rings by Radio Occultation," *Icarus*, vol. 68, pp. 120–166, 1986.
- [10] G. Lindal, J. Lyons, D. Sweetnam, V. Eshleman, D. Hinson, and G. Tyler, "The Atmosphere of Uranus: Results of Radio Occultation Measurements with Voyager 2," *Journal of Geophysical Research*, vol. 92, pp. 14,987–15,001, 1987.
- [11] G. Tyler, "Radio Propagation Experiments in the Outer Solar System with Voyager," *Proceedings of the IEEE*, vol. 75, pp. 1404–1431, 1987.

- [12] G. Lindal, "The Atmosphere of Neptune: An Analysis of Radio Occultation Data Acquired by Voyager 2," *The Astronomical Journal*, vol. 103, pp. 967–982, 1991.
- [13] O. Yakovlev, I. Matyugov, and I. A. Vilkov, "Radio-Wave Phase and Frequency Fluctuations as Observed in Radio Occultation Experiments on the Satellite-to-Satellite Link," *Journal of Communications Technology and Electronics*, vol. 41, no. 11, pp. 993–998, 1996.
- [14] A. S. Liu, *On the Determination and Investigation of the Terrestrial Ionospheric Refractive Indices Using GEOS-3/ATS-6 Satellite-to-Satellite Tracking Data; Final Report*, NASA-CR-156848, November 1, 1978.
- [15] T. Yunck, G. Lindal, and C. H. Liu, "The Role of GPS in Precise Earth Observations," *Proceedings of the IEEE: Position Location and Navigation Symposium*, Orlando, Florida, 1988.
- [16] R. Ware, M. Exner, D. Feng, M. Gorbunov, K. Hardy, B. Herman, Y.-H. Kuo, T. Meehan, W. Melbourne, C. Rocken, W. Schreiner, S. Sokolovskiy, F. Solheim, X. Zou, R. Anthes, S. Businger, and K. Trenberth, "GPS Sounding of the Atmosphere from Low Earth Orbit: Preliminary Results," *Bulletin of the American Meteorological Society*, vol. 77, pp. 19–40, 1996.
- [17] E. Kursinski, G. Hajj, W. Bertiger, S. Leroy, T. Meehan, L. Romans, J. Schofield, D. McCleese, W. Melbourne, C. Thornton, T. Yunck, J. Eyre, and R. Nagatani, "Initial Results of Radio Occultation Observations of Earth's Atmosphere Using the Global Positioning System," *Science*, vol. 271, pp. 1107–1110, 1996.
- [18] G. A. Hajj, E. R. Kursinski, W. I. Bertiger, S. S. Leroy, T. Meehan, L. J. Romans, and J. T. Schofield, "Initial Results of GPS–LEO Occultation Measurements of Earth's Atmosphere Obtained with the GPS/MET Experiment," *Proceedings of the IAG Symposium, G1, GPS Trends in Precise Terrestrial, Airborne, and Spaceborne Applications, IUGG XXI General Assembly, Boulder, Colorado, July 2–14, 1995*, Springer-Verlag, pp. 144–154, 1996.
- [19] C. Rocken, R. Anthes, M. Exner, D. Hunt, S. Sokolovskiy, R. Ware, M. Gorbunov, W. Schreiner, D. Feng, B. Herman, Y.-H. Kuo, and X. Zou, "Analysis and Validation of GPS/MET Data in the Neutral Atmosphere," *Journal of Geophysical Research*, vol. 102, pp. 29,849–29,866, 1997.

- [20] J. Wickert, T. Schmidt, C. Marquardt, Ch. Reigber, K.-H. Neumayer, G. Beyerle, R. Galas, and L. Grunwaldt, "GPS Radio Occultation with CHAMP: First Results and Status of the Experiment," *Proceedings of IAG Scientific Assembly, Vistas for Geodesy in the New Millennium*, September 2–7, 2001, Budapest, Hungary, Springer Series, vol. 125, pp. 273–278, 2002.
