University of Calgary

PRISM Repository **https://prism.ucalgary.ca**

The Vault Open Theses and Dissertations

2018-05-08

New GNSS Navigation Messages for Inherent Fast TTFF and High Sensitivity - Underlying Theory Study and System Analysis

Zhang, Wentao

Zhang, W.. (2018). New GNSS Navigation Messages for Inherent Fast TTFF and High Sensitivity - Underlying Theory Study and System Analysis (Doctoral thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca. doi:10.11575/PRISM/31910 http://hdl.handle.net/1880/106628

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

New GNSS Navigation Messages for Inherent Fast TTFF and High Sensitivity - Underlying Theory Study and System Analysis

by

Wentao Zhang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN GEOMATICS ENGINEERING

CALGARY, ALBERTA

April, 2018

© Wentao Zhang 2018

Abstract

In the navigation applications on mobile devices, the extreme demands for fast Time To First position Fix (TTFF) and high sensitivity have been driving the technology innovations in these areas in recent years. Assisted GNSS (AGNSS) and Ephemeris Extension (EE) technologies constitute the efforts to improve the TTFF and sensitivity.

However, it is challenging for both AGNSS and EE. For AGNSS, while it attempts to improve TTFF on sensitivity on mobile devices, it is subject to frequent ephemeris expiration and therefore it requires the mobile devices to be always or frequently connected to the assisting networks. For EE - a technology complementary to the AGNSS to improve TTFF, although it requires little connectivity to assisting networks by directly using some extended ephemerides (valid for days) in the first position fix, such extended ephemerides can be hardly used as the aiding data for tracking weak signals.

In the analysis of the challenges, this thesis points out that, such challenges are originated from the weakness in the fundamental design of the existing GNSSs – the life expectancy of ephemeris is too short. Then this thesis proposes an alternative solution for future GNSSs, to fundamentally resolve the above issues by broadcasting some new navigation (NAV) messages with validity for up to 24 hours instead of those used by current GNSSs. Through the study of the TTFF and sensitivity fundamentals, this thesis fully explains how the ephemeris life expectancy relates to TTFF and sensitivity; and through fundamental study on orbital determination theories and ephemeris extension practices, this thesis confirms the feasibility to

obtain long-validity ephemerides; and through some simulated uses of the long-validity ephemerides in some typical scenarios, this thesis further confirms the navigation availability and accuracy using the proposed new NAV messages are comparable to those using the current NAV messages. Therefore, for a GNSS that deploys the proposed NAV messages, the capability to achieve fast TTFF and high sensitivity on a mobile device is inherently enhanced, with minimum or even no reliance on assisting infrastructures.

Table of Contents

Abstract ii
Table of Contents iv
List of Tables ix
List of Figures and Illustrations x
List of Symbols xiv
List of Abbreviations and Nomenclature xv

CHAPTER 1 INTRODUCTION 1
1.1 Background 1
1.2 Problem Statement 2
1.3 Research Objectives and Contribution of this thesis 4
1.4 Literature Review 6
1.5 Thesis Outline 9

CHAPTER 2 FUNDAMENTALS OF GNSS TTFF AND SENSITIVITY 12
2.1 NAV Messages of Existing GNSSs 12
2.1.1 Navigation Data 12
2.1.2 Super-frame Structures 13
2.2 Position Fix on a Standalone GPS UE 15
2.2.1 Different Start-up Modes 15
2.2.2 Signal Acquisition 16
2.2.3 Signal Tracking 18
2.2.4 Time Synchronizations 18
2.2.5 Ephemeris Downloading 22
2.2.6 Measurements Computation and Position Fix 23
2.3 TTFF and the Challenges 24
2.4 Sensitivity and the Challenges 25
2.4.1 Signal Levels 25
2.4.2 Sensitivity Categories 26
2.5 TTFF/Sensitivity and Validity Period of GNSS Ephemeris/NAV Messages 28
2.6 Existing Technologies for Improving TTFF and Sensitivity 30
2.6.1 Assisted GNSS (AGNSS) 30
2.6.1.1 Protocols 31
2.6.1.2 Operation Modes 32
2.6.2 Ephemeris Extension (EE) 33
2.6.3 Limitations of the Existing Assistance Technologies 35

CHAPTER 3 GNSS EPHEMERIS WITH LONG VALIDITY PERIOD 38
3.1 Time Systems 38
3.1.1 Terrestrial Time (TT) 38
3.1.2 International Atomic Time (TAI) 38
3.1.3 Universal Time (UT1) 39
3.1.4 Coordinated Universal Time (UTC) 39

3.1.5 GPS Time (GPST) 39
3.2 Coordinate Frames 40
3.2.1 Earth Center Earth Fixed (ECEF) Frame 40
3.2.2 Earth Center Inertial (ECI) Frame 41
3.2.3 Satellite Body Frames 42
3.2.3.1 Origin of Satellite Body Frame 43
3.2.3.2 Y-D-B Frame 44
3.2.3.3 X-Y-Z Frame 45
3.2.4 Radial-Tangent-Normal (RTN) Frame 46
3.3 Perturbation forces acting on a satellite 47
3.3.1 Overall 47
3.3.2 Gravitational Forces 49
3.3.2.1 Point-mass perturbations 49
3.3.2.2 The Geopotential 50
3.3.3 Non-gravitational Forces 53
3.3.3.1 Cannon Ball (CB) Model 54
3.3.3.2 Cannon Ball + Y Bias (CBY) Model 56
3.3.3.3 SPHRC Model 57
3.3.3.4 SRDYZ Model 59
3.3.3.5 SRXYZ Model 60
3.3.3.6 SRDYB Model 61
3.3.3.7 BERNE Model 62
3.3.4 Eclipse Impact 64
3.4 Numerical integration 66
3.4.1 Runge-Kutta (RK) Integrator 66
3.4.2 Adams-Moulton (AM) Integrator 68
3.4.3 Numerical integration of satellite orbits 69
3.5 Estimation of Initial Conditions (ICs) 70
3.5.1 Observations 70
3.5.2 Derivatives at $\mathbf{t_k}$ 70
3.5.3 Time transition of Partials 71
3.5.4 State Vector 72
3.5.5 Observations and Observation Equation 73
3.5.6 Normal equation and WLS estimation of ICs 73
3.6 Satellite Clock Modeling 75
3.7 Ephemeris Extension with Estimated Parameters 77
3.7.1 Basic Steps 78
3.7.2 Flow chart 79
3.7.3 Evaluation of Extended Ephemeris 80
3.7.3.1 Orbital error in ECEF 80
3.7.3.2 Orbital error in RTN 81

CHAPTER 4 FAST POSITION FIX WITH LONG-VALIDITY EPHEMERIS 83
4.1 Fast Acquisition 83
4.1.1 Prediction of Code Phase and Doppler 83

4.1.2 Components Needed for Acquisition Prediction 86
4.1.3 Uncertainties in Satellite Position and Velocity 87
4.2 Fast Time Synchronization 89
4.2.1 BitSync 89
4.2.2 FrameSync 91
4.2.3 Soft TimeSync 94
4.2.4 Significance of long-validity ephemeris on Fast TimeSync 94
4.3 Obtaining Pseudorange Measurements 94
4.4 Fast First Position Fix 96
4.4.1 Position Fix before Ephemeris is Decoded Over The Air (OTA) 96
4.4.2 Position Fix before Full TimeSync 96
4.4.3 Position Fix with Snapshot Measurements 97
4.4.4.1 Determine approximate user location 98
4.4.4.2 Construct full pseudoranges 101
4.4.4.3 Estimate time offset and accurate user location 105
4.4.4 Feed-forward the Estimated Time for Fast Time Synchronization 109
4.5 Position Fix with 3 SVs 114
4.6 Concluding remarks 120

CHAPTER 5 HIGH SENSITIVITY WITH LONG-VALIDITY NAV MESSAGES 121
5.1 Overview 121
5.1.1 Sensitivity of GNSS Receiver 122
5.1.2 Obtaining High Sensitivity 125
5.2 Signal Acquisition 128
5.2.1 Theory of Signal Detection 129
5.2.2 Signal Acquisition 131
5.2.3 Prediction of Present Signals 133
5.2.4 Coherent and Non-coherent Integration 134
5.3 Signal Tracking with Long Validity NAV Messages 137
5.3.1 Tracking of Weak signals 137
5.3.2 Prediction and Maintenance of Alignment with Bit Boundaries 140
5.3.3 Bit Decoding in Weak Signal Condition 146
5.4 Position Fix with Long Validity Ephemeris under Weak Signal Condition 149
5.4.1 Challenge in Discriminating Reflected and Direct Signals 151
5.4.2 Challenge in Outlier Detection Algorithms 152
5.4.3.1 Guard range for measurements to be used in positioning 155
5.4.3.2 Classification of Residual Clusters 157

CHAPTER 6 PROPOSAL OF NEW GNSS NAV MESSAGES 161
6.1 Weakness in the Design of NAV Messages of Existing GNSS 161
6.2 Considering New GNSS NAV Messages 163
6.2.1 Proposal of New GNSS NAV Messages 164
6.2.2 Advantages of a GNSS with the Proposed New NAV Messages 171
6.2.2.1 Standalone UEs, New GNSS vs. GPS 172
6.2.2.2 Assisted UEs, New GNSS vs. GPS 172

6.2.2.3 Current GNSS, Standalone vs. Assisted 173
6.2.2.4 New GNSS, Standalone vs. Assisted 173
6.3 Concerns of Deploying the New NAV Messages 174
6.3.1 Accuracies of Satellite Orbit and Clock 174
6.3.2 Broadcast of the New NAV Messages 177
6.3.3 Satellite Orbital Maneuver and Clock Adjustment 179
6.4 Equivalence of Cartesian and Keplerian Expression in Satellite Orbit 182
6.4.1 Conversion between Cartesian and Keplerian 182
6.4.1.1 Conversion from Keplerian to Cartesian 182
6.4.1.2 Conversion from Cartesian to Keplerian 183
6.4.2 Convenience of sanity checking using Keplerian 183
6.5 New Method for User Range Accuracy (URA) 185
6.5.1 Bounds of LOS errors from orbital error 186
6.5.2 URE Algorithm and the Problems 188
6.5.3 Deterministic Calculation on URA based on Symmetry of Orbit Errors 190
6.5.3.1 Orbital Prediction 190
6.5.3.2 Symmetry of the Orbital Errors from Forward and Backward Integration 191
6.5.3.3 URE based on the Error Symmetry 193
6.5.4 Advantage of the New URA Method 193
6.5.5 Numerical Example 194

CHAPTER 7 SIMULATION AND ANALYSIS ON GNSS THAT DEPLOYS THE NEW NAV MESSAGES 201
7.1 Deployment of the New NAV Messages in a GNSS 201
7.1.1 System Architecture of a GNSS Deploying the New NAV Messages 202
7.1.2 Deploying the New NAV Messages on CS 203
7.1.3 Deploying the New NAV Messages on SS 205
7.1.4 Deploying the New NAV Messages on UE 206
7.2 Simulation and Validation of the Deployment of the New NAV Messages 208
7.2.1 Simulation and Validation of the New Ephemeris 208
7.2.2 Simulation of the Use of the New NAV Messages 209
7.3 Results and Analysis 211
7.3.1 Evaluation of the New Ephemeris in Satellite Orbital Domain 211
7.3.2 Evaluation of the New Ephemeris in UE Navigation Domain 216
7.3.2.1 Trajectories 216
7.3.2.2 Statistics of Cross Track Errors 217
7.3.2.3 LOS Errors of Used SVs during the Usage period 218
7.3.3 Typical Benefits from the Use of Long Validity Ephemeris 220
7.3.3.1 UE Startup in Open sky 220
7.3.3.2 UE Startup in Urban Canyon 222
7.3.3.3 TTFF of Using the New NAV Messages 224
7.3.4 Simulated Use of the New NAV Messages in Weak-Signal Scenario 225

CHAPTER 8 CONCLUSIONS AND FUTURE WORK 234

8.1 Thesis Summary 234
8.2 Conclusions.....236
8.3 Future work.....238

APPENDIX.....240

REFERENCES 242

List of Tables

Table 2-1 Components of ToT in Different Time Scales 21

Table 2-2 Comparison of NAV Messages among Existing GNSSs on Length of Validity Period 28

Table 2-3 EE Products in GNSS Products 34

Table 2-4 Comparison of AGNSS and EE in improving TTFF and Sensitivity 35

Table 3-5 Unit Vectors related to Satellite Body Frame 42

Table 3-6 Summary of Perturbation Forces Acting on a GPS Satellite in Space 48

Table 3-7 Derivatives of Defined Unit Vectors 55

Table 4-8 Preconditions on Position Fix with Snapshot Measurements 98

Table 4-9 Deriving BitSync and FrameSync from Estimated ToT 114

Table 5-10 Ideal SNR at Different Noise Bandwidth 123

Table 5-11 Sources of Large Errors in High Sensitivity Scenario 150

Table 6-12 Proposed Contents of New GNSS Navigation Messages 169

Table 6-13 Comparison of the NAV messages for GPS/GLO/BDS/GAL/New NAV 171

Table 6-14 Statistics of the Correlation for all Studied GPS SVs 199

Table 6-15 Statistics of the Correlation with the worst SV excluded 200

Table 7-16 Statistics of LOS Orbital Errors 215

Table 7-17 Cross Track Errors for the Trajectories using EE of Different Ages 218

Table 7-18 Position Accuracies and Availabilities, with Startup in Open Sky 221

Table 7-19 Position Accuracies and Availabilities, with Startup in Urban Canyon 223

