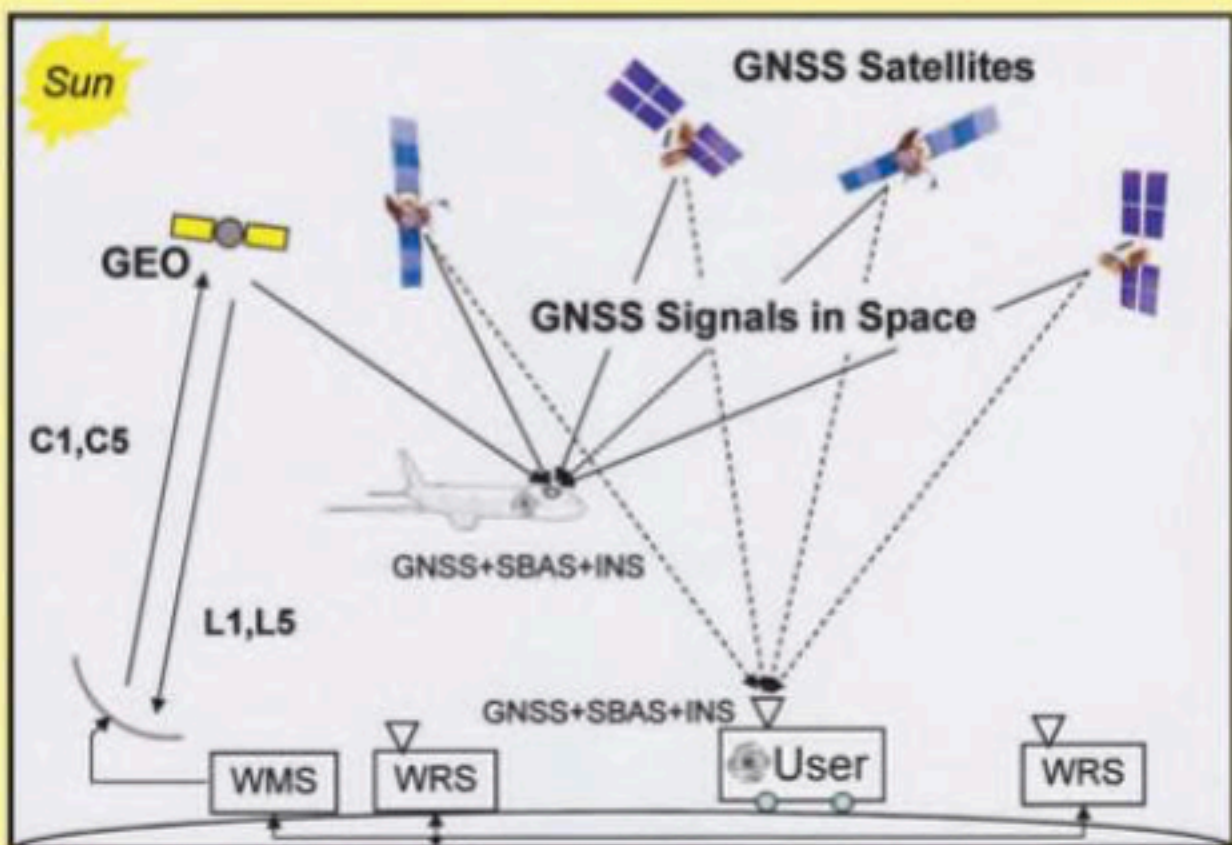


Third Edition

Global Navigation Satellite Systems, Inertial Navigation, and Integration

Mohinder S. Grewal, Angus P. Andrews, and Chris G. Bartone



**GLOBAL NAVIGATION
SATELLITE SYSTEMS,
INERTIAL NAVIGATION,
AND INTEGRATION**

GLOBAL NAVIGATION SATELLITE SYSTEMS, INERTIAL NAVIGATION, AND INTEGRATION

THIRD EDITION

**MOHINDER S. GREWAL
ANGUS P. ANDREWS
CHRIS G. BARTONE**

 **WILEY**

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2013 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Grewal, Mohinder S.

Global navigation satellite systems, inertial navigation, and integration / Mohinder S. Grewal, Angus P. Andrews, Chris G. Bartone. – Third edition.

pages cm

Includes index.

Originally published under title: Global positioning systems, inertial navigation, and integration.

ISBN 978-1-118-44700-0 (cloth)

1. Global Positioning System. 2. Inertial navigation. 3. Kalman filtering. I. Andrews, Angus P. II. Bartone, Chris G. III. Title.

G109.5.G74 2013

910.285–dc23

2012032753

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

M.S.G. dedicates this book to the memory of his parents, Livlin Kaur and Sardar Sahib Sardar Karam Singh Grewal.

A.P.A. dedicates his contributions to his wife, Jeri, without whom it never would have happened.

C.G.B dedicates this work to his wife, Shirley, and two sons, Christopher and Stephen, for their never-ending support over the years.

GNSS DATA ERRORS

7.1 DATA ERRORS

Ranging errors are typically grouped into six classes:

1. *Ionosphere.* Errors in corrections of pseudorange measurements caused by ionospheric effects (free electrons in the ionosphere).
2. *Troposphere.* Errors in corrections of pseudorange measurements caused by tropospheric effects; temperature, pressure, and humidity contribute to variations in the speed of light.
3. *Multipath.* Errors caused by reflected signals entering the receiver antenna.
4. *Ephemeris.* Ephemeris data errors in transmitted parameters in navigation messages for satellites' true positions.
5. *Satellite Clock.* Clock errors in the transmitted clock data for GNSS.
6. *Receiver Errors.* Errors in the receiver's measurement of range caused by thermal noise, software accuracy, and interchannel biases.

These are described in detail in Sections 7.2–7.8.

7.2 IONOSPHERIC PROPAGATION ERRORS

The ionosphere, which extends from approximately 50–1000 km above the surface of the earth, consists of gases that have been ionized by solar radiation. The ionization produces clouds of free electrons that act as a dispersive medium for GNSS signals in which propagation velocity is a function of frequency. A particular location within the ionosphere is alternately illuminated by the sun and shadowed from the sun by the earth in a daily cycle; consequently, the characteristics of the ionosphere exhibit a diurnal variation in which the ionization is usually maximum late in midafternoon and minimum a few hours after midnight. Additional variations result from changes in solar activity.

The primary effect of the ionosphere on GNSS signals is to change the signal propagation speed as compared to that of free space. A curious fact is that the signal modulation (the code and data stream) is delayed, while the carrier phase is advanced by the same amount. Thus, the measured pseudorange using the code is larger than the correct value, while that using the carrier phase is equally smaller. The magnitude of either error is directly proportional to the total electron content (TEC) in a tube of 1 m² cross section along the propagation path. The TEC varies spatially due to spatial nonhomogeneity of the ionosphere. Temporal variations are caused not only by ionospheric dynamics but also by rapid changes in the propagation path due to satellite motion. The path delay for a satellite at zenith typically varies from about 1 m at night to 5–15 m during late afternoon. At low elevation angles, the propagation path through the ionosphere is much longer, so the typical corresponding delays can increase to several meters at night and as much as 50 m during the day.

Since ionospheric error is usually greater at low elevation angles, the impact of these errors could be reduced by not using measurements from satellites below a certain elevation mask angle. However, in difficult signal environments, including blockage of some satellites by obstacles, the user may be forced to use low-elevation satellites. Mask angles of 5°–10° offer a good compromise between the loss of measurements and the likelihood of large ionospheric errors.

The L1-only receivers in nondifferential operation can reduce ionospheric pseudorange error by using a model of the ionosphere broadcast by the satellites, which reduces the uncompensated ionospheric delay by about 70% on the average. During the day, errors as large as 10 m at midlatitudes can still exist after compensation with this model and can be much worse with increased solar activity. Other recently developed models offer somewhat better performance. However, they still do not handle adequately the daily variability of the TEC, which can depart from the modeled value by 25% or more.