- [21] G. Hajj, C. Ao, B. Iijima, D. Kuang, E. Kursinski, A. Mannucci, T. Meehan, L. Romans, M. de la Torre Juarez, and T. Yunck, "CHAMP and SAC-C Atmospheric Occultation Results and Intercomparisons," submitted to *Journal of Geophysical Research, Atmosphere* August 2, 2002.
- [22] R. Anthes, C. Rocken, and Y. H. Kuo, "Applications of COSMIC to Meteorology and Climate," *Terrestrial, Atmospheric and Ocean Sciences*, vol. 11, pp. 115–156, 2000.
- [23] Y.-H. Kuo, X. Zou, F. Vandenberghe, W. Huang, and R. Anthes, "Assimilation of GPS/MET Data for Numerical Weather Prediction," *Terrestrial, Atmospheric and Ocean Sciences*, vol. 11, pp. 157–186, 2000.
- [24] C. Rocken, Y.-H. Kuo, W. Schreiner, D. Hunt, S. Sokolovskiy, and C. McCormick, "COSMIC System Description," *Terrestrial, Atmospheric and Oceanic Science*, vol. 11, pp. 21–52, 2000.
- [25] T. Yunck, C. H. Liu, and R. Ware, "A History of GPS Sounding," *Terrestrial, Atmospheric and Ocean Sciences*, vol. 11, pp. 1–20, 2000.
- [26] T. P. Yunck, G. A. Hajj, E. R. Kursinski, J. A. LaBrecque, S. T. Lowe, M. M. Watkins, and C. McCormick, "AMORE: An Autonomous Constellation Concept for Atmospheric and Oceanic Observation," *Acta Astronautica*, vol. 46, no. 2-6, pp. 355–364, 2000.
- [27] B. Ahmad and G. Tyler, "The Two-Dimensional Resolution Kernel Associated with Retrieval of Ionospheric and Atmospheric Refractivity Profiles Using the Abelian Inversion of Radio Occultation Ohase Data," *Radio Science*, vol. 33, pp. 469–478, 1997.
- [28] A. Pavelyev, A. Zakharov, A. Kucherjavenkov, E. Molotov, I. Siderenko, I. Kucherjavenkova and D. Pavelyev, "Propagation of Radio Waves Reflected from Earth's Surface at Grazing Angles between a Low-Orbit Space Station and a Geostationary Satellite," *Journal of Communications Technology and Electronics*, vol. 42, no. 1, pp. 51–57, 1997.
- [29] G. Beyerle, K. Hocke, J. Wickert, T. Schmidt, C. Marquardt, and Ch. Reigber, "GPS Radio Occultations with CHAMP: A Radio Holographic Analysis of GPS Signal Propagation in the Troposphere and Surface Reflections," *Journal of Geophysical Research*, vol. 107, no. D24, 4802, doi: 10.1029/2001JD001402, 2002.

- [30] *Global Positioning System*, ed. P. M. Janiczek, vols. 1–3, Washington, DC: Institute of Navigation, 1980.
- [31] B. Parkinson and J. Spilker, eds., *Global Positioning System: Theory and Application*, American Institute of Aeronautics and Astronautics, Washington, DC, 1996.
- [32] G. Gendt, P. Fang, and J. Zumberge, “Moving IGS Products Towards Real-Time,”
ftp://igsceb.jpl.nasa.gov/IGSCB/resource/Pubs/99_tec/99tr_AC_WS_7.pdf
1999.
- [33] R. Galas, J. Wickert, and W. Burghardt, “High Rate Low Latency GPS Ground Tracking Network for CHAMP,” *Physics and Chemistry of the Earth*, part A, vol. 26, issues 6–8, pp. 649–652, 2001.
- [34] G. Hajj, E. Kursinski, L. Romans, W. Bertiger, and S. Leroy, “A Technical Description of Atmospheric Sounding by GPS occultation,” *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 64, pp. 451–469, 2002.
- [35] J. Zumberge, and G. Gendt, “The Demise of Selective Availability and Implications for the International GPS Service,” *Physics and Chemistry of the Earth*, part A, vol. 26, issues 6–8, pp. 637–644, 2001.