Table 7-20 Horizontal Position Errors in the Trajectories Using the Simulated New NAV Msg at Different Ages 232

List of Figures and Illustrations

Figure 2-1 Super-frame Structure of GPS NAV messages 13
Figure 2-2 Super-frame Structure of GLONASS NAV messages 14
Figure 2-3 Super-frame Structure of BDS D1 NAV messages 14
Figure 2-4 Super-frame Structure of GALILEO F/NAV messages 15
Figure 2-5 Different Time Scales in GPS Signal (after Kaplan, 2005) 21
Figure 2-6 Typical process of position fix on a standalone GPS UE 24
Figure 2-7 Theoretical BER with GPS Signal Strength 27
Figure 3-8 Relationship among Different Time Systems 40
Figure 3-9 Relation between Satellite MC and APC 44
Figure 3-10 Illustration of Satellite Y-D-B body frame 45
Figure 3-11 Illustration of SV body frame X-Y-Z 46
Figure 3-12 Illustration of R-T-N Frame 47
Figure 3-13 Illustration of Perturbation Forces Acting on a Satellite in Space 48
Figure 3-14 Chances for a GPS Satellite to Cross Eclipse 65
Figure 3-15 Illustration of Estimation Iterations - Case 1 75
Figure 3-16 Illustration of Estimation Iterations - Case 2 75
Figure 3-17 Illustration of Satellite Clock Fitting 77
Figure 3-18 Basic Steps in Ephemeris Extension 79
Figure 3-19 Flowchart of Parameter Estimation and Orbital Prediction 80
Figure 3-20 Orbital Errors Expressed in ECEF and RTN 82
Figure 4-21 Maximum LOS Velocity due to SV Dynamics 84
Figure 4-22 Prediction of Code Phase and Doppler of a SV Signal 86
Figure 4-23 Typical Orbital Errors vs. Age for Alm, Eph and Extended Eph 88

Figure 4-24 Relative Bit Edge between SVs 91
Figure 4-25 FrameSync and Relative Bit Edges between Different SVs 93
Figure 4-26 List of Candidate Locations for LSA 100
Figure 4-27 Resolving Msec Ambiguities in obtained PRs 104
Figure 4-28 Illustration of the Fast Position Fix 109
Figure 4-29 Iterative Estimation Process without FrameSync 111
Figure 4-30 Process of Regular Position Fix 112
Figure 4-31 Process of Position Fix without FrameSync 112
Figure 4-32 Process of Position Fix with Feed-forward of Estimated Time 113
Figure 4-33 Metrics of PR Residuals for a 3-SV scenario with PDOP 2.2 116
Figure 4-34 Metrics of PRR Residuals for a 3-SV scenario with PDOP 2.2 117
Figure 4-35 Metrics of PR Residuals for a 3-SV scenario with PDOP 53.4 118
Figure 4-36 Metrics of PRR Residuals for a 3-SV scenario with PDOP 53.4 119
Figure 5-37 Simplified Block Diagram of a Single Channel for Baseband Processing 126
Figure 5-38 Search Space in Weak Signal Acquisition 129
Figure 5-39 Alignment between Aided NAV Bits and Msec Samples during Data Wipe-Off 141
Figure 5-40 Fast Time Synchronization Through Pattern Matching of A Few Decoded Bits in the Known NAV bits 143
Figure 5-41 Maintaining the Alignment of NAV Bits and Ms Samples 145
Figure 5-42 Taking Care of the Drift of Bit Boundary 145
Figure 5-43 CN0 Distribution in Urban Canynon from a Realistic Driving Test 147
Figure 5-44 Statistics of Time for CNo being Continuously Stronger than 27 dB-Hz 148
Figure 5-45 GPS Signal Blockage in Urban Caynon 150
Figure 5-46 Example of Measurement Errors vs. CN0 in Urban Canyon 154
Figure 5-47 Illustration of Measurement Residuals 158

Figure 5-48 Sorting of Measurement Residuals 159

Figure 6-49 Initial Condition of a Satellite Orbit..... 166

Figure 6-50 Keplerian Elements of a Satellite Orbit 168

Figure 6-51 Super-frame Structure of New GNSS NAV messages 170

Figure 6-52 Progress of Atomic Clock Stability (Courtesy of Meiser, 2014) 177

Figure 6-53 New NAV Messages, Time of Update (TOU) and Reference Time of Ephemeris (TOE) 178

Figure 6-54 Example of GPS Orbital Maneuver - Impact on Orbit Altitude 180

Figure 6-55 Example of GPS Orbital Maneuver - Impact on Orbit Speed..... 181

Figure 6-56 Example of Variations in Keplerian Elements (Ω, w) 184

Figure 6-57 Example of Variations in Keplerian Elements (*a, e, i*) 184

Figure 6-58 Lower and Upper Bounds of LOS Error due to Orbital Error 187

Figure 6-59 Statistics for User Range Errors (URE) 189

Figure 6-60 Concept of 'Approximate Accuracy Symmetry' 192

Figure 6-61 ECEF position errors for forward and backward orbits 196

Figure 6-62 Comparison of error magnitudes for forward and backward orbits 197

Figure 6-63 ECEF velocity errors for forward and backward orbits 198

Figure 6-64 Comparison of velocity error magnitudes for forward and backward orbits 199

Figure 7-65 System Architecture of a GNSS Deploying the New NAV Messages 202

Figure 7-66 Deployment of the New NAV Message on CS..... 205

Figure 7-67 Deployment of the New NAV Messages on SS 206

Figure 7-68 Deployment of the New Messages on UE 207

Figure 7-69 Errors of Extended Orbit for SV 1 Expressed in ECEF and RTN 212

Figure 7-70 Errors of Extended Orbit for SV 2 Expressed in ECEF and RTN 213

Figure 7-71 Errors of Extended Orbits for Different SVs in Radial Direction..... 213

Figure 7-72 Errors of Extended Orbits for Different SVs in Tangent Direction 214

Figure 7-73 Errors of Extended Orbits for Different SVs in Normal Direction 214

Figure 7-74 URE of Extended Ephemeris at Ages from 1 to 5 days 215

Figure 7-75 Trajectories of Using EE at Different Ages 217

Figure 7-76 LOS Error bounds in Extended Orbits for the Used Satellites in above testing 219

Figure 7-77 Actual LOS Errors in Extended Orbits for the Used Satellites during the test 220

Figure 7-78 Trajectories of Using BE/EE, w/ Startup in Open Sky 221

Figure 7-79 Comparison of Used SVs for Using BE and EE, with Startup in Open Sky 222

Figure 7-80 Trajectories of Using BE and EE, with Startup in Urban Canyon 223

Figure 7-81 Comparison of Used SVs for Using BE and EE, w/ Startup in Urban Canyon 224

Figure 7-82 Theoretical WarmStart TTFF for GPS and future GNSS with the new NAV Msg 225

Figure 7-83 Comparison of Trajectories from Simulated Use of the New NAV Messages at Different Ages (< 24 hrs) 226

Figure 7-84 Average Signal Strength during the Test in Urban Canyon 227

Figure 7-85 Street View of the Signal Environment where the Large Exursion Occurred 228

Figure 7-86 Signal Condition whereat the Large Exursion Occurred 229

Figure 7-87 Trajectory Errors in East/North for the Simulated Use of New NAV Messages in a Weak Signal Scenario 231

Figure 7-88 Difference in the Trajectories with New NAV Messages in a Weak Signal Scenario 232

Figure 7-89 Horizontal Position Errors in the Trajectories Using the Simulated New NAV Msg at Different Ages 233

List of Symbols

Symbol	Definition
a_u	astronomical unit
$\mathbf{C}_{ECEF}^{ECI}$	transition matrix from ECEF frame to ECI frame
$\mathbf{C}_{ECI}^{RTN}$	transition matrix from ECEF frame to ECI frame
C_R	radiation pressure coefficient
$\mathbf{e}_B$	unit vector in the solar panel, perpendicular to the Y axis
$\mathbf{e}_D$	unit vector toward satellite from the Sun
$\mathbf{e}_H$	unit normal vector of the orbital plane
$\mathbf{e}_N$	unit normal vector of the orbit, a different expression of $\mathbf{e}_H$
$\mathbf{e}_R$	unit vector along radial direction, positive from the Earth
$\mathbf{e}_T$	unit vector tangent to the orbit
$\mathbf{e}_V$	unit velocity vector
$\mathbf{e}_X$	unit vector perpendicular to the Y-Z plane
$\mathbf{e}_Y$	unit vector of the rotation axis of the satellite's solar panel
$\mathbf{e}_Z$	unit vector toward the Earth center
$\mathbf{I}_{3\times3}$	3x3 identity matrix
m	mass of the satellite
$\mathbf{N}$	Earth nutation matrix
$\mathbf{p}$	vector of solar radiation parameters
$\mathbf{P}$	Earth procession matrix
$\mathbf{r}$	position vector
$\lvert\mathbf{r}\rvert$	magnitude of vector $\mathbf{r}$
$\dot{\mathbf{r}}$	velocity vector
$\ddot{\mathbf{r}}$	acceleration vector
$\mathbf{r}_1 \times \mathbf{r}_2$	cross product of $\mathbf{r}_1$ and $\mathbf{r}_2$
$\mathbf{r}_1 \cdot \mathbf{r}_2$	dot product of $\mathbf{r}_1$ and $\mathbf{r}_2$
$\mathbf{S}$	Earth rotation matrix
$\mathbf{W}$	Earth wobble matrix

List of Abbreviations and Nomenclature

ABDS	Assisted/Assisting BDS
AGPS	Assisted/Assisting GPS
AGNSS	Assisted/Assisting GNSS
ALM	Almanac
AM	Adams-Moulton
AB	Adam-Bashforth
APC	Antenna Phase Center
BDS	BeiDou System
BDT	BeiDou Time
BE	Broadcast Ephemeris
BER	Bit Error Rate
CODE	Ceter of Orbit Determination in Europe
DE	Development Ephemerides
DGPS	Differential GPS
DGNSS	Differential GNSS
DLL	Delay Locked Loop
ECEF	Earth Center Earth Fixed Frame
ECI	Earth Center Inertial Frame
EE	Extended Ephemeris, or Ephemeris Extension
EGM	Earth Gravity Model
EOP	Earth Orientation Parameters
EPH	Ephemeris
ER	Earth Rotation
FLL	Frequency Locked Loop
GEO	Geostationary Orbit
GPST	GPS Time
HS	High Sensivitity
HOW	Hand Over Word
IAU	International Astronomical Union
IC	Initial Condition(s)
IF	Intermediate Frequency
IGS	International GNSS Service
IGSO	Inclined Geostationary Orbit
IERS	International Earth Rotation and Reference Systems Service
JPL	Jet Propulsion Laboratory
LS	Leap Seconds
LSA	Location Search Algorithm
MC	Mass Center
MEO	Medium Ellipse Orbit
MS	Mobile Station
MSA	Mobile Station Assisted

MSB	Mobile Station Based
NAGU	Notice Advisory to Glonass/Galileo/GNSS Users
NANU	Notice Advisory to Navstar Users
NAV	Navigation
NGA	National Geospatial-Intelligence Agency
NOAA	National Oceanic and Atmospheric Administrator
NTP	Network Time Protocol
OTA	Over The Air
PDI	Pre-Detection Integration
PECE	Predictor-Corrector
PLL	Phase Locked Loop
PTP	Precise Time Protocol
PR	Pseudorange
PRN	Pseudo Random Number
PRR	Pseudorange Rate
RK	Runge-Kutta
RF	Radio Frequency
RTN	Radial-Tangent-Normal Frame
SNR	Signal to Noise Ratio
SV	Satellite Vehicle
SR	Solar Radiation
SL	Squaring Losss
SRP	Solar Radiation Parameter(s)
TAI	International Atomic Time
ToR	Time of Reception
ToT	Time of Transmission
TOW	Time of Week
TT	Terrestrial Time
TTFF	Time to First Position Fix
UE	User Equipement(s)
URA	User Range Accuracy
URE	User Range Error
UT1	Univeral Time
UTC	Coordinated Univeral Time
WLS	Weighted Least Squares
WN	Week Number

CHAPTER 1 INTRODUCTION

1.1 Background

When GPS was initially designed in 1970's, it might have never been foreseen that it would be so popular in everyday life one day. The boom of GPS navigation applications did not arrive until around the year 2000, after the cancellation of GPS SA policy in 2000 (The White House, 2000) and the mandate of E911 (FCC, 2001). The boom continues till today, and the demands for high quality location information under all different difficult conditions have driven the technology advancement in GNSS industry, not only in accuracy and reliability, but also fast time to first position fix (TTFF) and high sensitivity.

Position fix on a standalone GNSS device normally consists of signal acquisition, tracking, bit synchronization, frame synchronization, ephemeris downloading, measurements taking and position computation, which could be a process varying from a few seconds to tens of seconds.

It is a big challenge for the standalone GNSSs to provide seamless, timely and reliable positioning information in some extremely weak-signal environments because of the difficulties of GNSS user equipment (UE) in either obtaining the ranging information or decoding the navigation (NAV) messages from the weak signals. In the past decade, great efforts have been spent by the industry and research institutes to get fast TTFF and high sensitivity on GNSS UEs by developing ground or space assisting systems. However, such assisting systems require the support from dedicated infrastructures, which usually are only available in the vicinity of urban areas.