The L1L2 receivers in nondifferential operation can take advantage of the dependence of delay on frequency to remove most of the ionospheric error. A relatively simple analysis shows that the group delay varies inversely as the square of the carrier frequency. This can be seen from the following model of the code pseudorange measurements at the L1 and L2 frequencies:

$$\rho_i = \rho \pm \frac{k}{f_i^2}, \quad (7.1)$$

where ρ is the error-free pseudorange, ρ_i is the measured pseudorange, and k is a constant that depends on the TEC along the propagation path. The subscript $i = 1, 2$ identifies the measurement at the L1 or L2 frequencies, respectively, and the plus or minus sign is identified with respective code or carrier phase measurements, respectively. The two equations can be solved for both ρ and k . The solution for ρ for ionosphere free code pseudorange measurements is

$$\rho = \frac{f_1^2}{f_1^2 - f_2^2} \rho_1 - \frac{f_2^2}{f_1^2 - f_2^2} \rho_2, \quad (7.2)$$

where f_1 and f_2 are the L1 and L2 carrier frequencies, respectively, and ρ_1 and ρ_2 are the corresponding pseudorange measurements.

An equation similar to Eq. 7.2 can be obtained for carrier phase measurements. However, in a nondifferential operation, the residual carrier phase error can be greater than either an L1 or L2 carrier wavelength, making ambiguity resolution difficult.

With a differential operation, ionospheric errors can be nearly eliminated in many applications because ionospheric errors tend to be highly correlated when the base and roving stations are in sufficiently close proximity. With two L1-only receivers separated by 25 km, the unmodeled differential ionospheric error is typically at the 10- to 20-cm level. At a 100-km separation, this can increase to as much as a meter. Additional error reduction using an ionospheric model can further reduce these errors by 25–50%.

7.2.1 Ionospheric Delay Model

J. A. Klobuchar's model [1, 2] for vertical ionospheric delay in seconds is given by

$$T_g = \text{DC} + A \left[1 - \frac{x^2}{2} + \frac{x^4}{24} \right] \text{ for } |x| \leq \frac{\pi}{2}, (\text{sec}) \quad (7.3)$$

where

$$x = \frac{2\pi(t - T_p)}{P}, (\text{rad})$$

DC = 5 ns (constant offset)

T_p = phase

= 50,400 s

A = amplitude

P = period

t = local time of the earth subpoint of the signal intersection with mean ionospheric height(s)

The algorithm assumes this latter height to be 350 km. The DC and phasing T_p are held constant at 5 ns and 14 h (50,400 s) local time.

Amplitude (A) and period (P) are modeled as third-order polynomials:

$$A = \sum_{n=0}^3 \alpha_n \phi_m^n(s) \text{ for } A \geq 0, \text{ otherwise } = 0$$

$$P = \sum_{n=0}^3 \beta_n \phi_m^n(s) \text{ for } P > 72,000 \text{ s, otherwise } P = 72,000 \text{ s}$$

where ϕ_m is the geomagnetic latitude of the ionospheric subpoint and α_n, β_n are coefficients selected (from 370 such sets of constants) by the GPS master control station and placed in the satellite navigation upload message for downlink to the user.

A typical value of coefficients is

$$\alpha_n = [0.8382 \times 10^{-8}, -0.745 \times 10^{-8}, -0.596 \times 10^{-7}, -0.596 \times 10^{-7}],$$

$$\beta_n = [0.8806 \times 10^{-5}, -0.3277 \times 10^{-5}, -0.1966 \times 10^{-6}, -0.1966 \times 10^{-6}].$$

The parameter ϕ_m is calculated as follows with example values:

1. Subtended earth angle (EA) between user and satellite is given by the approximation

$$\text{EA} = \left(\frac{445}{\text{el} + 20} \right)^4 (\text{deg}),$$

where el is the elevation of the satellite and, with respect to the user, equals 15.5°.

2. Geodetic latitude (lat) and longitude (long) of the ionospheric subpoint are found using the approximations

$$\text{Iono lat } \phi_1 = \phi_{\text{user}} + \text{EA} \cos \text{AZ} (\text{deg}),$$

$$\text{Iono long } \lambda_1 = \lambda_{\text{user}} + \frac{\text{EA} \sin \text{AZ}}{\cos \phi_1} (\text{deg}),$$

$$t = 4.32 \times 10^{-4} \lambda_1 + \text{GPS time (sec)}$$

where ϕ_{user} is geodetic latitude = 41°, λ_{user} is geodetic longitude = -73°, and AZ is azimuth of the satellite with respect to the user = 112.5°.

3. The geodetic latitude is converted to a geomagnetic coordinate system using the approximation

$$\phi_m = \phi_1 + 11.6^\circ (\lambda_1 - 291^\circ) (\text{deg}).$$

4. The final step in the algorithm is to account for elevation angle effect by scaling with an obliquity scale factor (SF):

$$SF = 1 + 2 \left[\frac{96^\circ - el}{90^\circ} \right]^3 \text{ (unitless).}$$

With scaling, time delay due to ionospheric becomes

$$T_g = \begin{cases} SF \left[(DC) + A \left(1 - \frac{x^2}{2} + \frac{x^4}{24} \right) \right] & |x| < \frac{\pi}{2}, \\ SF(DC), & |x| > \frac{\pi}{2}, \end{cases}$$

$$T_G = CT_g$$

C = speed of light

$$t = \frac{\lambda_f}{15} + UTC,$$

where T_g is in seconds and T_G is in meters.

The MATLAB® programs Klobuchar fix.m and Klobuchar pseudorandom noise (PRN) for computing ionospheric delay (for PRN = satellite number) are described in Appendix A.

7.2.2 GNSS SBAS Ionospheric Algorithms

The ionospheric correction computation (ICC) algorithms enable the computation of the ionospheric delays applicable to a signal on L1 and to the GPS and wide-area reference station (WRS) L1 and L2 interfrequency biases. These algorithms also calculate grid ionospheric vertical errors (GIVEs), empirically derived error bounds for the broadcast ionospheric corrections. The ionospheric delays are employed by the space-based augmentation system (SBAS) user to correct the L1 measurements, as well as internally to correct the WRSs' L1 geostationary earth orbit (GEO) measurement for orbit determination if dual-frequency corrections are not available from GEOs. The interfrequency biases are needed internally to convert the dual-frequency-derived SBAS corrections to single-frequency corrections for the SBAS users. The vertical ionospheric delay and GIVE information is broadcast to the SBAS user via message types 18 and 26. See the Minimum Operational Performance Standards (MOPs) for details on the content and usage of the SBAS messages [3].