- [36] G. Beutler, M. Rothacher, S. Schaer, T. A. Springer, J. Kouba, and R. E. Neilan, “The International GPS Service (IGS): An Interdisciplinary Service for Earth Sciences,” *Advances in Space Research*, vol. 23, no. 4, pp. 631–653, 1999.
- [37] R. Neilan, J. Zumberge, G. Beutler, and J. Kouba, “The International GPS Service: A Global Resource for GPS Applications and Research,” *Proc. ION-GPS-97*, The Institute of Navigation, pp. 883–889, 1997.
- [38] B. Ahmad and G. Tyler, “Systematic Errors in Atmospheric Profiles Obtained from Abelian Inversion of Radio Occultation Data: Effects of Large Scale Horizontal Gradients,” *Journal of Geophysical Research*, vol. 104, no. D4, pp. 3971–3992, 1999.
- [39] S. Sokolovskiy and D. Hunt, “Statistical Optimization Approach for GPS/MET Data Inversions,” *URSI GPS/MET Workshop*, Tucson, Arizona, February 1996.
- [40] R. Phinney and D. Anderson, “On the Radio Occultation Method for Studying Planetary Atmospheres,” *Journal of Geophysical Research*, vol. 73, no. 5, pp. 1819–1827, 1968.
- [41] G. Fjeldbo and V. Eshelman, “The Atmosphere of Mars Analyzed by Integral Inversion of the Mariner IV Occultation Data,” *Planetary Space Science*, vol. 16, pp. 1035–1059, 1968.

- [42] E. Kursinski, G. Hajj, J. Schofield, R. Linfield, and K. Hardy, "Observing Earth's Atmosphere with Radio Occultation Measurements Using the Global Positioning System," *Journal of Geophysical Research*, vol. 102, pp. 23,429–23,465, 1997.
- [43] O. Tetens, "Über Einige Meteorologische Begriffe," *Zeitschrift für Geophysik*, vol. 6, pp. 297–309, 1930.
- [44] J. Eyre, "Assimilation of Radio Occultation Measurements into a Numerical Weather Prediction System," ECMWF Technical Memorandum 199, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, 1994.
- [45] X. Zou, Y.-H. Kuo, and Y.-R. Guo, "Assimilation of Atmospheric Radio Refractivity Using a Nonhydrostatic Adjoint Model," *Monthly Weather Review*, vol. 123, pp. 2229–2249, 1995.
- [46] X. Zou, F. Vandenberghe, B. Huang, M. Gorbunov, Y.-H. Kuo, S. Sokolovskiy, J. Chang, J. Sela, and R. Anthes, "A Ray-Tracing Operator and Its Adjoint for the Use of GPS/MET Refraction Angle Measurements," *Journal of Geophysical Research, Atmospheres*, vol. 104, no. D18, pp. 22,301–22,318, 1999.
- [47] Y.-H. Kuo, X. Zou, F. Vandenberghe, W. Huang, and R. Anthes, "Assimilation of GPS/MET Data for Numerical Weather Prediction," *Terrestrial, Atmospheric and Oceanic Sciences*, 1999.
- [48] M. Born and E. Wolf, *Principles of Optics*, 6th ed., Oxford, United Kingdom: Pergamon Press, 1980.
- [49] R. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals*, New York: McGraw-Hill, 1965.
- [50] S. Leroy, "Measurement of Geopotential Heights by GPS Radio Occultation," *Journal of Geophysical Research*, vol. 102, pp. 6971–6986, 1997.
- [51] K. Hocke, "Inversion of GPS Meteorology Data," *Annales Geophysicae*, vol. 15, pp. 443–450, 1997.
- [52] K. Hocke, T. Tsuda, and A. de la Torre, "A Study of Stratospheric GW Fluctuations and Sporadic E at Midlatitudes with Focus on Possible Orographic Effect of Andes," *Journal of Geophysical Research*, vol. 107, no. D20, 4428, doi: 10.1029/2001JD001330, October 2002.