1.2 Problem Statement

The representative technologies of ground assisting systems are Assisted GPS (AGPS) and Ephemeris Extension (EE). Along with the implementation of additional GNSS, such as GLONASS and BEIDOU System (BDS) in the assisting systems, AGPS is evolving to Assisted GNSS (AGNSS). Both AGNSS and EE technologies attempt to provide alternative ephemeris from ground assisting systems to UEs when the GNSS signals are too weak to allow a timely downloading of the NAV messages from the GNSS satellites. The difference between the two technologies is that the ephemerides sent to UEs by AGNSS are authentic and collected elsewhere, whereas those provided by EE are synthesized and have much longer life expectancy. Moreover, AGNSS is able to provide additional aiding data, including location and time information, to expedite the signal acquisition on UEs. It enables fast TTFF and high sensitivity on UEs, as long as the UEs are under the coverage of ground assisting systems, and aiding data is updated from AGNSS every 2 hours (taking GPS for example) or from EE every a few days. As to space assisting systems, the only difference to traditional AGNSS is that the assisting infrastructure is space-borne, and a representative system is using Iridium satellites to relay GNSS ephemerides to ground UEs (Gibbons, 2008; Iridium, 2016). All these assisting systems together ensure that fast TTFF and high sensitivity can be achieved on the ground GNSS UEs under different challenging environments.

However, both the ground and space assisting systems are subject to significant limitations when obtaining fast TTFF and high sensitivity on UEs, including the following aspects:

- Firstly, developing and maintaining additional infrastructure for assisting GNSS is a tremendous effort, which is complicated and costly.

- Secondly, additional communication modules are required on the GNSS UEs to get aiding data from the assisting systems, which result in increased cost, system complexity, and power consumption.

- Thirdly, the GNSS UEs have to keep connected to the assisting systems to get up-to-date aiding data for fast TTFF and high sensitivity, as the NAV messages from the existing GNSSs change frequently, say every 2 hours for GPS. Therefore, additional cost will be incurred to UEs for frequent data transferring from ground assisting systems, and the cost will be significant if to use the space assisting systems.

- Last but not least, fast TTFF and high sensitivity are only obtainable within the areas covered by the signals of the assisting systems.

For all the existing GNSSs (GPS/GLONASS) including the regionally operational BDS and GALILEO, it is an expectation that fast TTFF and high sensitivity are always challenging for the standalone UEs, since the NAV messages for all these systems are designed to change too frequently - every 15 min to 3 hours. So the reliance on assisting system is a fact for all the GNSSs (GPS/GLONASS/BDS/GALILEO), to obtain fast TTFF and high sensitivity on the UEs.

1.3 Research Objectives and Contribution of this thesis

From the perspective of TTFF and sensitivity, there is significant weakness in the design of NAV messages in existing GNSSs. Instead of spending efforts on developing and improving the assisting systems, this research is to focus on improving the design of GNSS NAV messages with an extension in the update period – say once per day, so that with the improved design, future standalone GNSS UEs will have TTFF and sensitivity performance equivalent to what can be achieved in today's GNSS UEs only with assisting systems.

More specifically, the objectives of this thesis are:

- To research the relationship between TTFF/sensitivity and the validity period of GNSS ephemeris/NAV messages;
- To study the methods to extend the validity period for GNSS ephemeris;
- To study the algorithms for TTFF improvements on standalone GNSS UEs with long validity GNSS ephemeris;
- To study the algorithms for sensitivity improvements on standalone GNSS UEs with long validity GNSS NAV messages;
- To come up with the design of new GNSS NAV messages with long validity period and low update rate;
- To investigate the performance of the system that deploys the new NAV messages.

The primary contributions from this thesis will consist of the following aspects:

- An innovative proposal for design of GNSS NAV messages with long validity period and low update rate (Section 6.3);
- A new method for providing the User Range Accuracy (URA) for systems that deploy the proposed NAV messages (Section 6.6);
- A new method for fast position fix with snapshot measurements immediately after signal acquisition (Section 4.4.3);
- A new method to feed-forward estimated time for fast time synchronization (Section 4.4.4);
- A new method for detection of the image solution, which often happens when only 3 SVs are used to determine a fast first position fix (Section 4.5);
- Fast time synchronization with bit pattern matching (5.3.2);
- Two new methods for dealing with measurement outliers in weak signal conditions (Sections 5.4.3.1 and 5.4.3.2);

As an alternative to AGNSS and EE, the solution proposed in this research will enable future standalone GNSS to provide the performance that is currently achievable only with the help of additional assisting infrastructure, and to provide even superior performance if assisted. For systems deploying the proposed solution, there are also additional benefits, such as:

- Less uploading load for ground GNSS control center;
- Less downloading load, simplified design, and lower cost on GNSS UEs;

- No or minimum reliance on additional assisting systems, therefore reduced cost in developing and maintaining assisting systems, and reduced cost on GNSS UEs for using such assisted data;
- No or minimum limitation on areas of coverage.

Moreover, considering the enormous number of patents and frequent infringe on patents in GNSS industry, the research of this thesis can help break some patent barriers, as some critical patents related to AGNSS and EE will be automatically invalidated with the deployment of the new proposal.

1.4 Literature Review

The efforts of improving the TTFF and sensitivity on GNSS UEs can be categorized into three different aspects: (1) developing assisting systems, (2) inventing and implementing new algorithms for standalone UEs, and (3) improving and upgrading the design of existing GNSS systems.

The concept of assisted GPS (AGPS) was brought up by industry in late 1990's, when lots of patents were filed, and then granted in early 2000's (King et al., 2001; Krasner, 2001; McBurney et al., 2001; Paul et al., 2001; Zadeh et al., 2001). Seeing the challenges of TTFF and sensitivity on standalone GPS devices, the general idea from the patents is to provide assisting information to GNSS UEs, such as time, rough location, a list of satellites in view, Doppler of each satellite, ephemeris, and etc., in a way to speed up each stage in the process of position fix. From

system/infrastructure perspective, AGPS has been well described in Kaplan et al. (2005), Diggelen (2009) and LaMance (2003). In a typical network-assisted AGPS system, a Serving Mobile Location Center (SMLC) is established to process the data (navigation messages) collected by the Wide Area Reference Network (WARN) and package the assistance data in Information Elements (IEs). Upon requests from User Equipments (UEs), the SMLC sends out IEs through Mobile Switching Center (MSC) and Radio Network Controller (RNC), until finally arriving at the UEs. With series of AGPS specifications embodied and continuously revised in the 3GPP and Open Mobile Alliance (OMA) standards since 2001 (3GPP, 2001), AGPS-enabled products then became popular in the GNSS market.

It is also reported that the Iridium satellites are being upgraded to assist GPS, with the underlying principle similar to AGPS (Iridium, 2016). When the connections between the Iridium satellites and their UEs are built and maintained, GPS ephemerides can be relayed from Iridium satellites to the UEs. Meanwhile, the UE time can be synchronized to Iridium time at a precise level, and the UEs can be located by Iridium satellites at the accuracy level of 10 km (Iridium, 2012), which all together can greatly speed up the GNSS TTFF and improve acquisition/tracking performance.

With additional assistance, although performance of fast TTFF and sensitivity is obtainable on the GNSS UEs, it is still a challenge when network connectivity is not available. A technology referred to as ephemeris extension (EE) was introduced by two pioneer companies GlobalLocate (now a part of Broadcom) (LaMance, 2003), and SiRF (used to be a part of CSR) (Garin et al., 2008; Han 2009), which enables fast TTFF and high sensitivity on GNSS UEs even without

network connectivity. According to the introduction in Broadcom (2014) and Lundgren (2005) for their product called Long Term Orbit (LTO), and CSR (2014) for InstantFix, both are based on orbital determination theories (Ash, 1972; Montenbruck et al., 2000; Su, 2000), and provide alternative ephemeris with validity period extended to a few days, rather than 2 hours for regular GPS ephemeris. As of today, a variety of EE products have become available from other additional companies or institutes, including EASY/Hotstill/Embedded Predicted Orbit (EPO) from MediaTek (MediaTek, 2014), AssistNow from uBlox (uBlox, 2014), GPStream/PGPS from RxNetworks (RxNetworks, 2009), and Embedded Autonomous Ephemeris Prediction (EAP) for JPL (JPL, 2014), etc..

In order to improve TTFF performance on standalone GNSS UEs, some critical techniques were also introduced in industry (Paul et al., 2001; Diggelen et al., 2002) along with the efforts for more advanced IC design. One key stage in the process of position fix is frame synchronization, which is to obtain the time information from the frames of GNSS NAV messages. Each subframe of GPS NAV messages lasts for 6 seconds, which means obtaining frame synchronization takes at least 6 seconds. In order to save the time waiting for decoding the time, these techniques treat the time as an unknown parameter and have it estimated before frame synchronization, and therefore speed up TTFF by at least 6 seconds. More details of such techniques will be further introduced and studied in Chapter 4. In order to improve the sensitivity on standalone GNSS UEs, GPS signal processing has been an intensive research area since early 2000's, focusing on hardware design or algorithm studies on baseband signal acquisition and tracking, with the representative methodologies well summarized in Kaplan et al. (2005).

Foreseeing the heavy demands for GNSS, and being aware of the limitation in existing design, the GPS modernization program led by the USAF has been underway since 2002 (Reaser, 2002; GPS IS, 2004). To favor the high sensitivity on UEs, a pilot channel, namely a data-less channel is added on L1C, so that coherent integration can be inherently facilitated to obtain high sensitivity, without worrying about the erroneous data bits or expiration of data bits. GALILEO has also adopted similar design of pilot channel in E1 (GALILEO ICD, 2008) to facilitate signal tracking under weak signal conditions.

However, successful GPS bit decoding is subject to limitation of ~27 dB-Hz (Braasch et al., 1999). Even though the weak signals can be well tracked with the help of the pilot channel, position fix still might not be obtained if new ephemeris cannot be decoded, once expired. Instead of focusing on improving the ground assisting infrastructure, it was proposed in Zhang (2009) to improve the design of GNSS NAV messages, so that the downloading of ephemeris from satellites can be minimized to once or twice a day, which therefore essentially facilitate fast TTFF and high sensitivity on UEs.

The work done in this thesis constitutes the extension of the efforts in Zhang (2009), which attempts to systematically study the technical fundamentals, take in depth analysis on potential issues, and simulate and illustrate achievable performance based on the new proposal.

1.5 Thesis Outline

This thesis will consist of the following chapters:

- Chapter 1: Introduction

This chapter will cover background introduction, statement of problem, purpose of this research and methodology, literature review, and thesis outline.

- Chapter 2: Fundamentals of GNSS TTFF and Sensitivity

This chapter will extend the discussion of GNSS TTFF and sensitivity into details, which will cover the introduction of the working procedures of a standalone GPS UE, categorization of the TTFF and sensitivity according to different working phases, discussion of the normally expected performance, identification of the key factors that lead to fast TTFF and high sensitivity on a GNSS UE, analysis of the relationship between TTFF/sensitivity and the lengths of validity periods of NAV messages, and briefing of the methods to obtain NAV messages with long validity.

- Chapter 3: GNSS Ephemeris with Long Validity Period

This chapter is to focus on studying the feasibility of GNSS ephemeris with longer validity period compared to what is deployed in existing GNSS, through a deep diving into the theory of satellite orbit determination, including different perturbation forces, solar radiation models, numerical integration, and estimation methods. Methods to obtain ephemeris with extended validity will be further studied and quantitatively analyzed.

- Chapter 4: Fast Position Fix with Long-validity Ephemeris

This chapter will focus on the algorithm study for position fix at different aforementioned working stages, under the condition that GNSS ephemeris is valid for long.

- Chapter 5: High Sensitivity with Long-validity NAV Messages

This chapter will study the techniques that are commonly used in obtaining higher acquisition and tracking sensitivity in existing GNSS UEs, identify the challenges, and investigate how the sensitivity on GNSS UEs can be facilitated with long-validity NAV messages.

- Chapter 6: Proposal of New GNSS NAV messages

This chapter will come up with the details of the proposal for new GNSS NAV messages with long validity period, and further study the best update rate of NAV messages, failure detection algorithms with the new NAV messages, and practical implementation issues on GNSS UEs.

- Chapter 7: Simulation and Analysis on GNSS Deploying the New NAV Messages

This chapter will study the sub-systems that deploy such a design, discuss related challenges and associated issues, and identify the essential problems through simulation. In addition, this chapter will conduct some use case analysis for such a new GNSS by comparing to existing GNSSs, in terms of different performance metrics including TTFF and sensitivity.

- Chapter 8: Conclusions and Future Work

This chapter will summarize the research work in this thesis, including the feasibility of the proposal and performance of GNSS that deploys the proposal.

CHAPTER 2 FUNDAMENTALS OF GNSS TTFF AND SENSITIVITY

This chapter will extend the discussion of GNSS TTFF and sensitivity into more details, which will cover the introduction of the working procedures of a standalone GPS UE, categorization of the TTFF and sensitivity according to different working phases, discussion of the normally expected performance, identification of the key factors that lead to fast TTFF and high sensitivity on a GNSS UE, analysis of the relationship between TTFF/sensitivity and the lengths of validity periods of NAV messages, and briefing of the methods to obtain NAV messages with long validity.

2.1 NAV Messages of Existing GNSSs

In this section, the definitions of NAV messages of the existing GNSSs are aggregated and visualized in a unified way, although they are defined and explained in much different ways in their ICDs.

2.1.1 Navigation Data

For different GNSSs, like GPS, GLONASS, BDS and GALILEO, despite of the difference (more or less) in the structures of NAV messages, it is very similar that a super set of NAV messages is comprised of immediate (primarily ephemeris and SV clock) and non-immediate (primarily almanac) data. The immediate data is repeated at a much shorter period than the non-immediate data, and expires much sooner than the non-immediate data.