The algorithms used to compute ionospheric delays and interfrequency biases are based on those originated at the Jet Propulsion Laboratory [4]. The ICC models assume that ionospheric electron density is concentrated on a thin shell of height 350 km above the mean earth surface. The estimates of interfrequency biases and ionospheric delays are derived using a pair of

Kalman filters, herein referred to as the L1L2 and ionosphere (IONO) filters. The purpose of the L1L2 filter is to estimate the interfrequency biases, while the purpose of the IONO filter is to estimate the ionosphere delays. The inputs to both filters are leveled WRS receiver slant delay measurements (L2 minus L1 differential delay), which are output from the data. Both filters perform their calculations in total electron count units (TECU) (1 m of L1 ranging delay = 6.16 TECU, and 1 m of L1 – L2 differential delay = 9.52 TECU). Conceptually, the measurement equation is (neglecting the noise term)

$$\tau_{TECU} = 9.52 \times \tau_m \quad (7.4)$$

$$= 9.52 \times (t_{L2,m} - t_{L1,m}) \quad (7.5)$$

$$= 9.52 \times (b_m^r + b_m^s) + TEC_{TECU} \quad (7.6)$$

$$= b_{TECU}^r + b_{TECU}^s + TEC_{TECU}, \quad (7.7)$$

where τ is the differential delay, b^r and b^s are the interfrequency biases of the respective receiver and satellite, and TEC is the ionospheric delay. The subscripts m (meters) and TECU denote the corresponding units of each term. The ionospheric delay in meters for a signal on the L1 frequency is

$$\tau_m^{L1} = 1.5457 \times \frac{1}{9.52} TEC_{TECU} \quad (7.8)$$

$$= \frac{1}{6.16} TEC_{TECU}. \quad (7.9)$$

Both Kalman filters contain the vertical delays at the vertices of a triangular spherical grid of height 350 km fixed in the solar-magnetic (SM) coordinate frame as states. The L1L2 filter also contains interfrequency biases as states. In contrast, the IONO filter does not estimate the interfrequency biases, but instead they are periodically forwarded to the IONO filter, along with the variances of the estimates, from the L1L2 filter. Each slant measurement is modeled as a linear combination of the vertical delays at the three vertices surrounding the corresponding measurement pierce point (the intersection of the line of sight and the spherical grid), plus the sum of the receiver and satellite biases, plus noise. The ionospheric delays computed in the IONO filter are eventually transformed to a latitude–longitude grid that is sent to the SBAS users via message type 26. Because SBAS does not have any calibrated ground receivers, the interfrequency bias estimates are all relative to a single receiver designated as a reference, whose L1L2 interfrequency bias filter covariance is initialized to a small value, and to which no process noise is applied.

The major algorithms making up the ICC discussed here are

Initialization. The L1L2 and IONO filters are initialized using either the Klobuchar model or using previously recorded data.

the logic associated with restarting the estimation using recorded data must be analyzed. The capabilities to perform thread switches and to detect anomalies must be examined, and the special cases necessitating a change of reference receiver. In each phase of validation, the critical test is whether there is any significant degradation in accuracy as compared to nominal performance, and whether the nominal performance itself is adequate.

7.2.2.4 Calculation of Ionospheric Delay Using Pseudoranges The calculation of ionospheric propagation delay from P-code and C/A-code can be formulated in terms of the following measurement equalities:

$$\rho_{RL1} = \rho + L1_{iono} - c\tau_{RX1} - c\tau_{GD}, \quad (7.10)$$

$$\rho_{RL2} = \rho + \frac{L1_{iono}}{(f_{L2}/f_{L1})^2} - c\tau_{RX2} - \frac{c\tau_{GD}}{(f_{L2}/f_{L1})^2}, \quad (7.11)$$

where

$$\left. \begin{aligned} \rho_{RL1} &= \text{L1 pseudorange} \\ \rho_{RL2} &= \text{L2 pseudorange} \\ \rho &= \text{geometric distance between GPS satellite} \\ &\quad \text{transmitter and GPS receiver, including} \\ &\quad \text{nondispersive contributions such as} \\ &\quad \text{tropospheric refraction and user clock error} \\ f_{L1} &= \text{L1 frequency} \\ &= 1572.42 \text{ MHz (GPS)} \\ f_{L2} &= \text{L2 frequency} \\ &= 1227.6 \text{ MHz (GPS)} \\ \tau_{RX1} &= \text{receiver noise as manifested in code} \\ &\quad \text{(receiver and calibration biases) at L1 (ns)} \\ \tau_{RX2} &= \text{receiver noise as manifested in code} \\ &\quad \text{(receiver and calibration biases) at L2 (ns)} \\ \tau_{GD} &= \text{satellite group delay (interfrequency bias)} \\ c &= \text{speed of light} \\ &= 0.299792458 \frac{\text{m}}{\text{ns}} \\ L1_{iono} &= \text{delay at L1 (m)} \end{aligned} \right\} \quad (7.12)$$

Subtracting Eq. 7.10 from Eq. 7.11, we get

$$L1_{iono} = \frac{\rho_{RL1} - \rho_{RL2}}{1 - (f_{L1}/f_{L2})^2} - \frac{c(\tau_{RX1} - \tau_{RX2})}{1 - (f_{L1}/f_{L2})^2} - c\tau_{GD}. \quad (7.13)$$

What is actually measured in the ionospheric delay is the sum of receiver bias and interfrequency bias. The biases are determined and taken out from the ionospheric delay calculation. These biases may be up to 10 ns (3 m) [6, 7].

However, the presence of ambiguities N_1 and N_2 in carrier phase measurements of L1 and L2 preclude the possibility of calculating the ionosphere delay directly and would involve the solution of the carrier cycle ambiguities iteratively.

The MATLAB® program Iono_delay(PRN#) (described in Appendix A) uses pseudorange and carrier phase data from L1 and L2 signals.

7.3 TROPOSPHERIC PROPAGATION ERRORS

The lower part of the earth's atmosphere is composed of dry gases and water vapor, which lengthen the propagation path due to refraction. The magnitude of the resulting signal delay depends on the refractive index of the air along the propagation path and typically varies from about 2.5 m in the zenith direction to 10–15 m at low satellite elevation angles. The troposphere is nondispersive at the GNSS frequencies, so that delay is not frequency dependent. In contrast to the ionosphere, tropospheric path delay is consequently the same for code and carrier signal components. Therefore, this delay cannot be measured by utilizing both L1 and L2 pseudorange measurements, and either models and/or differential techniques must be used to reduce the error.

The refractive index of the troposphere consists of that due to the dry-gas component and the water vapor component, which respectively contribute about 90% and 10% of the total delay. Knowledge of the temperature, pressure, and humidity along the propagation path can determine the refractivity profile, but such measurements are seldom available to the user. However, using standard atmospheric models for the dry delay permits determination of the zenith delay to within about 0.5 m and with an error at other elevation angles that approximately equals the zenith error times the cosecant of the elevation angle. These standard atmospheric models are based on the laws of ideal gases and assume spherical layers of constant refractivity with no temporal variation and an effective atmospheric height of about 40 km. Estimation of dry delay can be improved considerably if surface pressure and temperature measurements are available, bringing the residual error down to within 2–5% of the total.

The component of tropospheric delay due to water vapor (at altitudes up to about 12 km) is much more difficult to model because there is considerable spatial and temporal variation of water vapor in the atmosphere. Fortunately, the wet delay is only about 10% of the total, with values of 5–30 cm in continental midlatitudes. Despite its variability, an exponential vertical profile model can reduce it to within about 2–5 cm.

In practice, a model of the standard atmosphere at the antenna location would be used to estimate the combined zenith delay due to both wet and dry components. Such models use inputs such as the day of the year and the

latitude and altitude of the user. The delay is modeled as the zenith delay multiplied by a factor that is a function of the satellite elevation angle. At zenith, this factor is unity, and it increases with decreasing elevation angle as the length of the propagation path through the troposphere increases. Typical values of the multiplication factor are 2 at 30° elevation angle, 4 at 15°, 6 at 10°, and 10 at 5°. The accuracy of the model decreases at low elevation angles, with decimeter level errors at zenith and about 1 m at 10° elevation.

Much research has gone into the development and testing of various tropospheric models. Excellent summaries of these appear in the literature [8–10].

Although a GNSS receiver cannot measure pseudorange error due to the troposphere, differential operation can usually reduce the error to small values by taking advantage of the high spatial correlation of tropospheric errors at two points within 0–100 km on the earth's surface. However, exceptions often occur when storm fronts pass between the receivers, causing large gradients in temperature, pressure, and humidity.

7.4 THE MULTIPATH PROBLEM

Multipath propagation of the GNSS signal is a dominant source of error in positioning, especially in differential GNSS architectures. Objects in the vicinity of a receiver antenna (notably the ground) can easily reflect GNSS signals, resulting in one or more secondary propagation paths. These secondary-path signals, which are superimposed on the desired direct-path signal, always have a longer propagation time and can significantly distort the amplitude and phase of the direct-path signal.