- [53] K. Igarashi, A. Pavelyev, K. Hocke, D. Pavelyev, and J. Wickert, "Observation of Wave Structures in the Upper Atmosphere by Means of Radio Holographic Analysis of the Radio Occultation Data," *Advances in Space Research*, vol. 27, no. 6–7, pp. 1321–1326, 2001.

- [54] A. Pavelyev, K. Igarashi, Ch. Reigber, K. Hocke, J. Wickert, G. Beyerle, S. Matyugov, A. Kucherjavenkov, D. Pavelyev, and O. I. Yakovlev, "First Application of the Radio Holographic Method to Wave Observations in the Upper Atmosphere," *Radio Science*, vol. 37, no. 3, pp. 15-1–15-11, 2002.
- [55] *Atmospheric Profiling Mission*, ESA Report SP-1196, no. 7, European Space Agency, April 1996.
- [56] Y.-H. Kuo, X. Zou, and W. Huang, "The Impact of GPS Data on the Prediction of an Extratropical Cyclone: An Observing System Simulation Experiment," *Dynamics of Atmospheres and Oceans*, vol. 27, pp. 439–470, 1997.
- [57] E. K. Smith, Jr., and S. Weintraub, "The Constants in the Equation for Atmospheric Refractive Index at Radio Frequencies," *Proceedings of the IRE*, vol. 41, pp. 1035–1037, 1953.
- [58] D. Thayer, "An Improved Equation for the Radio Refractive Index in Air," *Radio Science*, vol. 9, pp. 803–807, 1974.
- [59] G. A. Hajj and L. J. Romans, "Ionosphere Electron Density Profiles Obtained with the Global Positioning System: Results from the GPS/MET Experiment," *Radio Science*, vol. 33, no. 1, pp. 175–190, 1998.
- [60] M. E. Gorbunov, S. V. Sokolovsky, and L. Bengtsson, *Space Refractive Tomography of the Atmosphere: Modeling of Direct and Inverse Problems*, Report 210, Max Planck Institute for Meteorology, Hamburg, Germany, 1996.
- [61] T. Meehan, personal communication, Jet Propulsion Laboratory, Pasadena, California, 1995.
- [62] A. G. Pavelyev, A. V. Volkov, A. I. Zakharov, S. A. Krutikh, and A. I. Kucherjavenkov, "Bistatic Radar as a Tool for Earth Investigation Using Small Satellites," *Acta Astronautica*, vol. 39, no. 9–12, pp. 721–730, 1996.
- [63] K. Hocke, A. Pavelyev, O. Yakovlev, L. Barthes, and N. Jakowski, "Radio Occultation Data Analysis by the Radioholographic Method," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 61, pp. 1169–1177, 1999.
- [64] K. Igarashi, A. Pavelyev, K. Hocke, D. Pavelyev, I. Kucherjavenkov, S. Matyugov, A. Zakharov, and O. Yakovlev, "Radio Holographic Principle for Observing Natural Processes in the Atmosphere and Retrieving Meteorological Parameters from Radio Occultation Data," *Earth, Planets and Space*, vol. 52, no. 11, pp. 893–899, 2000.

- [65] M. E. Gorbunov, A. S. Gurvich, and L. Kornbluh, "Comparative Analysis of Radioholographic Methods of Processing Radio Occultation Data," *Radio Science*, vol. 35, no. 4, pp. 1025–1034, 2000.
- [66] S. Sokolovsky, "Modeling and Inverting Radio Occultation Signals in the Moist Troposphere," *Radio Science*, vol. 36, no. 3, pp. 441–458, 2001.
- [67] A. Pavelyev, Y.-A. Liou, Ch. Reigber, J. Wickert, K. Igarashi, K. Hocke, and C.-Y. Huang, "GPS Radio Holography as a Tool for Remote Sensing of the Atmosphere and Mesosphere from Space," *GPS Solutions*, vol. 6, pp. 100–108, 2002.
- [68] K. Igarashi, K. Pavelyev, J. Wickert, K. Hocke, and D. Pavelyev, "Application of Radio Holographic Method for Observation of Altitude Variations of the Electron Density in the Mesosphere/Lower Thermosphere Using GPS/MET Radio Occultation Data," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 64, no. 8–11, pp. 959–969, 2002.