2.1.2 Super-frame Structures

The GPS L1 C/A NAV messages are broadcasted at 50 bps and consist of five sub-frames (with each lasting 6 seconds) and one super-frame (lasting for 750 seconds), as depicted in Figure 2-1. The first 3 sub-frames, necessary for the receiver position fixing, contain the satellite ephemeris, with the content repeated every 30 seconds and updated usually every 2 hours; while the last 2 sub-frames provide almanac for each satellites in the form of 25 pages, with the content updated nominally every 6 days (GPS IS 2004) and actually on a daily basis. Considering that there are different types of NAV messages in GPS system on different frequencies (Navipedia, 2018), to avoid confusion, the term "GPS NAV messages" used in this thesis always refers to the GPS L1 C/A NAV messages.

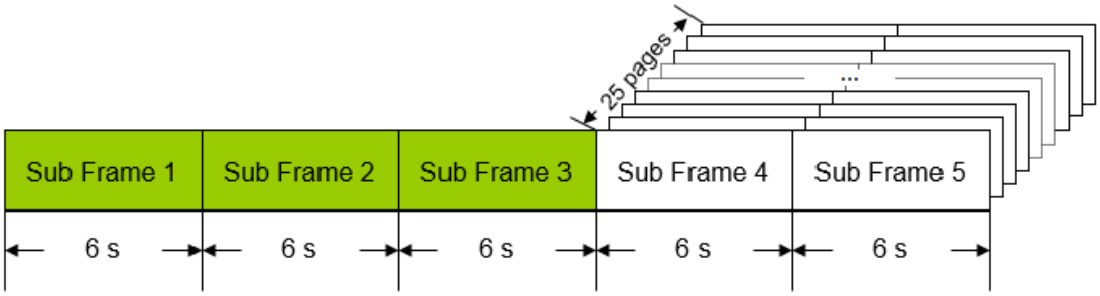


Figure 2-1 Super-frame Structure of GPS NAV messages

Also broadcasted at 50 bps, the super-frames of GLONASS NAV messages consists of 5 frames, with each lasting for 30 seconds (Figure 2-2). Each frame further consists of 15 strings, each lasting for 2 seconds, with the content repeated every 30 seconds and updated every half an hour. The first 4 strings in each frame carries the so-called immediate data – actually the ephemeris

(including SV clock), and the rest of the strings in each frame carry non-immediate data – actually the almanacs for 4 to 5 GLONASS SVs (GLONASS IS 2008).

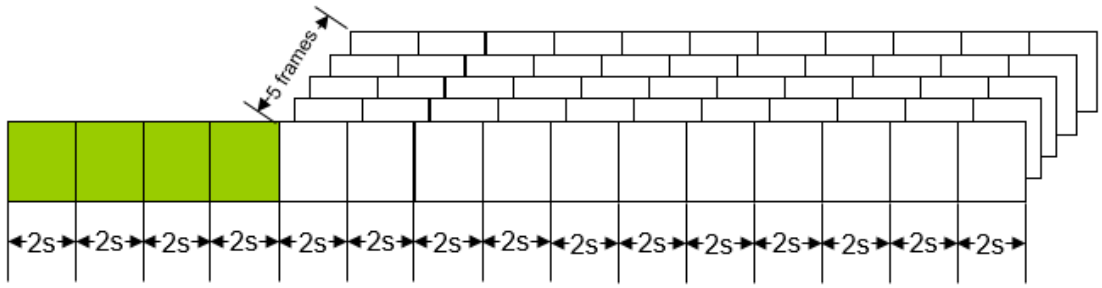


Figure 2-2 Super-frame Structure of GLONASS NAV messages

The BDS includes both MEO and GEO/IGSO satellites, with the NAV messages broadcasted at 500 bps and 50 bps respectively, and referred to as D1 and D2 messages respectively. For the ease of comparison with GPS NAV messages, only the D1 NAV messages are discussed in this thesis. As depicted in Figure 2-3, the super-frame structure of BDS NAV messages (D1) is very similar to that of GPS NAV messages, except there is 1 less page, and therefore the BDS D1 super-frame only lasts for 720 seconds. Although like in GPS NAV messages, the contents of sub-frames 1 - 3 in BDS D1 messages, carrying ephemeris, are repeated every 30 seconds, they are updated every 1 hour, much shorter than the 2 hours in GPS (BDS IS 2012).

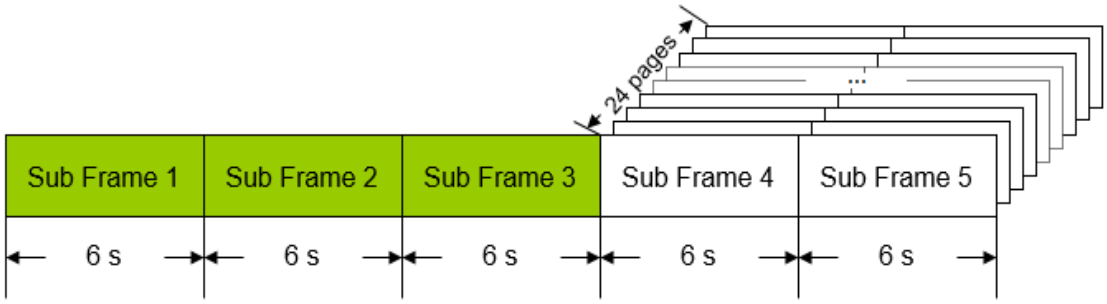


Figure 2-3 Super-frame Structure of BDS D1 NAV messages

The design of NAV messages in GALILEO is a little complicated. For ease of comparison to other GNSSs, only the GALILEO F/NAV messages are discussed in this thesis. As depicted in Figure 2-4, a super-frame of GALILEO F/NAV consists of 12 sub-frames and lasts for 600 seconds. Each sub-frame is further splitted into 5 pages, with the first 4 pages broadcasting ephemeris (including SV clock) and repeated every 50 seconds (GALILEO IS 2010).

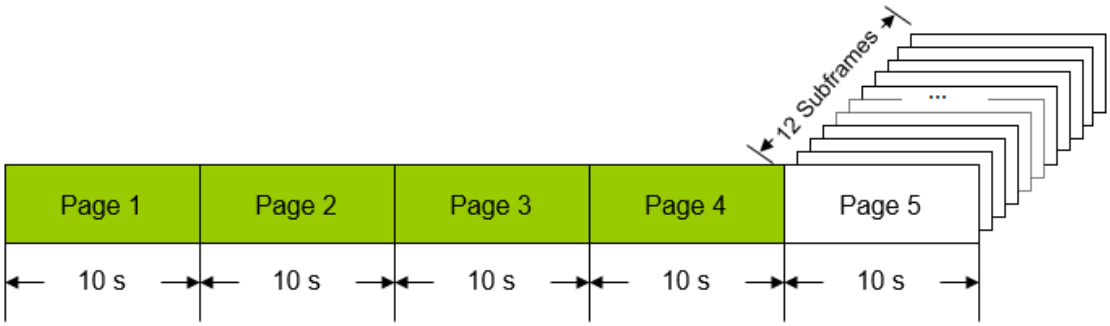


Figure 2-4 Super-frame Structure of GALILEO F/NAV messages

2.2 Position Fix on a Standalone GPS UE

As depicted in Figure 2-6, position fix on a standalone GNSS UE normally consists of signal acquisition, tracking, bit synchronization, frame synchronization, ephemeris downloading, measurements computation and position estimation.

2.2.1 Different Start-up Modes

Depending on the availability and accuracy of the data that is required for a GNSS position fix when a GPSS UE is started, including time, UE position, ephemeris/almanac, the start of the

GNSS UE is usually referred as a cold, warm or hot start, among which warm start is the most common case.

In the widely accepted definition, cold start usually refers to the case that time and UE position are totally unknown, and there are only hard-coded almanacs when the UE is started. However, warm start usually refers to the case that time is known at uncertainty of a few seconds, position is known at uncertainty of a few tens of kilometers to a few hundred kilometers, and recently downloaded almanacs (say within a few days) are available when the UE is started. As to hot start, it usually refers to the case that the accuracy of available time is less than 1 second, the accuracy of available position is within a few hundred meters, and fresh ephemerides are available.

2.2.2 Signal Acquisition

The GPS signal acquisition usually refers to acquiring the PRN codes of the GPS satellites in view. At the moment when GPS signals are transmitted from different satellites, they are precisely time synchronized. However, when arriving at the UE antenna, the signals have experienced different time delays and Doppler shifts. The signal from each GPS SV uses a unique PRN code. The tasks of signal acquisition are to detect which PRN codes are present if it is not known, and to find the exact code phase and carrier frequency for the signal of each visible SV arriving at the UE antenna at the sampling moment. So acquiring the C/A codes involves search in both time (code phase) and frequency dimensions. As the C/A code consists of 1023 chips and is repeated every millisecond, the search of code phases is up to 1023 chips. The

Doppler shift is a superimposition of satellite dynamics, UE dynamics and receiver clock drift, with the total range of up to ± 10 kHz (Tsui, 2000). The way the search is usually performed is to, choose a size of the bins in the 2-dimensional space (time and frequency), and a replicated C/A code with the code phase and frequency from each bin to correlate with 1-ms samples of the incoming signals. Signal acquisition is declared in the bin where the correlation is at maximum and also passes the detection criterion (Elliott et al., 2005).

In the case it is unknown which satellites are visible, each PRN of the 32 GPS satellites has to be examined in the received signals. Moreover, there are also cases that, some signals are present but not acquired using only 1-ms samples in the above process because the signals are too weak. In such case, the integration length of incoming samples has to be increased until obtaining adequate processing gain in Signal-to-Noise Ratio (SNR) for signal detection. In the context of high sensitivity UE, unless the signal level is known, it adds one more search dimension in signal acquisition. Moreover, when both strong signals and weak signals are present, false acquisition could happen due to cross correlation. In such case, a signal could be either detected in the name of a wrong PRN, or declared present at the wrong code phase or frequency bin.

The performance of signal acquisition on a GNSS UE is quantified by the time needed to acquire a signal at specific signal level and the lowest signal level that a signal could be acquired. The full range search in the multi-dimensional space for a signal is time consuming. In order to speed up the process, the straight-forward techniques are to either improve the processing speed by using more advanced ASIC, or to lower the processing load by narrowing down the search space through the use the best knowledge of initial time, location, satellites position and velocity. In

order to acquire weaker signal, smaller bin sizes in time/frequency and longer integration of correlation samples have to be used, which therefore incurs significantly higher processing load.

2.2.3 Signal Tracking

Once a signal is acquired, it is immediately transitioned to the signal tracking stage, and a code tracking loop followed by a carrier tracking loop are started with gradually reduced tracking bandwidths in order to lock to the signal at better resolution (accuracies) in time and frequency. The pull-in stage of the tracking loops usually takes a few hundred milliseconds, depending on the accuracy in the initial acquisition.

The tasks in this stage are to keep signals in stable tracking status, despite of changes in signal level and UE dynamics, so that snapshots of code phase and carrier frequency can be taken, the bit stream of NAV messages can be decoded, precise time can be obtained and full ranging information can be computed. Certainly, it is challenging to accomplish each of the tasks under weak signal condition, and lot of indispensable techniques will be discussed later in this thesis.

2.2.4 Time Synchronizations

Different time scales used in the GPS signals are depicted in Figure 2-5, which together enable the GPS UE to obtain highly precise measurements for position fixes.

With a signal remaining in stable tracking status, the snapshot of code phase can be taken at the time when tracking loop is updated. In the C/A code where the sampling moment situates, the exact location in the 1023 chips is the code phase, consisting of integer chips and sub-chip parts. For the sub-chip part, it is obtained by filtering the DLL discriminator results from some Early (E), Prompt (P) and Late (L) correlators. The code phase provides time resolution at nanosecond level, with the portion of time above milliseconds remaining ambiguous.

Considering that the length of a data bit is 20 ms, when C/A codes are continuously received for multiple bits, it is possible to find where the present C/A is located in the present data bit, which is often referred to as epoch number. Finding the epoch number corresponds to a process of finding the edge of a bit, which is referred to as Bit Synchronization (Sync). The Bit Sync provides time resolution of 20 ms, with the portion of time above 20 ms remaining ambiguous.

Once Bit Sync is obtained, data bits can be decoded one by one continuously. In the GPS NAV messages, every 30 bits constitute a word. However, in the bit stream, the start of a word has to be located. Locating the start of a word is accomplished in a subframe context, as every 10 words constitute a subframe, and the first word in each subframe is a Telemetry word (TLM) containing a preamble 10001011 as the indication of the start of a subframe. In other words, locating the preamble in the bit stream is equivalent to finding the start of a subframe, and therefore the start of words. Such a process takes more than 6 seconds, and is usually referred to as Frame Synchronization (Sync).

With the start of a subframe located, the Time Of Week (TOW) and subframe ID can be decoded from the 2nd word – the Handover Word (HOW). Then by referring to the definition of each subframe, the contents of the NAV data can be decoded word by word (GPS IS, 2004).

By now, it is possible to get the precise time that the sampling moment corresponds to given the above time scales that a signal is synchronized to. At the sampling instant, the start time of current subframe is known as soon as the TOW is decoded, the start time of current word can be counted with respect to the TOW considering each word lasts for 0.6 seconds, the start time of current bit is counted in a similar way from the start of current word considering each bit lasts for 20 ms, the start time of current C/A code is counted from determined epoch number (0 to 19), as soon as the code phase (integer and fractional chips) is available, the precise GPS time that the signal is aligned to is derived.