Errors due to multipath cannot be reduced by the use of differential GNSS since they depend on local reflection geometry near each receiver antenna. In a receiver without multipath protection, C/A-code ranging errors of 10 m or more can be experienced. Multipath cannot only cause large code ranging errors but can also severely degrade the ambiguity resolution process required for carrier phase ranging such as that used in precision surveying applications.

Multipath propagation can be divided into two classes: static and dynamic. For a stationary receiver, the propagation geometry changes slowly as the satellites move across the sky, making the multipath parameters essentially constant for perhaps several minutes. However, in mobile applications, there can be rapid fluctuations in fractions of a second. Therefore, different multipath mitigation techniques are generally employed for these two types of multipath environments. Most current research has been focused on static applications, such as surveying, where greater demand for high accuracy exists. For this reason, we will concentrate our attention to the static case.

7.4.1 How Multipath Causes Ranging Errors

To facilitate an understanding of how multipath causes ranging errors, several simplifications can be made that in no way obscure the fundamentals involved.

We will assume that the receiver processes only the C/A-code and that the received signal has been converted to complex (i.e., analytic) form at baseband (nominally zero frequency), where all Doppler shift has been removed by a carrier tracking phase-lock loop. It is also assumed that the GNSS navigation data modulation has been removed from the signal, which can be achieved by standard techniques. When no multipath is present, the received waveform is represented by

$$r(t) = ae^{j\phi}c(t-\tau) + n(t), \quad (7.14)$$

where $c(t)$ is the normalized, undelayed C/A-code waveform as transmitted; r is the signal propagation delay; a is the signal amplitude; ϕ is the carrier phase; and $n(t)$ is the Gaussian receiver thermal noise having flat power spectral density. Pseudoranging consists of estimating the delay parameter τ . As we have previously seen, an optimal estimate (i.e., a minimum-variance unbiased estimate) of τ can be obtained by forming the cross-correlation function

$$R(\tau) = \int_{t_1}^{t_2} r(t)c_r(t-\tau)dt, \quad (7.15)$$

of $r(t)$ with a replica $c_r(t)$ of the transmitted C/A-code and choosing as the delay estimate that value of τ that maximizes this function. Except for an error due to receiver thermal noise, this occurs when the received and replica waveforms are in time alignment. A typical cross-correlation function without multipath for C/A-code receivers having a 2-MHz precorrelation bandwidth is shown by the solid lines Fig. 7.2 (these plots ignore the effect of noise, which would add small random variations to the curves).

If multipath is present with a single secondary path, the waveform of Eq. 7.14 changes to

$$r(t) = ae^{j\phi_1}c(t-\tau_1) + be^{j\phi_2}c(t-\tau_2) + n(t), \quad (7.16)$$

where the direct and secondary paths have respective propagation delays τ_1 and τ_2 , amplitudes a and b , and carrier phases ϕ_1 and ϕ_2 . In a receiver not designed expressly to handle multipath, the resulting cross-correlation function will now have two superimposed components, one from the direct path and one from the secondary path. The result is a function with a distortion depending on the relative amplitude, delay, and phase of the secondary-path signal, as illustrated at the top of Fig. 7.2 for an in-phase secondary path and at the bottom of Fig. 7.2 for an out-of-phase secondary path. Most importantly, the location of the peak of the function has been displaced from its correct position, resulting in a pseudorange error.

In vintage receivers employing standard code tracking techniques (early and late codes separated by one C/A-code chip), the magnitude of pseudorange error caused by multipath can be quite large, reaching 70–80 m for a secondary-path signal one-half as large as the direct-path signal and having a relative delay of approximately 250 m. Further details can be found in Ref. 11.

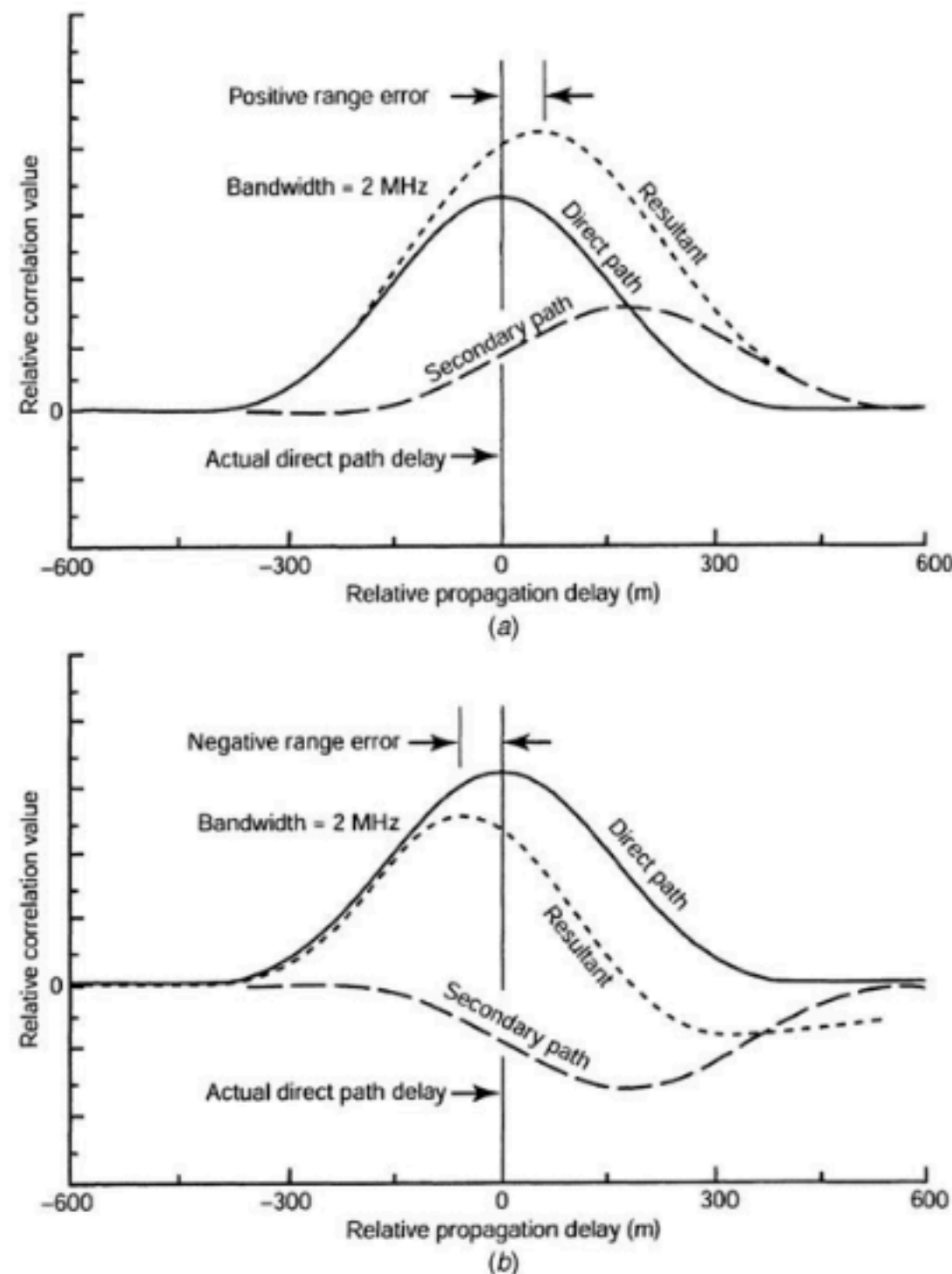


Fig. 7.2 Effect of multipath on C/A-code cross-correlation function.