- [69] O. I. Yakovlev, A. G. Pavelyev, S. S. Matyugov, D. A. Pavelyev, and V. A. Anufriev, *Radioholographic Analysis of Radiooccultation Data*, Scientific Technical Report 02/05, GeoForschungsZentrum-Potsdam, Potsdam, Germany, 2002.
- [70] J. Goodman, *Introduction to Fourier Optics*, New York: McGraw-Hill, 1968.
- [71] E. N. Bramley, "The Accuracy of Computing Ionospheric Radiowave Scintillation by the Thin-Phase Screen Approximation," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 39, pp. 367–373, 1977.
- [72] E. Marouf, G. Tyler, and P. Rosen, "Profiling Saturn's Rings by Radio Occultation," *Icarus*, vol. 68, pp. 120–166, 1986.
- [73] W. G. Melbourne, E. S. Davis, C. B. Duncan, G. A. Hajj, K. R. Hardy, E. R. Kursinski, T. K. Meehan, L. E. Young, and T. P. Yunck, *The Application of Spaceborne GPS to Atmospheric Limb Sounding and Global Change Monitoring*, JPL Publication 94-18, Jet Propulsion Laboratory, Pasadena, California, April 1994.
- [74] M. Meincke, *Inversion Methods for Atmospheric Profiling with GPS Occultation*, Scientific Report No. 99-11, Danish Meteorological Institute, Copenhagen, Denmark, 1999.
- [75] A. J. Devaney, "A Filtered Backpropagation Algorithm for Diffraction Tomography," *Ultrasonic Imaging*, vol. 4, pp. 336–350, 1982.
- [76] M. E. Gorbunov, A. S. Gurvich, and L. Bengtsson, *Advanced Algorithms of Inversion of GPS/MET Satellite Data and Their Application to Reconstruction of Temperature and Humidity*, Report 211, Max Planck Institute for Meteorology, Hamburg, Germany, 1996.

- [77] E. Karayel and D. Hinson, "Sub-Fresnel Scale Vertical Resolution in Atmospheric Profiles from Radio Occultation," *Radio Science*, vol. 32, no. 2, pp. 411–423, 1997.
- [78] R. H. Hardin and F. D. Tappert, "Applications of the Split-Step Fourier Method to the Numerical Solution of Nonlinear and Variable Coefficient Wave Equations," *SIAM Review*, vol. 15, p. 423, 1973.
- [79] D. Knepp, "Multiple Phase-Screen Calculation of the Temporal Behavior of Stochastic Waves," *Proceedings of the IEEE*, vol. 71, no. 6, pp. 722–737, 1983.
- [80] J. Martin and S. Flatté, "Intensity Images and Statistics from Numerical Simulation of Wave Propagation in 3-D Random Media," *Applied Optics*, vol. 27, no. 11, pp. 2111–2126, 1988.
- [81] M. Levy, *Parabolic Equation Methods for Electromagnetic Wave Propagation*, London, United Kingdom: The Institution for Electrical Engineers, 2000.
- [82] G. Beyerle, M. Gorbunov, and C. Ao, "Simulation Studies of GPS Radio Occultation Measurements," *Radio Science*, in press, 2003.
- [83] C. Ao, G. Hajj, T. Meehan, S. Leroy, E. Kursinski, M. de la Torre Juárez, B. Iijima, and A. Mannucci, "Backpropagation Processing of GPS Radio Occultation Data," in *CHAMP Science Meeting*, Ch. Reigber, H. Lühr, and P. Schwintzer, eds., January 21–24, 2002, Springer-Verlag, Berlin, Germany, pp. 415–422, 2003.
- [84] M. E. Gorbunov, "Canonical Transform Method for Processing Radio Occultation Data in the Lower Troposphere," *Radio Science*, vol. 37, no. 5, 1076, doi:10.1029/2000RS002592, 2002.