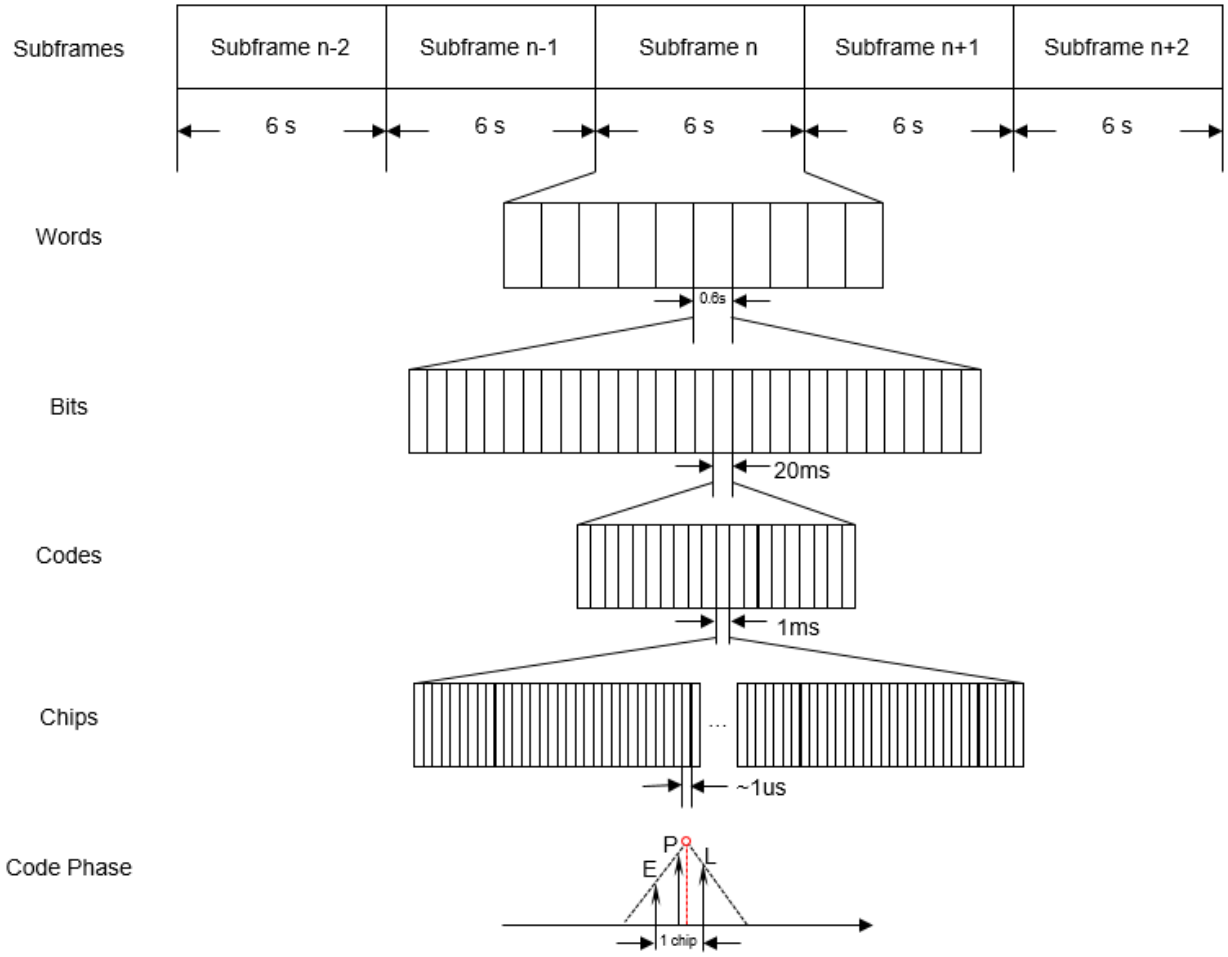


Figure 2-5 Different Time Scales in GPS Signal (after Kaplan, 2005)

At the sampling moment, the above derived GPS time actually corresponds to the Time of Transmission (ToT) of the signal from the satellite antenna, and the time of signal arrival on UE antenna is often referred to as Time of Reception (ToR). As the description of ToT, the measurements by means of the different time scales are further summarized in the following table:

Table 2-1 Components of ToT in Different Time Scales

ToT	Time	Range	Notes

Components	Scale/Resolution		
iPage	30 s ~ 750 s	1 ~ 25, integer	Describe current page number
iSubframe	6s ~ 30 s	1 ~ 5, integer	Describe current subframe
iWord	0.6 s ~ 6 s	1 ~ 10, integer	Describe the location of current Word in current Subframe
iBit	20 ms ~ 600 ms	1 ~ 30, integer	Describe the location of current Bit in current Word
iEpochNum	1 ms ~ 20 ms	0 ~ 19, integer	Describe the location of current C/A in current Bit
iChip	1 us ~ 1 ms	0 ~ 1022, integer	Describe the location of ToT in current C/A code, in integer chips
ChipPhase	0 ~ 1 us	0.0 ~ 1.0, float	The fractional part of chips, as the location of ToT in current C/A code

2.2.5 Ephemeris Downloading

The ephemeris in different literature may refer to satellite orbit part only, or both satellite orbit and clock parts. In this thesis, ephemeris usually refers to both satellite orbit and clock unless otherwise specified.

According to the super-frame structure introduced in Figure 2-1, the downloading of the full ephemeris would take 18 seconds if luckily the first collected data bit corresponds to the start of sub-frame 1 and otherwise would take 30 seconds, assuming in open sky where signal strength is strong enough to allow bit decoding at sufficiently low bit error rate (BER).

Similarly, it would take 8 to 30 seconds for GLONASS, 18 to 30 seconds for BDS, and 40 to 50 seconds for GALILEO F/NAV to download a copy of full ephemeris.

2.2.6 Measurements Computation and Position Fix

For different signals arriving at the UE antenna at ToR, once the ToT of each is derived in the process that is described in Section 2.2.4, the pseudorange (PR) measurement from a satellite can be computed as:

$$\boldsymbol{PR(i) = c \cdot [ToR - ToT(i)]} \tag{2-1}$$

Where i represents the i-th satellite, and c represents the speed of light in meters. For the multiple signals, the common time instant ToR is actually not precisely known yet when the PR is computed. That's why the computed range is given a prefix 'pseudo'. Before the PRs are computed through Equation (1), usually a rough ToR is given as:

$$ToR = ToT(0) + 0.076 \tag{2-2}$$

where 0.076 seconds is the nominal time for a signal to travel from a GPS satellite to the UE antenna on the Earth surface. For all signals that arrive at the same sampling moment, the common time instant derived in Equation (2) obviously has time error, and the truth time instant has to be resolved through estimation along with the position components in the position fix once four or more than four PRs are available.

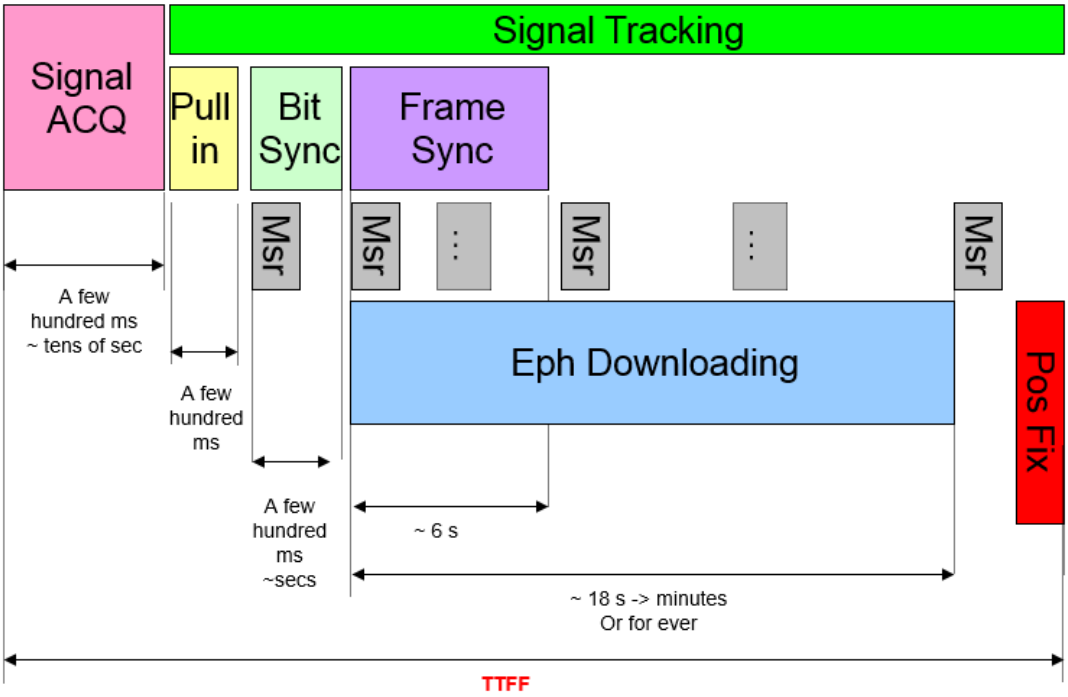


Figure 2-6 Typical process of position fix on a standalone GPS UE

2.3 TTFF and the Challenges

In doing a regular warm start, the signal acquisition usually takes only a few hundred milliseconds for a GPS device in open sky environments. However, under weak signal conditions, the signal acquisition might take much longer (say a few tens of seconds) to finish the three-dimensional search in the time, frequency and signal level domains. Once the signal gets acquired, the tracking loop is activated, and immediately after signal is pulled in, the process of data bit synchronization is started. This process takes a few hundred milliseconds to several seconds depending on signal strengths and algorithm efficiency. In the stable tracking status, the navigation bits are collected one by one. Collecting a complete copy of ephemeris (sub-frames 1 – 3) takes about 18 seconds in open sky, and minutes or forever in weak signal environments due

to increased bit error rate (BER) (see Figure 2-6). As soon as the ephemeris downloading from 3-4 satellites is completed and the measurements are taken, the user position usually can be fixed immediately.

The overall time from UE start-up to the first position fix is often referred to as the time to first fix (TTFF). In weak signal environments, both the signal acquisition and ephemeris downloading take time and thus slow the TTFF, whereas in open sky, the primary obstacle to fast TTFF is the time needed for ephemeris downloading.

2.4 Sensitivity and the Challenges

2.4.1 Signal Levels

For a GNSS UE in open sky on the earth surface, the received minimum signal level from GPS L1 is around -130 dBm, and a higher level is possible, but is not expected to exceed -125.5 dBm (GPS IS, 2004). For other GNSS signals, the nominal received signal levels are approximately same comparing to GPS (GLONASS ICD 2008; BDS ICD 2013; GALILEO ICD 2008). However, in some extreme cases, such as urban canyon, foliage and indoor environments, the signals finally arriving at receiver antenna could be heavily attenuated down by -30 dB or even more because of signal reflection, blockage and penetration. Working under such conditions is almost a mission impossible for receivers with normal tracking sensitivity, but is required for a high sensitivity receiver. The minimum TTFF and sensitivity performance conformance tests for AGPS has been given in 3GPP specifications (3GPP 2009), and it is actually common for some

GNSS chips in industry to have solid navigation at signal level of -160 dBm and even further down to -163 dBm by different means of assistance (GPS Business News, 2009).

GNSS UEs capable of working work under normal signal levels are regarded as normal sensitivity, and those capable of working with signals significantly below (say 30 dB or more) the nominal signal level are regarded as high sensitivity.

2.4.2 Sensitivity Categories

Each stage of the positioning process described in Section 2.2 is highly impacted by the signal level. When the GNSS signal strength drops, it might not be easily acquired even if it is present in the samples. The lowest signal level at which the signal can be acquired on a GNSS UE is referred to as the acquisition sensitivity of the UE.

When the GNSS signal strength drops to a certain level, it causes immediate difficulties in the GNSS receiver tracking loop and ephemeris downloading. The parameters of the tracking loop, designed for normal signal strengths, are no longer good to either obtaining enough gain for the signal detection or maintaining signal tracking. So the lowest signal level at which the signal can be remained in stable tracking status is referred to as the tracking sensitivity of the UE.

From perspective of NAV message decoding over the air (OTA), bit error rate (BER) increases with the decrease of signal strength, as depicted in Figure 2-7. When the signal level drops below

27 dB-Hz, even if the signal tracking is maintained, it would be difficult for successful decoding of navigation messages (Braasch et al., 1999), due to the high level of BER.