7.5 METHODS OF MULTIPATH MITIGATION

Processing against slowly changing multipath can be broadly separated into two classes: spatial processing and time-domain processing. Spatial processing uses antenna design in combination with known or partially known characteristics of signal propagation geometry to isolate the direct-path received signal. In contrast, time domain processing achieves the same result by operating only on the multipath-corrupted signal within the receiver.

7.5.1 Spatial Processing Techniques

7.5.1.1 Antenna Location Strategy Perhaps the simplest form of spatial processing is to locate the antenna where it is less likely to receive reflected signals. For example, to obtain the position of a point near reflective objects, one can first use GNSS to determine the position of a nearby point “in the clear” and then calculate the relative position of the desired point by simple distance and/or angle measurement techniques. Another technique that minimizes ever-present ground signal reflections is to place the receiver antenna directly at ground level. This causes the point of ground reflection to be essentially coincident with the antenna location so that the secondary path has very nearly the same delay as the direct path. Clearly, such antenna location strategies may not always be possible but can be very effective when feasible.

7.5.1.2 Ground Plane Antennas The most common form of spatial processing is an antenna designed to attenuate signals reflected from the ground. A simple design uses a metallic ground plane disk centered at the base of the antenna to shield the antenna from below. A deficiency of this design is that when the signal wave fronts arrive at the disk edge from below, they induce surface waves on the top of the disk that then travel to the antenna. The surface waves can be eliminated by replacing the ground plane with a choke ring, which is essentially a ground plane containing a series of concentric circular troughs one-quarter wavelength deep. These troughs act as transmission lines shorted at the bottom ends so that their top ends exhibit very high impedance at the GNSS carrier frequency. Therefore, induced surface waves cannot form, and signals that arrive from below the horizontal plane are significantly attenuated. However, the size, weight, and cost of a choke ring antenna are significantly greater than those of simpler designs. Most importantly, the choke ring cannot effectively attenuate secondary-path signals arriving from above the horizontal, such as those reflecting from buildings or other structures. Nevertheless, such antennas have proven to be effective when signal ground bounce is the dominant source of multipath, particularly in GNSS surveying applications.

7.5.1.3 Directive Antenna Arrays A more advanced form of spatial processing uses antenna arrays to form a highly directive spatial response pattern with high gain in the direction of the direct-path signal and attenuation in directions from which secondary-path signals arrive. However, inasmuch as signals from different satellites have different directions of arrival and different multipath geometries, many directivity patterns must be simultaneously operative, and each must be capable of adapting to changing geometry as the satellites move across the sky. For these reasons, highly directive arrays seldom are practical or affordable for most applications.

7.5.1.4 Long-Term Signal Observation If a GNSS signal is observed for sizable fractions of an hour to several hours, one can take advantage of changes

in multipath geometry caused by satellite motion. This motion causes the relative delays between the direct and secondary paths to change, resulting in measurable variations in the received signal. For example, a periodic change in signal level caused by alternate phase reinforcement and cancellation by the reflected signals is often observable. Although a variety of algorithms have been proposed for extracting the direct-path signal component from measurements of the received signal, the need for long observation times rules out this technique for most applications. However, it can be an effective method of multipath mitigation at a fixed site, such as at a differential GNSS base station. In this case, it is even possible to observe the same satellites from one day to the next, looking for patterns of pseudorange or phase measurements that repeat daily.

Multipath Calculation from Long-Term Observations Delays can be computed as follows by using pseudoranges and carrier phases over long signal observations (one day to next). This technique may be ruled out for most applications. Ambiguities and cycle slips have been eliminated or mitigated. Let

$$\left. \begin{array}{l} \lambda_1 = 19.03 \text{ cm, wavelength of L1 (GPS)} \\ \lambda_2 = 24.42 \text{ cm, wavelength of L2 (GPS)} \\ \phi_{L1} = \text{carrier phase for L1 (cycles)} \\ \phi_{L2} = \text{carrier phase for L2 (cycles)} \\ f_{L1} = \text{L1 frequency} = 1575.42 \text{ MHz (GPS)} \\ f_{L2} = \text{L2 frequency} = 1227.6 \text{ MHz (GPS)} \\ \rho = \text{error free pseudorange} \\ \rho_{RL1} = \text{pseudorange L1 (m)} \\ \rho_{RL2} = \text{pseudorange L2 (m)} \\ I = \text{ionospheric delay} \\ I_{L1} = \text{ionospheric delay in L1} \\ MP_{L1} = \text{multipath in L1} \\ MP_{L2} = \text{multipath in L2} \end{array} \right\}; \quad (7.17)$$

for dual-frequency GNSS receivers, one obtains

$$\lambda_1 \phi_{L1} = \rho - \frac{I}{(f_{L1})^2}, \quad (7.18)$$

$$\lambda_2 \phi_{L2} = \rho - \frac{I}{(f_{L2})^2}. \quad (7.19)$$

Subtracting Eq. 7.19 from Eq. 7.18, one can obtain

$$\begin{aligned} \lambda_1 \phi_{L1} - \lambda_2 \phi_{L2} &= \frac{I(f_{L1})^2 - I(f_{L2})^2}{(f_{L1})^2 (f_{L2})^2}, \\ I_{L1} &= \frac{(\lambda_1 \phi_{L1} - \lambda_2 \phi_{L2})(f_{L2})^2}{(f_{L1})^2 - (f_{L2})^2}, \end{aligned} \quad (7.20)$$

$$\begin{aligned} K &= \frac{(f_{L2})^2}{(f_{L1})^2 - (f_{L2})^2}, \\ I_{L1} &= K(\lambda_1 \phi_{L1} - \lambda_2 \phi_{L2}), \\ \rho_{RL1} &= \rho + \frac{I}{(f_{L1})^2}. \end{aligned} \quad (7.21)$$

Subtracting Eq. 7.18 from Eq. 7.21, one obtains the multipath as

$$MP_{L1} = \rho_{RL1} - \lambda_1 \phi_{L1} - 2I_{L1}, \quad (7.22)$$

where

$$I_{L1} = \frac{I}{(f_{L1})^2}.$$

Substitute Eq. 7.20 into Eq. 7.22 to obtain

$$\begin{aligned} MP_{L1} &= \rho_{RL1} - \lambda_1 \phi_{L1} - 2K(\lambda_1 \phi_{L1} - \lambda_2 \phi_{L2}) \\ &= \rho_{RL1} - [(1 - 2K)\lambda_1 \phi_{L1} + 2K\lambda_2 \phi_{L2}]. \end{aligned} \quad (7.23)$$

7.5.2 Time-Domain Processing

Although time-domain processing against GNSS multipath errors has been the subject of active research for at least two decades, there is still much to be learned, both at theoretical and practical levels. Most of the practical approaches have been developed by receiver manufacturers, who are often reluctant to explicitly reveal their methods. Nevertheless, enough information about multipath processing exists to gain insight into its recent evolution.