- [85] M. Abramowitz and I. Stegun, eds., *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards Series 55, Washington, DC, 1964.
- [86] F. Olver, *Introduction to Asymptotics and Special Functions*, New York: Academic Press, 1974.
- [87] W. G. Melbourne, *Sensing Atmospheric and Ionospheric Boundaries in GPS Radio Occultation Observations from a Low Earth Orbiter, Part 1*, JPL Publication 99-5, Jet Propulsion Laboratory, Pasadena, California, July 1998.
- [88] W. G. Melbourne, *Sensing Atmospheric and Ionospheric Boundaries in GPS Radio Occultation Observations from a Low Earth Orbiter, Part 2*, JPL Publication 99-5, Jet Propulsion Laboratory, Pasadena, California, April 1999.

- [89] F. Abelès, “Investigations on the Propagation of Sinusoidal Electromagnetic Waves in Stratified Media. Applications to Thin Films,” *Annales de Physique*, vol. 5, pp. 596–640, September–October 1950 (in French).
- [90] F. Abelès, “Investigations on the Propagation of Sinusoidal Electromagnetic Waves in Stratified Media. Application to Thin Films. II. Thin Films,” *Annales de Physique*, vol. 5, pp. 706–782, November–December 1950 (in French).
- [91] J. Wickert, G. Beyerle, T. Schmidt, C. Marquardt, R. König, L. Grunwald, and C. Reigber, “GPS Radio Occultation with CHAMP,” *First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies*, eds. C. Reigber, H. Luhr, and P. Schwintzer, proceedings from the First CHAMP Science Meeting, Potsdam, Germany, January 21–24, 2002, Springer-Verlag, 2002.
- [92] <http://genesis.jpl.nasa.gov/>
- [93] G. Hajj, L. Lee, X. Pi, L. Romans, W. Schreiner, P. Strauss, and C. Wang, “COSMIC GPS Ionosphere Sensing and Space Weather,” *Terrestrial, Atmospheric and Ocean Sciences*, vol. 11, pp. 235–272, 2000.
- [94] S. Bassiri and G. Hajj, “Higher-Order Ionospheric Effects on the Global Positioning System Observables and Means of Modeling Them,” *Manuscripta Geodaetica*, vol. 18, pp. 280–289, 1993.
- [95] V. V. Vorob’ev and T. G. Krasil’nikova, “Estimation of the Accuracy of the Atmospheric Refractive Index Recovery from Doppler Shift Measurements at Frequencies Used in the NAVSTAR System,” *Physics of the Atmosphere and Ocean*, vol. 29, no. 5, pp. 602–609, 1994.
- [96] J. Jackson, *Classical Electrodynamics*, 2nd ed., New York: John Wiley & Sons, Inc., 1975.
- [97] T. D. Moyer, *Formulation for Observed and Computed Values of Deep Space Network Data Types for Navigation*, Monograph 2, Deep Space Communications and Navigation Series, Jet Propulsion Laboratory, Pasadena, California, October 2000.
- [98] W. Hubbard, “Ray Propagation in Oblate Atmospheres,” *Icarus*, vol. 27, pp. 387–388, 1976.
- [99] A. J. Kliore, P. M. Woiceshyn, and W. B. Hubbard, “Temperature of the Atmosphere of Jupiter from Pioneer 10/11 Radio Occultations,” *Geophysical Research Letters*, vol. 3, pp. 113–116, 1976.
- [100] G. F. Lindal, D. N. Sweetnam, and V. R. Eshleman, “The Atmosphere of Saturn: An Analysis of the Voyager Radio Occultation Measurements,” *The Astronomical Journal*, vol. 90, pp. 1136–1146, 1985.

- [101] J. B. Thomas, *Signal-Processing Theory for the TurboRogue Receiver*, JPL Publication 95-6, Jet Propulsion Laboratory, Pasadena, California, 1995.
- [102] C. Ao, T. Meehan, G. Hajj, A. Mannucci, and G. Beyerle, "Lower Troposphere Refractivity Bias in GPS Occultation Retrievals," submitted to *Journal of Geophysical Research*, 2003.