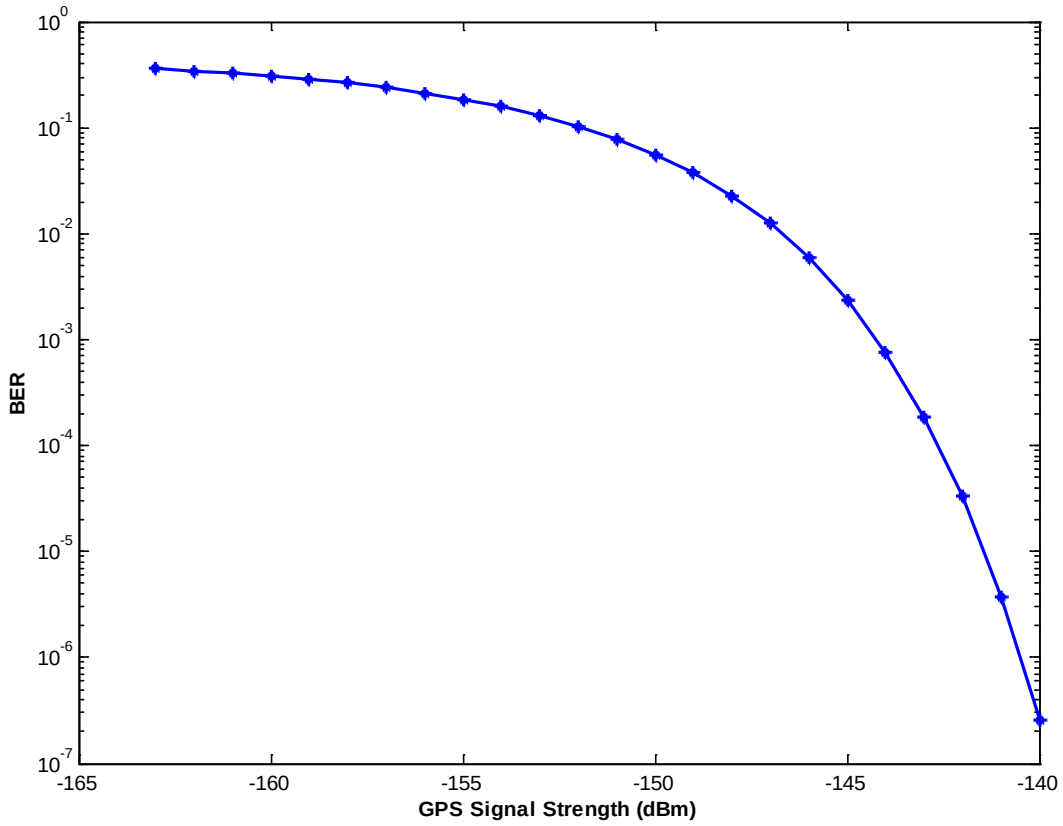


Figure 2-7 Theoretical BER with GPS Signal Strength

Under extremely weak signal conditions, even if the signals are remained in stable tracking, they might not be usable in providing a reliable position fix. Therefore, the lowest signal level at which reliable fix can be provided is referred to as the positioning sensitivity of a GNSS UE.

The achievable sensitivities in the above categories are highly dependent on the deployed algorithms in the GNSS UEs, except the Bit decoding sensitivity, which is limited to ~27 dB-Hz by the BPSK theory and the parity check algorithms implemented in the GNSS NAV message.

2.5 TTFF/Sensitivity and Validity Period of GNSS Ephemeris/NAV Messages

Taking GPS as the example, the TTFF and sensitivity challenges will be addressed through the introduction of different stages as depicted in Figure 2-6. The overall TTFF is dependent on the time spent in each of the stages, among which ephemeris downloading often takes most of the time. For sensitivity, each stage is subject to different limitations, depending on different conditions when the UE is started. Therefore, each stage will be introduced in details, including the purpose, the operation principles, the bottlenecks and the limitations. In terms of sensitivity and required time in each stage, introduction of regularly achievable performance, the best performance and preconditions will also be included.

Then the study will relate the challenges to the length of validity period in the GNSS ephemeris/NAV messages (Table 2-2), and extend the study to other existing GNSSs, including GLONASS, BDS and GALILEO. The key factors that lead to fast TTFF and high sensitivity in GNSS UEs will be investigated. By examining the designs of the GNSSs from the perspective of NAV messages, the common design of short validity and high update rate will be pointed out, which is of innate weakness and result into the common challenges in fast TTFF and high sensitivity on the standalone UEs for each GNSS.

Table 2-2 Comparison of NAV Messages among Existing GNSSs on Length of Validity Period

	GPS	GLONASS	GALILEO (E5a)	BDS (D1)

Data rate	50 bps	50 bps	50 bps	50 bps
Length of super-frame	750 s	150 s	600 s	720 s
Length of sub-frame	6 s	30 s	50 s	6 s
Length of Eph	18 s	10 s	~30 s	18 s
Repeat of Eph	30 s	30 s	50 s	30 s
Update of Eph	2 hrs	0.5 hr	3 hrs	1 hr
Orbit	Keplerian	Cartesian	Keplerian	Keplerian
Length of Validity Period	4 hrs	0.5 hr	4 hrs	2 hrs

As introduced in Table 2-2, for ephemeris, the primary content included in the NAV messages, one point in common for all the existing GNSSs (GPS, GLONASS, GALILEO and BDS) is that the length of validity period is short (0.5 hour to 4 hours), and is updated frequently (every 0.5 hour to 2 hours). Because of the short validity period for the ephemeris, the content of the NAV messages is changed every a short period of time (0.5 hour to 2 hours), and therefore the GNSS UEs need to download the NAV messages from the air every a short period of time, which brings the following challenges to sensitivity on GNSS UEs:

(1) NAV bits decoding

Bit Error Rate (BER) increases with the decrease of signal strength. The tolerance of BER is decided by the implemented code-words and parity algorithm in the NAV messages, corresponding to a limitation of ~27 dB-Hz in signal strength for successful GPS NAV message decoding (Braasch et al., 1999). The need for frequent NAV message decoding (every 2 hours

for GPS) results in higher risk of decoding failures due to change of signal conditions. In such cases, even if the GNSS signals are tracked and measurements are available, position fix still may not be possible because of the failure in ephemeris decoding.

(2) Acquisition and tracking

One of the primary techniques for improving acquisition and tracking sensitivity is through long coherent integration, which requires the data bits during the integration period to be wiped off first. It is often possible, if the NAV bits have ever been successfully decoded, or provided from external sources. However, when there are frequent changes in the NAV messages, long coherent integration would become challenging.

Therefore, some remedy technologies are necessary to obtain fast TTFF and high sensitivity in GNSS UEs, including AGNSS and especially EE.

2.6 Existing Technologies for Improving TTFF and Sensitivity

2.6.1 Assisted GNSS (AGNSS)

The existing efforts of improving the TTFF and sensitivity on GNSS UEs can be categorized into two different approaches: (1) developing assisting systems, (2) inventing and implementing new algorithms for standalone UEs.

The concept of assisted GPS (AGPS) was brought up by industry in late 1990's, when lots of patents were filed, and then granted in early 2000's (King 2001; Zadeh 2001). Seeing the challenges of TTFF and sensitivity on standalone GPS devices, the general idea from the patents is to provide assisting information to GNSS UEs, such as time, rough location, a list of satellites in view, Doppler of each satellite, ephemeris, etc., in a way to speed up each stage in the process of position fix (Figure 2-6). The AGPS system and infrastructure have been well described in Diggelen (2009). In a typical network-assisted AGPS system, a Serving Mobile Location Center (SMLC) is established to process the data (navigation messages) collected by the Wide Area Reference Network (WARN) and package the assistance data in Information Elements (IEs). Upon requests from GPS UEs, the SMLC sends out IEs through Mobile Switching Center (MSC) and Radio Network Controller (RNC), until finally arriving at the UEs. With a series of AGPS specifications embodied and continuously revised in the 3GPP and Open Mobile Alliance (OMA) standards since 2001 (3GPP, 2001), AGPS-enabled products then became popular in the GNSS market.

2.6.1.1 Protocols

As part of positioning protocols, the A-GPS protocols consist of protocols for Control Plane (C-Plane) and User Plane (U-Plane), which are defined by two different standardization bodies 3GPP and Open Mobile Alliance (OMA) respectively.

The C-Plane protocols are for private Signaling network and related infrastructure to provide wireless communication services to authorized subscribers, whereas the U-Plane protocols are

for the exchange of location data between user and location platform. As a key IP technology for Location-Based Services (LBS), the Secure User Plane Location (SUPL) protocols are deployed in the system that consisting of a server (network equipment stack) and a SUPL Enabled Terminal (SET) (SUPL enabled wireless handset). SUPL is operated as a separate network layer that requires minimal interaction with private signaling networks, using established data-bearing channels and positioning protocols for the exchange of location data between a SET and a SUPL Location Platform (SLP).

2.6.1.2 Operation Modes

The AGNSS is often operated in Mobile Station Assisted (MSA) or Mobile Station Based (MSB) modes.

In MSA mode, in order to help speed up the signal acquisition on Mobile Station (MS), usually the following Assistance Data (AD) is sent from the AGNSS server to the MS:

- Rough time and uncertainty
- Elevation/Azimuth of visible SVs
- Predicted code phases and uncertainties of visible SVs
- Predicted frequencies of visible SVs
- Navigation bits (possibly)

After the signals are acquired and tracked on the MS, measurements are sent from MS to the AGNSS Server for position computation. Depending on implementation, the measurements sent

to Server might be partial (code phases and carrier frequency) or full measurements (full PRs and and carrier frequency) of tracked signals. If partial measurements are sent to the AGNSS Server, the full measurements have to be recovered first, and then using the SV ephemerides obtained elsewhere, the position at the measurement instant is computed on the AGNSS server and then sent back to MS.

In MSB mode, in order to help speed up the signal acquisition on MS, the following AD may be sent from the AGNSS Server to MS:

- Rough time and uncertainty
- Rough location and uncertainty
- Ephemeris/Almanac, or EE

so that the MS is able to predict the list of visible SVs and the code phase and Doppler of each SV for fast signal acquisition with the AD. Once in stable signal tracking stage, depending on the uncertainties of assisted time and location, the Bit Sync or even the Frame Sync on MS could be skipped, and once full measurements are available, position is computed on the MS and sent to As if needed.

2.6.2 Ephemeris Extension (EE)

With additional assistance, the performance of fast TTFF and sensitivity on UEs can be further enhanced but it is a challenge when network connectivity is not available. A technology usually referred to as Ephemeris Extension (EE) was introduced by GlobalLocate (now a part of Broadcom) (LaMance, 2003), and SiRF (later a part of CSR) (Garin et al., 2008; Han 2009),

which enables fast TTFF and high sensitivity on GNSS UEs even without network connectivity. According to the introduction in Broadcom (2015) and Lundgren (2005) for the product Long Term Orbit (LTO), and in CSR (2015) for InstantFix, both are based on orbital determination theories, and provide alternative ephemeris with a validity period extended to a few days, rather than 2 hours for regular GPS ephemeris. As of today, a variety of EE products become available from many other companies and institutes (see Table 2-3), e.g. EASY/Hotstill/Embedded Predicted Orbit (EPO) from MediaTek (MediaTek, 2015), AssistNow from uBlox (uBlox, 2015), GPStream/PGPS from RxNetworks (RxNetworks 2015), and Embedded Autonomous Ephemeris Prediction (EAP) from JPL (JPL, 2015), etc.

Table 2-3 EE Products in GNSS Products

Company/Institute	EE Products
Broadcom	LTO (Long Term Orbit) • Server based
SiRF/CSR	InstantFix • SGEE: Server-Generated • CGEE: Client-Generated
MediaTek	EASY (Embedded Assist System) • Self-generated HotStill • Client based EPO (Extended Predict Orbit) • Server based
uBlox	AssistNow • Offline: Server based • Autonomous: Client based
RxN	GPStream PGPS • Autonomous • Connected
JPL	EAP (Embedded Autonomous Ephemeris Prediction)

2.6.3 Limitations of the Existing Assistance Technologies

In spite of the benefits on TTFF and sensitivity, the AGNSS and EE technologies have obvious limitations, as detailed in Table 2-4.

Building and maintaining the AGNSS infrastructure requires tremendous efforts and continuous cost. Any AGNSS-capable navigation devices, unlike standalone GNSS devices, are tied to good signals from the subscriber cellular phone networks in order to get assistance data on time, which substantially limit their area of operation.

The EE technologies consist of server based and client based modes. The client based EE is good for standalone devices, but the accuracy is subject to the validity of the embedded earth orientation parameters (EOP), and the quantity and quality of the local data collection. The server based EE is able to provide better accuracy, but it also needs the supports from the global infrastructure for data collection and is subject to network connectivity, see Lundgren (2005) for an example of such a system.

Table 2-4 Comparison of AGNSS and EE in improving TTFF and Sensitivity

	AGNSS	**EE**
Benefits	*TTFF* -Fast signal acquisition -Saving Eph downloading time *Sensitivity* -To provide Eph when local	**Server/Client Based** *TTFF* -Saving Eph downloading time *Sensitivity* -To provide Eph when local eph downloading

	Eph downloading is not possible	is not possible **Server Based** -Good accuracy **Client Based** -No reliance on network connectivity
Limitations	-Infrastructure development and maintenance -Reliance on network connectivity -Frequent data transferring over network; Subject to network bandwidth -Additional cost of data plan for end users -Transportation delay of assisting data -2 hour update rate	**Server Based** -Infrastructure development and maintenance -Reliance on network connectivity -Large data size to transfer over network; subject to network bandwidth -Additional cost of data plan for end users -Unusable for data aiding **Client Based** -Subject to quantity/quality of local data collection -Susceptible to orbit maneuver and clock adjustment -Subject to validity of local earth orientation parameters (EOP) -Unusable for data aiding

Table 2-4 clearly indicates that AGNSS and EE can only be beneficial under certain prerequisite conditions, such as network connectivity and data availability. Or in other words, even with the above technologies, fast TTFF and high sensitivity may still not be obtainable when those prerequisite conditions are not met, which does happen often in daily life. It is really a frustrating fact, since enormous efforts have been spent for years by the industry.