7.5.2.1 Narrow-Correlator Technology (1990–1993) The first significant means to reduce GPS multipath effects by receiver processing made its debut in the early 1990s. Until that time, most receivers had been designed with a 2-MHz precorrelation bandwidth that encompassed most, but not all, of the GPS C/A spread-spectrum signal power. These receivers also used one-chip spacing between the early and late reference C/A-codes in the code tracking loops. However, the 1992 paper [12] makes it clear that using a significantly larger bandwidth combined with much closer spacing of the early and late reference codes would dramatically improve the ranging accuracy both with

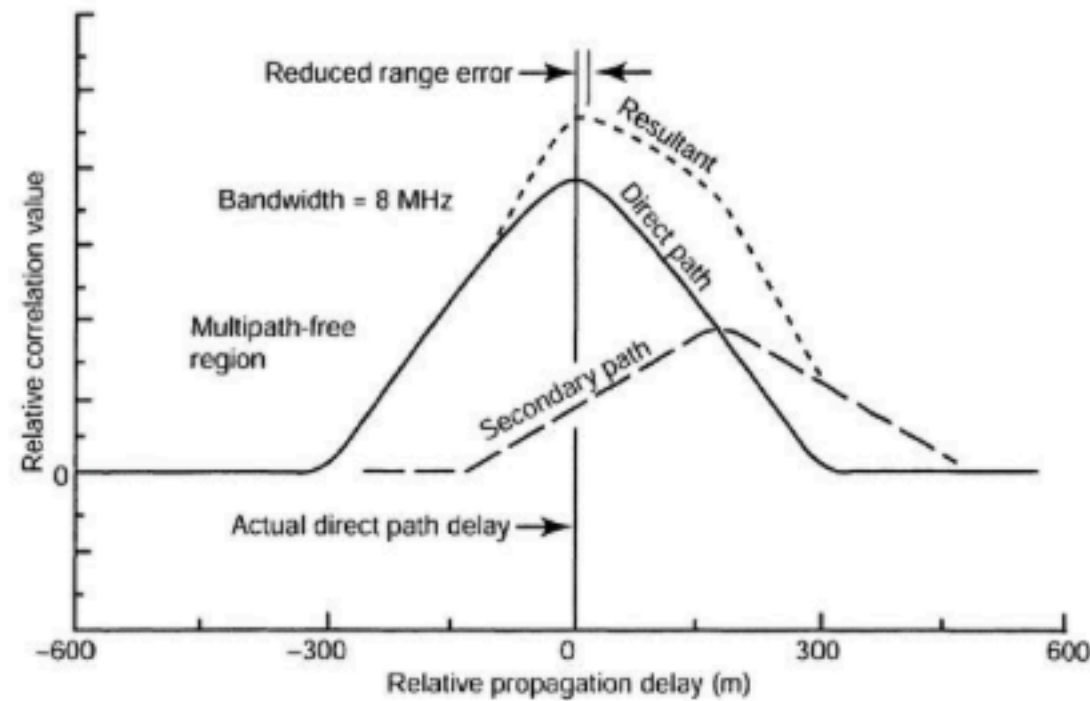


Fig. 7.3 Reduced multipath error with larger precorrelation bandwidth.

and without multipath. It is somewhat surprising that these facts were not recognized earlier by the GNSS community, given that they had been well known in radar circles for many decades.

A 2-MHz precorrelation bandwidth causes the peak of the direct-path cross-correlation function to be severely rounded, as illustrated in Fig. 7.2. Consequently, the sloping sides of a secondary-path component of the correlation function can significantly shift the location of the peak, as indicated in the figure. The result of using an 8-MHz bandwidth is shown in Fig. 7.3, where it can be noted that the sharper peak of the direct-path cross-correlation function is less easily shifted by the secondary-path component. It can also be shown that at larger bandwidths, the sharper peak is more resistant to disturbance by receiver thermal noise, even though the precorrelation signal-to-noise ratio is increased.

Another advantage of a larger precorrelation bandwidth is that the spacing between the early and late reference codes in a code tracking loop can be made smaller without significantly reducing the gain of the loop, hence the term narrow correlator. It can be shown that this causes the noises on the early and late correlator outputs to become more highly correlated, resulting in less noise on the loop error signal. An additional benefit is that the code tracking loop will be affected only by the multipath-induced distortions near the peak of the correlation function.

7.5.2.2 Leading-Edge Techniques Because the direct-path signal always precedes secondary-path signals, the leading (left-hand) portion of the correla-

tion function is uncontaminated by multipath, as is illustrated in Fig. 7.3. Therefore, if one could measure the location of just the leading part, it appears that the direct-path delay could be determined with no error due to multipath. Unfortunately, this seemingly happy state of affairs is illusory. With a small direct-/secondary-path separation, the uncontaminated portion of the correlation function is a minuscule piece at the extreme left, where the curve just begins to rise. In this region, not only is the signal-to-noise ratio relatively poor, but the slope of the curve is also relatively small, which severely degrades the accuracy of delay estimation.

For these reasons, the leading-edge approach best suits situations with a moderate to large direct-/secondary-path separation. However, even in these cases, there is the problem of making the delay measurement insensitive to the slope of the correlation function leading edge, which can vary with signal strength. Such a problem does not occur when measuring the location of the correlation function peak.

7.5.2.3 Correlation Function Shape-Based Methods Some GNSS receiver designers have attempted to determine the parameters of the multipath model from the shape of the correlation function. The idea has merit, but for best results, many correlations with different values of reference code delay are required to obtain a sampled version of the function shape. Another practical difficulty arises in attempting to map each measured shape into a corresponding direct-path delay estimate. Even in the simple two-path model (Eq. 7.14) there are six signal parameters, so that a very large number of correlation function shapes must be handled. An example of a heuristically developed shape-based approach called the early-late slope (ELS) method can be found in Ref. 13, while a method based on maximum-likelihood estimation (MLE) called the multipath-estimating delay-lock loop (MEDLL) is described in Ref. 14.

7.5.2.4 Modified Correlator Reference Waveforms Another new approach to multipath mitigation alters the waveform of the correlator reference PRN code to provide a cross-correlation function with inherent resistance to errors caused by multipath. Examples include the strobe correlator [15], the use of special code reference waveforms to narrow the correlation function developed in Refs. 16 and 17, and the gated correlator developed in Ref. 18. These techniques take advantage of the fact that the range information in the received signal resides primarily in the chip transitions of the C/A-code. By using a correlator reference waveform that is not responsive to the flat portions of the C/A-code, the resulting correlation function can be narrowed down to the width of a chip transition, thereby being almost immune to multipath having a primary/secondary-path separation greater than 30–40 m. An example of such a reference waveform and the corresponding correlation function are shown in Fig. 7.4.

methods and its ability to reach theoretical performance limits in many cases. Knowledge of such limits provides a valuable benchmark in receiver design by permitting an accurate assessment of the potential payoff in developing techniques that are better than those in current use. Of equal importance is the revelation of the signal processing operations that can reach performance bounds. Although it may not be feasible to implement the processing directly, its revelation often leads to a practical method that achieves nearly the same performance.

7.6.1.1 Optimality Criteria In discussing theoretical performance limits, it is important to define the criterion of optimality. In GPS, the optimal range estimator is traditionally considered to be the minimum-variance unbiased estimator (MVUE), which can be realized by properly designed receivers. However, in Ref. 20, it is shown that the standard deviation of a MVUE designed for multipath becomes infinite as the primary-to-secondary-path separation approaches zero. For this reason, it seems that a better criterion of optimality would be the minimum root mean square RMS error, which can include both random and bias components. Unfortunately, it can be shown that no estimator exists having minimum RMS error for every combination of true multipath parameters.

7.6.2 Minimum Mean-Squared Error (MMSE) Estimator

There is an estimator that can be claimed optimal in a weaker sense. The MMSE estimator has the property that no other estimator has a uniformly smaller RMS error. In other words, if some other estimator has smaller RMS error than the MMSE estimator for some set of true multipath parameter values, then that estimator must have a larger RMS error than the MMSE estimator for some other set of values.

The MMSE estimator also has an important advantage not possessed by most current multipath mitigation methods in that the RMS error decreases as the length of the signal observation interval is increased.

7.6.3 Multipath Modeling Errors

Although a properly designed estimation-theoretic approach such as the MMSE estimator will generally outperform other methods, the design of such estimators requires a mathematical model of the multipath-contaminated signal containing parameters to be estimated. If the actual signal departs from the assumed model, performance degradation can occur. For example, if the model contains only two signal propagation paths but in reality the signal is arriving via three or more paths, large bias errors in range estimation can result. On the other hand, poorer performance (usually in the form of random error cause by noise) can also occur if the model has too many degrees of freedom. Striking the right balance in the number of parameters in the model

can be difficult if little information exists about the multipath reflection geometry.

7.7 EPHEMERIS DATA ERRORS

Small errors in the ephemeris data transmitted by each satellite cause corresponding errors in the computed position of the satellite (here we exclude the ephemeris error component of SA, which is regarded as a separate error source). Satellite ephemerides are determined by the master control station of the GNSS ground segment based on monitoring of individual signals by four monitoring stations. Because the locations of these stations are known precisely, an “inverted” positioning process can calculate the orbital parameters of the satellites as if they were users. This process is aided by precision clocks at the monitoring stations and by tracking over long periods of time with optimal filter processing. Based on the orbital parameter estimates thus obtained, the master control station uploads the ephemeris data to each satellite, which then transmits the data to users via the navigation data message. Errors in satellite position when calculated from the ephemeris data typically result in range errors on the order of 1–2 m. Improvements in satellite tracking will undoubtedly reduce this error further.

7.8 ONBOARD CLOCK ERRORS

Timing of the signal transmission from each satellite is directly controlled by its own atomic clock without any corrections applied. This time frame is called *space vehicle (SV) time*. A schematic of a rubidium atomic clock is shown in Fig. 7.8. Although the atomic clocks in the satellites are highly accurate, errors can be large enough to require correction. Correction is needed partly because it would be difficult to directly synchronize the clocks closely in all the satellites. Instead, the clocks are allowed some degree of relative drift that is estimated by ground station observations and is used to generate clock correction data in the GNSS navigation message. When SV time is corrected using this data, the result is called *GNSS time*. The time of transmission used in calculating pseudoranges must be GNSS time, which is common to all satellites.

The onboard clock error is typically less than 1 ms and varies slowly. This permits the correction to be specified by a quadratic polynomial in time whose coefficients are transmitted in the navigation message. The correction has the form

$$\Delta t_{sv} = a_{f0} + a_{f1}(t_{sv} - t_{oc}) + a_{f2}(t_{sv} - t_{oc})^2 + \Delta t_r, \quad (7.47)$$

with

$$t_{GNSS} = t_{sv} - \Delta t_{sv}, \quad (7.48)$$

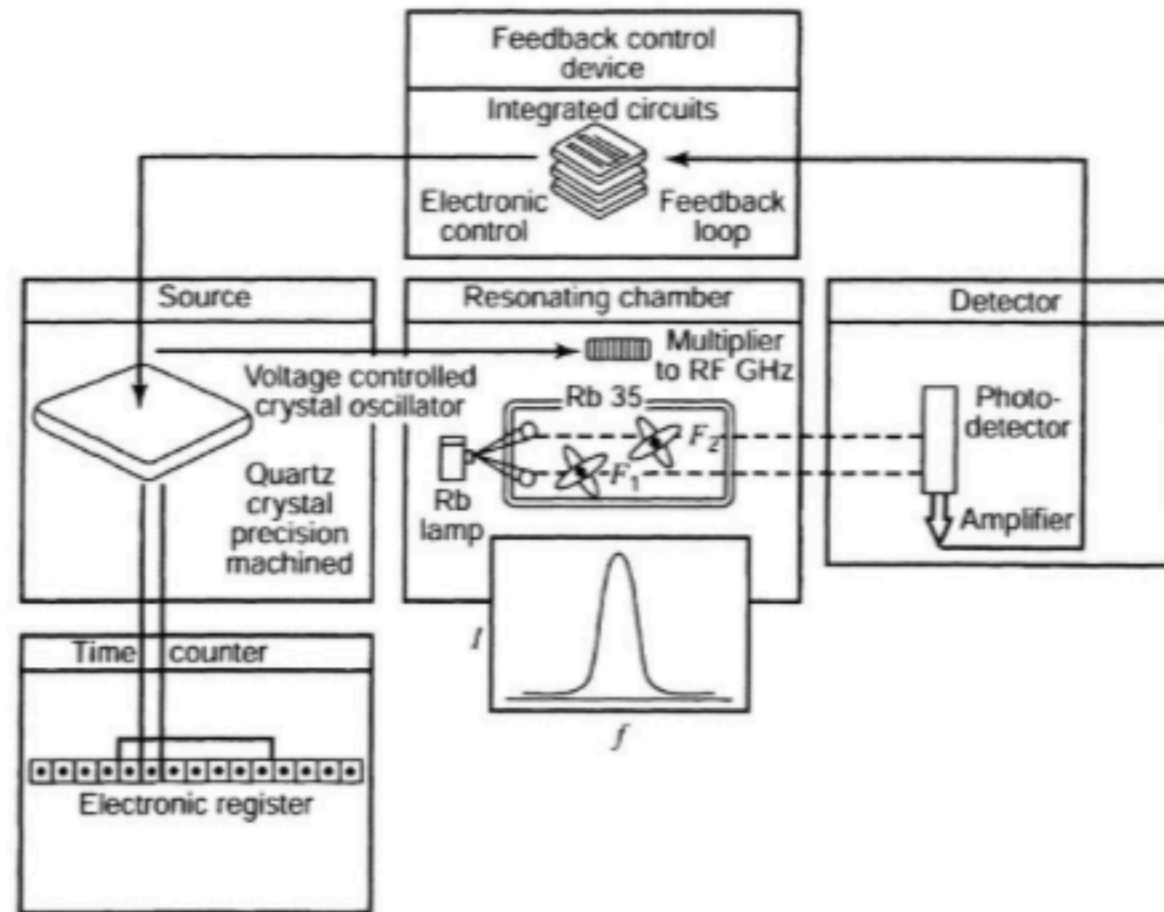


Fig. 7.8 Schematic of a rubidium atomic clock. RF, radio frequency.

where a_{f0} , a_{f1} , a_{f2} are the correction coefficients, t_{sv} is SV time, and Δt_r is a small relativistic clock correction caused by the orbital eccentricity. The clock data reference time t_{oc} , in seconds is broadcast in the navigation data message. The stability of the atomic clocks permits the polynomial correction given by Eq. 7.47 to be valid over a time interval of 4–6 h. After the correction has been applied, the residual error in GNSS time is typically less than a few nanoseconds, or about 1 m in range. Complete calculations of GNSS time are given as exercises in Chapter 4, Section 4.1.3.4.

7.9 RECEIVER CLOCK ERRORS

Because the navigation solution includes a solution for receiver clock error, the requirements for accuracy of receiver clocks is far less stringent than for GNSS satellite clocks. In fact, for receiver clocks, short-term stability over the pseudorange measurement period is usually more important than absolute frequency accuracy. In almost all cases, such clocks are quartz crystal oscillators with absolute accuracies in the 1- to 10-ppm range over typical operating temperature ranges. When properly designed, such oscillators typically have stabilities of 0.01–0.05 ppm over a period of a few seconds.

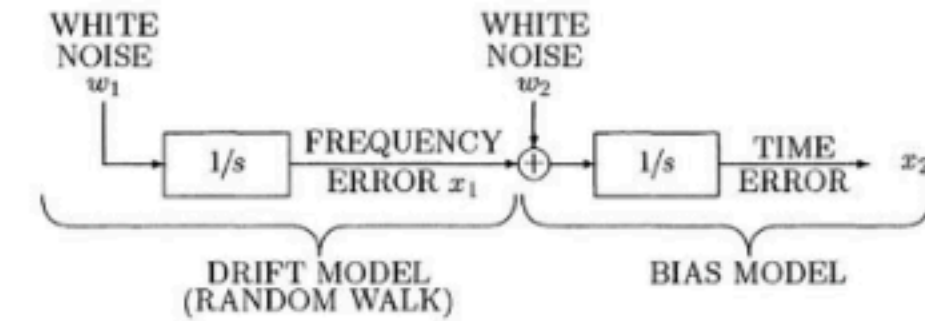


Fig. 7.9 Crystal clock error model.