The fundamental cause, in the author's view, lies in the congenital weakness of the design of the existing GNSS NAV messages. Taking GPS as an example, the contents in GPS sub-frames 1-3 are updated every 2 hours, although the ephemeris is valid for up to 4 hours. It is challenging and

questionable for standalone GPS devices in weak signal environments to catch up with such frequent ephemeris updates. Working properly in the past 2 hours does not mean that the UE can work properly in the next 2 hours, if ephemerides are not downloaded in time. The NAV messages received 2 hours ago cannot be used for the data aiding in the next 2 hours to improve the tracking sensitivity. For startups under normal signal conditions, the UEs, if missing the start of sub-frame 1, have to wait 30 s to get to the next sub-frame 1 in order to download a complete copy of the ephemeris. The successful startups 4 hours ago do not help much reduce the TTFF in the subsequent startups, as time needs to be spent again on downloading the ephemeris.

To some extent, all the efforts of the above technologies actually just attempt to improve the TTFF and sensitivity by overcoming the limited life period (2-4 hours) of the broadcast ephemeris. The longer the life of the broadcast ephemeris, the less the need for the above technologies.

CHAPTER 3 GNSS EPHEMERIS WITH LONG VALIDITY PERIOD

This chapter is to focus on studying the feasibility of GNSS ephemeris with longer validity periods compared to what is deployed in existing GNSS, through a deep diving into the theory of satellite orbit determination, including different perturbation forces, solar radiation models, numerical integration, and estimation methods. Methods to obtain ephemeris with extended validity will be further studied and quantitatively analyzed.

3.1 Time Systems

Different time systems used in the satellite orbital determination are summarized in this section, along with the relationship among them.

3.1.1 Terrestrial Time (TT)

TT is the coordinate time on the Earth surface, and it is used in describing the Earth's precession and nutation.

3.1.2 International Atomic Time (TAI)

TAI is an international time standard in a continuous time scale that is not connected to the Earth's rotation but based on atomic clock, with the origin selected at 1 January 1958 when UT1

- TAI is almost 0. TT is a successor of Ephemeris Time (ET). Because of the historical difference between TAI and ET when TT was introduced, TT is ~32.184 s ahead of TAI.

3.1.3 Universal Time (UT1)

It is a time that is determined by the Earth rotation around the Earth pole, and counted as 0 at midnight, with the unit of duration being a solar day. UT0 is the rotational time for a particular place of observation, and UT1 is the rotation time with the effect of polar motion corrected in UT0. For UT1, the difference to UTC, namely ($UT1 - UTC$), is distributed by IERS Bulletins A and B (IERS, 2017).

3.1.4 Coordinated Universal Time (UTC)

Defined by ITU, UTC is actually based on TAI. Leap seconds (LS) are inserted at irregular intervals when necessary to keep UTC within 0.9 seconds of UT1, namely $|UT1 - UTC| < 0.9\ s$.

3.1.5 GPS Time (GPST)

Each GNSS has its own time system. For GPS, the GPST takes the same time scale of TAI, with the origin selected at the midnight of January 6, 1980. Similarly, for GLONASS, the time (GLOT) is aligned with UTC, with 3 hours offset ahead; for BDS, the origin of BDS Time (BDT) is selected at 00:00:00 UTC of January 1, 2006; and for GALILEO, the time system

(GST) started at middle night of 21 August 1999 and is synchronized with TAI with a nominal offset within 50 ns.

The relationships among the above time systems are summarized in the following Figure 3-8:

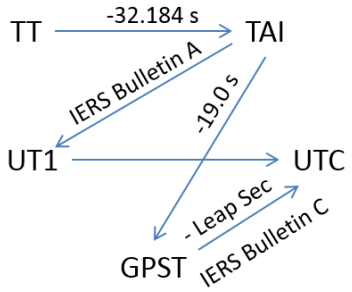


Figure 3-8 Relationship among Different Time Systems

3.2 Coordinate Frames

In ephemeris extension, the following coordinate frames are involved and satellite positions and velocities are frequently transformed from one to the others.

3.2.1 Earth Center Earth Fixed (ECEF) Frame

This system is listed here for completeness, with the detailed definition skipped considering that it is well known to Geomatics people. In ephemeris extension, the Earth gravity model, satellite position and velocity obtained from broadcast ephemeris are all expressed in ECEF.

3.2.2 Earth Center Inertial (ECI) Frame

The ECI frame is centered at the geocentre and not rotating with respect to distant galaxies, which is defined using the Earth's orbit plane and the orientation of the Earth's rotational axis in space. The J2000, a commonly used ECI frame, is used in this thesis for orbital determination related work. The ECI is where the numerical integration of orbit happens. For this purpose, the coordinates calculated for the Sun, the Moon, the satellites, and all the force models need to be transformed into ECI, so that the finally derived satellite accelerations can be expressed in ECI and then sent to numerical integration.

In ephemeris extension, the conversions between ECEF and ECI for the satellite position and velocity are often needed. Usually the transition matrix from ECEF to ECI consists of the following components:

$$\mathbf{C}_{ECEF}^{ECI} = \mathbf{W} \cdot \mathbf{S} \cdot \mathbf{N} \cdot \mathbf{P} \tag{3-1}$$

where

W is wobble matrix, describing the Earth's polar motion

S is rotation matrix, around the axis of rotation of the Earth

N is nutation matrix, relating the true equinox and mean equinox

P is precession matrix, relating the true equinox and mean equinox

with the details of each equation are fully explained in Ash (1972) and Montenbruck (2000).

3.2.3 Satellite Body Frames

The satellite body frame is needed when modeling some non-gravitational perturbing forces that act on the satellite, such as the solar radiation pressure. In order to define the satellite body frames, the following basic vectors are needed:

$\mathbf{r}$ the satellite's position vector, expressed in ECI

$\dot{\mathbf{r}}$ the satellite's velocity vector, expressed in ECI

$\mathbf{r_s}$ the Sun's position vector, expressed in ECI

And some additional unit vectors are further defined in Table 3-5:

Table 3-5 Unit Vectors related to Satellite Body Frame

Vector	Notes
$\mathbf{e}_D = \frac{\mathbf{r}-\mathbf{r_s}}{\lvert\mathbf{r}-\mathbf{r_s}\rvert}$	Unit vector pointing from the Sun to the satellite, expected to be perpendicular to the surface of the satellite's solar panel
$\mathbf{e}_Z = -\frac{\mathbf{r}}{\lvert\mathbf{r}\rvert}$	Unit vector pointing from the satellite to the Earth center
$\mathbf{e}_R = -\mathbf{e}_Z$	Unit vector along satellite radial direction, positive from earth center to the satellite
$\mathbf{e}_V = \frac{\dot{\mathbf{r}}}{\lvert\dot{\mathbf{r}}\rvert}$	Unit vector of the satellite's velocity
$\mathbf{e}_T = \mathbf{e}_N \times \mathbf{e}_R$ $= \mathbf{e}_V$	Unit vector tangent to the satellite orbit
$\mathbf{e}_H = \mathbf{e}_V \times \mathbf{e}_Z$	Unit vector perpendicular (normal vector) to the orbital plane
$\mathbf{e}_N = \mathbf{e}_R \times \mathbf{e}_V$ $= \mathbf{e}_H$	Unit vector perpendicular (normal vector) to the orbital plane
$\mathbf{e}_Y = \mathbf{e}_D \times \mathbf{e}_Z$	Unit vector along the axis of the satellite's solar panel, regarded as the rotation axis of the solar panel
$\mathbf{e}_B = \mathbf{e}_D \times \mathbf{e}_Y$	Unit vector in the solar panel, perpendicular to the Y axis
$\mathbf{e}_X = -\mathbf{e}_Z \times \mathbf{e}_Y$	Unit vector perpendicular to the Y-Z plane

Using different sets of above vectors, a few different satellite body frames can be established. For different solar radiation models, each is realized in a specific body frame.

3.2.3.1 Origin of Satellite Body Frame

The satellite measurements, and positions calculated from the broadcast ephemeris (BE) are referred to the satellite's antenna phase center (APC). However, the satellite positions from precise orbits are referred to the satellite mass center (MC), as the numerical integration involved in the satellite orbital determination is always with respect to the satellite MC. Therefore, the APC offset vector with respect to the MC should be accounted for when satellite body frames with different origins are involved.

The satellite APC is usually calibrated and provided by the satellite manufacturer, and then continuously monitored and estimated by National Geospatial-Intelligence Agency (NGA). The details of estimated APC for each GPS satellite is introduced in NGA (2013), and the process of how APC is modeled for GPS IIF-1 is explained in Dilssner (2010).

Per Sanz et al. (2011), the relation between the satellite APC and MC can be expressed as:

$$\mathbf{r}^{APC} = \mathbf{r}^{MC} + \mathbf{C}_{MC}^{APC} \cdot \Delta_{APC} \tag{3-2}$$

Where:

$\mathbf{r}^{APC}$ represents the position of satellite APC

$\mathbf{r}^{MC}$ represents the position of satellite MC

Δ_{APC} represents the vector of satellite APC offset w.r.t. the MC

$\mathbf{C}_{MC}^{APC}$ a transition matrix from the satellite MC to APC

And the relation is further illustrated in Figure 3-9, in which the Z axis is pointing from the satellite MC to the Earth center, Y axis is along the satellite solar panel crossing the satellite MC, and X axis is perpendicular to Y and Z axes to complete a right handed system. In this chapter, the satellite positions by default are referred to the MC unless otherwise specified.

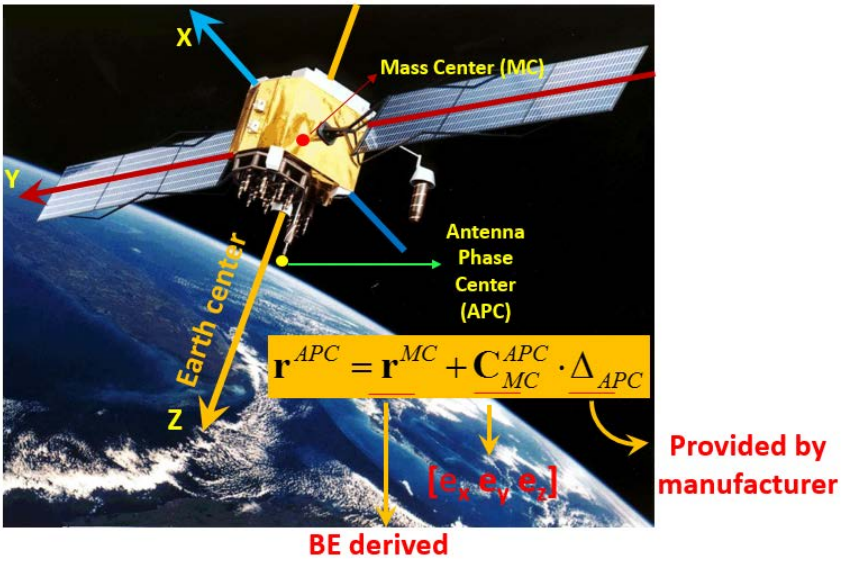


Figure 3-9 Relation between Satellite MC and APC

(Backgroud image: http://www.circuitstoday.com/wp-content/uploads/2011/12/Lockheed-Martin-Block-III-GPS-Satellite.jpg)

3.2.3.2 Y-D-B Frame

With the help of the 3 orthogonal unit vectors $\mathbf{e}_Y$, $\mathbf{e}_D$ and $\mathbf{e}_B$, the Y-D-B body fame can be established in Figure 3-10, in which all the vectors defined in Table 3-5 are illustrated, with the axes Y, D and B highlighted through line thickness.

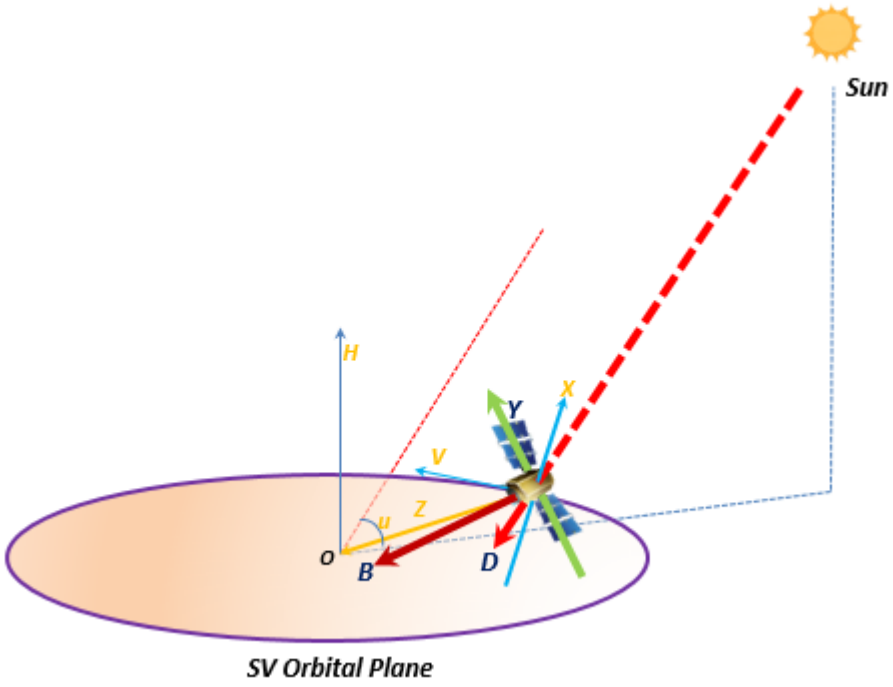


Figure 3-10 Illustration of Satellite Y-D-B body frame

3.2.3.3 X-Y-Z Frame

With the help of the 3 orthogonal vectors $\mathbf{e}_X$, $\mathbf{e}_Y$ and $\mathbf{e}_Z$, the X-Y-Z fame can be established in Figure 3-11 with the axes X, Y and Z highlighted through line thickness, which is actually the same to the body frame used in Figure 3-9. The reason to reintroduce this frame is to put it under the context of other unit vectors.