Receivers that incorporate receiver clock error in the Kalman filter state vector need a suitable mathematical model of the crystal clock error. A typical model in the continuous-time domain is shown in Fig. 7.9, which is easily changed to a discrete version for the Kalman filter. In this model, the clock error consists of a bias (frequency) component and a drift (time) component. The frequency error component is modeled as a random walk produced by integrated white noise. The time error component is modeled as the integral of the frequency error after additional white noise (statistically independent from that causing the frequency error) has been added to the latter. In the model, the key parameters that need to be specified are the power spectral densities of the two noise sources, which depend on characteristics of the specific crystal oscillator used.

The continuous-time model has the form

$$\dot{x}_1 = w_1, \quad (7.49)$$

$$\dot{x}_2 = x_1 + w_2, \quad (7.50)$$

where $w_1(t)$ and $w_2(t)$ are independent zero-mean white-noise processes with known variances.

The equivalent discrete-time model has the state vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad (7.51)$$

and the stochastic sequence model

$$\mathbf{x}_k = \begin{bmatrix} 1 & 0 \\ \Delta t & 1 \end{bmatrix} \mathbf{x}_{k-1} + \begin{bmatrix} w_{1,k-1} \\ w_{2,k-1} \end{bmatrix}, \quad (7.52)$$

where Δt is the discrete-time step and $\{w_{1,k-1}\}$, $\{w_{2,k-1}\}$ are independent zero-mean white-noise sequences with known variances.

7.10 SA ERRORS

Prior to May 1, 2000, SA was a mechanism adopted by the Department of Defense (DOD) to control the achievable navigation accuracy by nonmilitary GPS receivers. In the GPS SPS mode, the SA errors were specified to degrade navigation solution accuracy to 100 m (2D RMS) horizontally and 156 m (RMS) vertically. On May 1, 2000, the President of the United States announced the decision to discontinue this intentional degradation of GPS signals available to the public. The decision to discontinue SA was coupled with continuing efforts to upgrade the military utility of systems using GPS and was supported by threat assessments that concluded that setting SA to zero would have minimal impact on U.S. national security. The decision was part of an ongoing effort to make GPS more responsive to civil and commercial users worldwide.

7.11 ERROR BUDGETS

For purposes of analyzing the effects of the errors discussed above, it is convenient to convert each error into an equivalent range error experienced by a user, which is called the *user-equivalent range error* (UERE). In general, the errors from different sources will have different statistical properties. For example, satellite clock and ephemeris errors tend to vary slowly with time and appear as biases over moderately long time intervals, perhaps hours. On the other hand, errors due to receiver noise and quantization effects may vary much more rapidly, perhaps within seconds. Nonetheless, if sufficiently long time durations over many navigation scenarios are considered, all errors can be considered as zero-mean random processes that can be combined to form a single UERE. This is accomplished by forming the root sum square (RSS) of the UERE errors from all sources:

$$\text{UERE} = \sqrt{\sum_{i=1}^n (\text{UERE}_i)^2}. \quad (7.53)$$

Figure 7.10 depicts the various GPS UERE errors and their combined effect for both C/A-code and P(Y)-code navigation at the $1 - \sigma$ level.

The UERE for the C/A-code user is typically about 19 m. It can be seen that, for such a user, the dominant error sources in nondifferential operations are multipath, receiver noise/resolution, and ionospheric and tropospheric delay (however, recent advances in receiver technology have in some cases significantly reduced receiver noise/resolution errors). On the other hand, the P(Y)-code user has a significantly smaller UERE of about 6 m for the following reasons:

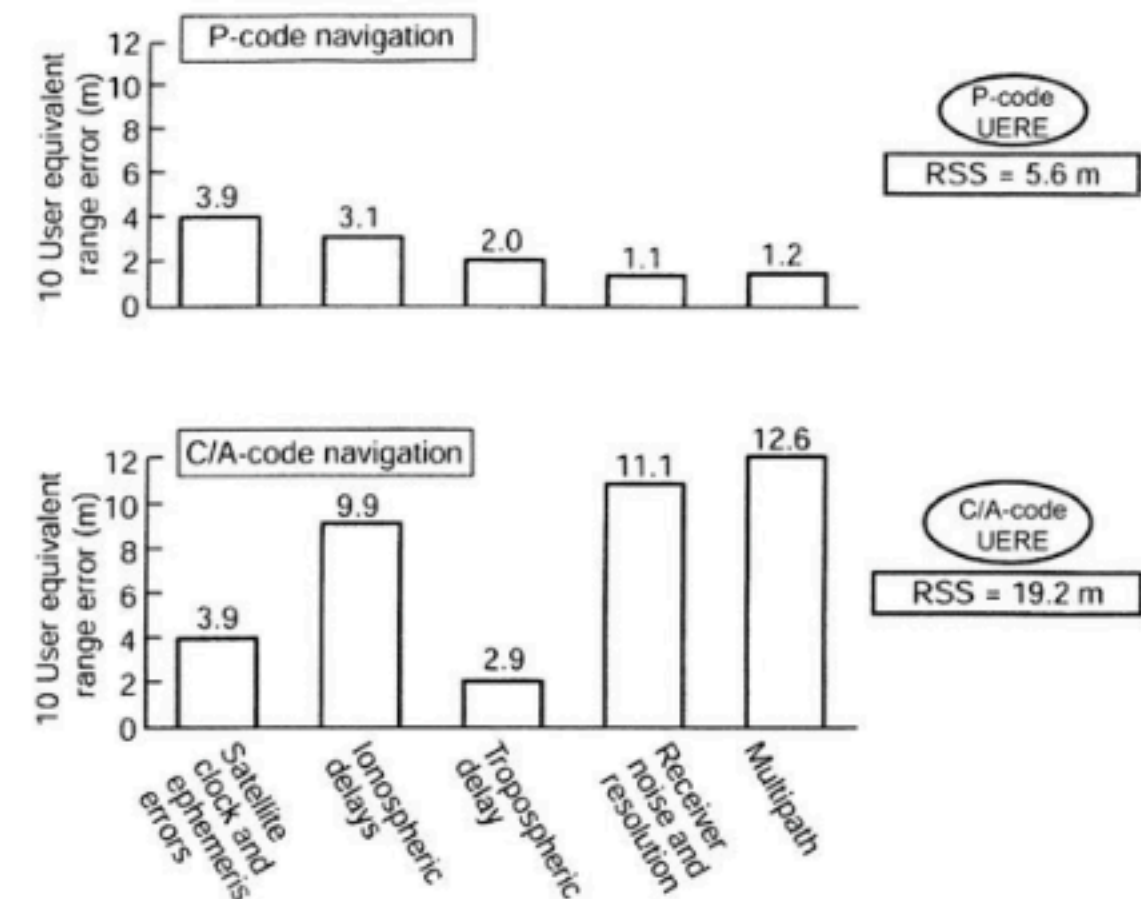


Fig. 7.10 GPS UERE budget.

1. The full use of the L1 and L2 signals permits significant reduction of ionospheric error.
2. The wider bandwidth of the P(Y)-codes greatly reduces errors due to multipath and receiver noise.

PROBLEMS

- 7.1 Using the values provided for Klobuchar's model in Section 7.2.1, calculate the ionospheric delay and plot the results.
- 7.2 Assume that a direct-path GNSS L1 C/A-code signal arrives with a phase such that all of the signal power lies in the baseband I channel, so that the baseband signal is purely real. Further assume an infinite signal bandwidth so that the cross correlation of the baseband signal with an ideal C/A reference code waveform will be an isosceles triangle 600 m wide at the base.
 - (a) Suppose that in addition to the direct-path signal there is a secondary-path signal arriving with a relative time delay of precisely 250 L1 carrier cycles (so that it is in phase with the direct-path signal) and with an amplitude one-half that of the direct path. Calculate the pseudorange error that would result, including its sign, under noiseless