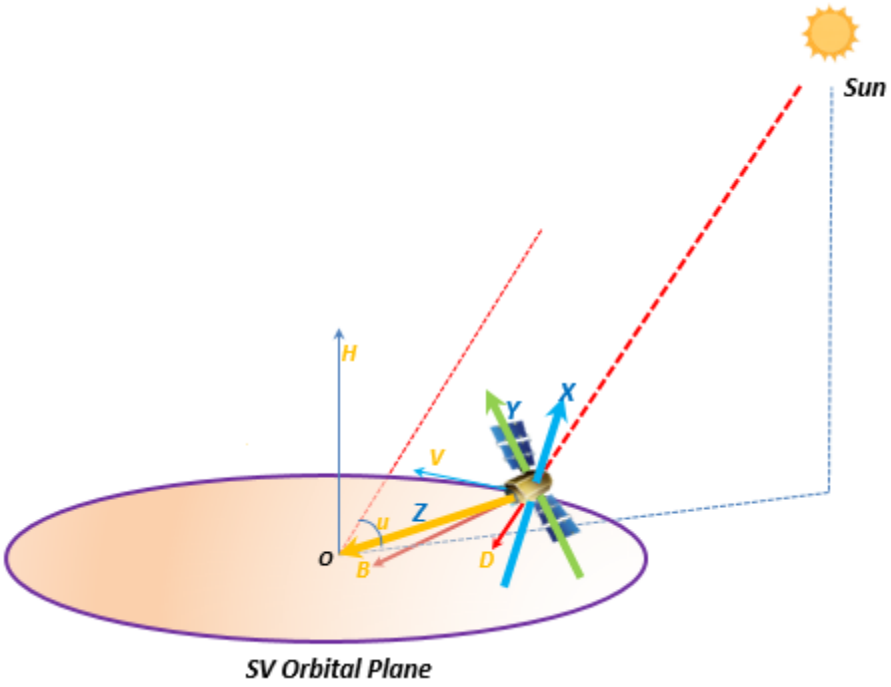


Figure 3-11 Illustration of SV body frame X-Y-Z

3.2.4 Radial-Tangent-Normal (RTN) Frame

This frame is usually defined in ECI, and is satellite position dependent, see the illustration plot below. At satellite position P_0, the RTN frame is defined in Figure 3-12 by the three vectors $\mathbf{e}_R$, $\mathbf{e}_R$ and $\mathbf{e}_R$ that are defined in Table 3-5.

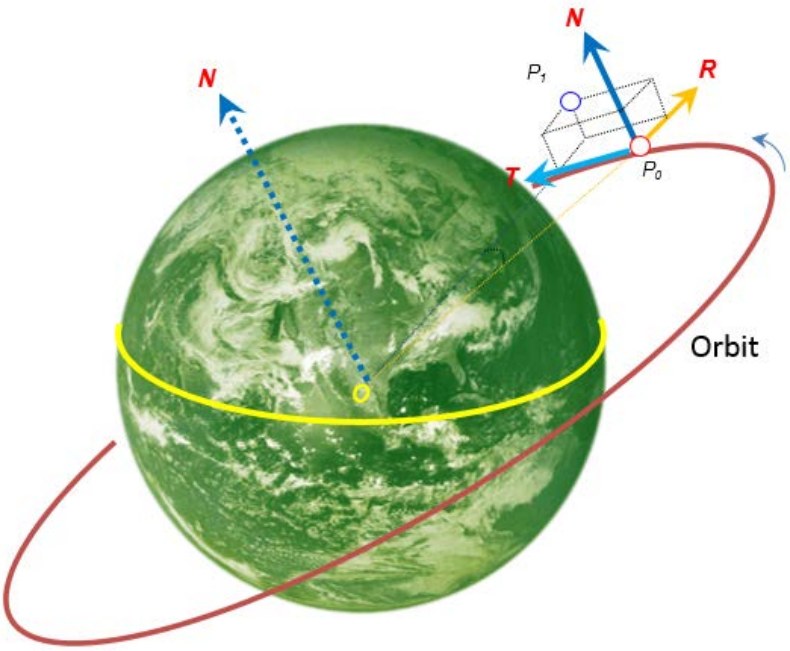


Figure 3-12 Illustration of R-T-N Frame

The RTN frame is helpful in analyzing orbital errors. So given orbital errors expressed in ECI, it is often necessary to transform to RTN frame with the help of a transition matrix:

$$\mathbf{C}_{ECI}^{RTN} = \begin{bmatrix} \mathbf{e}_R & \mathbf{e}_T & \mathbf{e}_N \end{bmatrix}^T \tag{3-3}$$

3.3 Perturbation forces acting on a satellite

3.3.1 Overall

Different perturbation forces that act on a satellite can be categorized as gravitational and non-gravitational forces, and are illustrated in Figure 3-13. In order to give an idea of how each

perturbation force affects the satellite motion, the typical magnitude of each force is further given in Table 3-6.

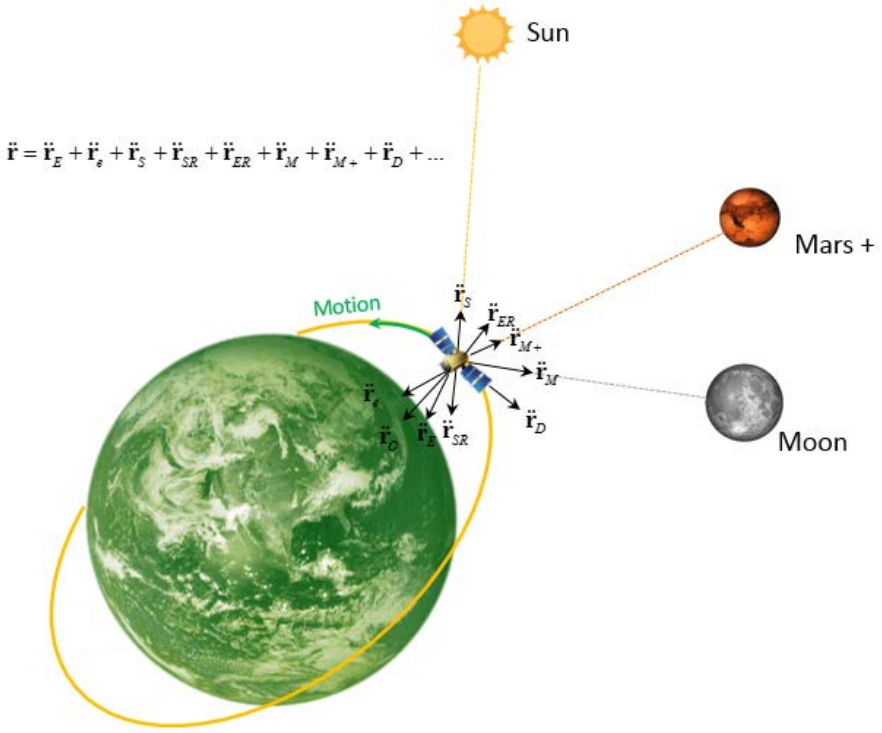


Figure 3-13 Illustration of Perturbation Forces Acting on a Satellite in Space

Table 3-6 Summary of Perturbation Forces Acting on a GPS Satellite in Space

Forces	**Magnitude (m/s^2)**	**Typical Errors on orbit (24 hrs)**
Earth central force ($\ddot{\mathbf{r}}_O$)	1×10^{-1}	3×10^5 km
Earth zonal harmonics ($\ddot{\mathbf{r}}_E$)	1×10^{-5}	37 km
Earth tesseral harmonics ($\ddot{\mathbf{r}}_\varepsilon$)	3×10^{-7}	1 km
Lunar gravitation ($\ddot{\mathbf{r}}_M$)	2.8×10^{-6}	10 km
Solar gravitation ($\ddot{\mathbf{r}}_S$)	1×10^{-6}	3 km

Additional planets($\ddot{\mathbf{r}}_{M+}$)	$<1 \times 10^{-14}$	~0 m
Tide ($\ddot{\mathbf{r}}_{T}$)	1×10^{-10}	0.4 m
Solar radiation ($\ddot{\mathbf{r}}_{SR}$)	1×10^{-7}	400 m
Earth radiation ($\ddot{\mathbf{r}}_{ER}$)	1×10^{-9}	4 m
Air drag ($\ddot{\mathbf{r}}_{D}$)	Negligible at GPS satellite altitude	
Thrust ($\ddot{\mathbf{r}}_{t}$)	Not considered in this thesis	
Satellite transmit radiation pressure	1 ~ 27 mm per orbit cycle Not considered in this thesis	

It should be noted that, for the perturbation from solar radiation in Table 3-6, although the magnitude is not large among other perturbation forces, it is the most difficult to handle. The solar radiation perturbation is dependent on not only solar flux (as a function of the distance from the Sun to the satellite), but also the area, orientation and reflectivity of the solar panel, which are different from satellite to satellite, change with the aging of solar panel, and should be always estimated.

3.3.2 Gravitational Forces

3.3.2.1 Point-mass perturbations

The perturbations of the Sun, the Moon, and additional planets are modeled as point-mass in the ECI frame:

$$\ddot{\mathbf{r}} = -GM_* \cdot \left(\frac{\mathbf{r} - \mathbf{r}_{s_*}}{\left| \mathbf{r} - \mathbf{r}_{s_*} \right|^3} + \frac{\mathbf{r}_{s_*}}{\left| \mathbf{r}_{s_*} \right|^3} \right) \tag{3-4}$$

Where

G represents the gravitational constant

M_* represents the mass of the Sun, the Moon, or other planets

$\mathbf{r}_{s_*}$ represents the coordinates of the Sun, the Moon, or other planets in ECI

And the partial derivatives (partials) of the above acceleration w.r.t. the satellite position **r** are:

$$\frac{\partial \ddot{\mathbf{r}}}{\partial \mathbf{r}} = -GM_* \cdot \left(\frac{1}{\left| \mathbf{r} - \mathbf{r}_{s_*} \right|^3} \cdot \mathbf{I}_{3\times 3} - 3(\mathbf{r} - \mathbf{r}_{s_*}) \cdot \frac{(\mathbf{r} - \mathbf{r}_{s_*})^T}{\left| \mathbf{r} - \mathbf{r}_{s_*} \right|^3} \right) \tag{3-5}$$

For the calculation of the coordinates of the Sun, the Moon and other planets in ECI, the Jet Propulsion Laboratory (JPL) provides a series of solar system ephemerides in the form of Chebyshev approximations. A version of the Development Ephemerides (DE), DE405 is widely adopted.

3.3.2.2 The Geopotential

Assuming the total mass of the Earth is concentrated in the Earth center, the Earth's gravitational force acting on a satellite is:

$$\ddot{\mathbf{r}} = -GM_E \frac{\mathbf{r}}{\left| \mathbf{r} \right|^3} \tag{3-6}$$

where M_E stands for the mass of the Earth.

However, considering the fact that the Earth mass is actually irregularly distributed, the Earth's gravitational force is usually modelled as the sum of contributions from many different individual mass elements. As the integral of the individual mass elements, the total Earth mass is:

$$U = G\int \frac{\rho(p)d^3\mathbf{p}}{|\mathbf{r}-\mathbf{p}|^3} \tag{3-7}$$

where $\rho(p)$ represents the mass density at point $\mathbf{p}$ inside of the Earth, and $|\mathbf{r}-\mathbf{p}|$ is the satellite's distance to this place.

To calculate the above integral, the inverse of $|\mathbf{r}-\mathbf{p}|$ can be expanded in a series of Legendre polynomials $\mathbf{P}_n(u)$ – degree of n, and further expressed with the help of an associated Legendre polynomial $\mathbf{P}_{nm}(u)$ of degree **n** and order **m**, so the Earth's gravity potential can be expressed as:

$$U = \frac{GM_E}{r}\sum_{n=0}^{\infty}\sum_{m=0}^{n}\frac{R_E^n}{r^n}P_{nm}(\sin\phi)\cdot\left(C_{nm}\cos(m\lambda)+S_{nm}\sin(m\lambda)\right) \tag{3-8}$$

Where

λ / ϕ represent longitude/latitude of the mass point r

C_{nm} / S_{nm} represent the geopotential coefficients that are provided by different gravity models, and can be normalized and noted as $\overline{C}_{nm} / \overline{S}_{nm}$

P_{nm} represent the associated Legendre polynomial, and can be normalized and noted as $\overline{P}_{nm}$

For different **n** and **m**, the geopotential coefficients can be categorized into:

(1) Zonal coefficients (m=0)

These coefficients describe part of the potential that does not depend on longitude. The notation $J_n = -C_{n0}$ is commonly used for zonal terms as S_{n0} all vanish. A two-body potential is actually only the 1st term in expansion of the Earth's potential (with only C_{00}); and the point-mass potential is a two-body potential, with the mass density not varying with longitude and latitude.

For zontal terms (C_{n0}), the mass distribution considered symmetric w.r.t. the rotation axis for an oblate rotational ellipsoid, with $J = -\frac{3}{2}C_{20}$ being the largest geo-potential coefficient aside from C_{00}.

(2) Tesseral coefficients (m<n)

(3) Sectorial coefficients (m=n)

The Tesseral and Sectorial coefficients are used to consider the irregularities in earth ellipsoid.

The partial derivatives of the geopotential w.r.t. the satellite position **r** are often part of interest:

$$\frac{\partial \ddot{\mathbf{r}}}{\partial \mathbf{r}} = \sum_{n,m} \frac{\partial \ddot{\mathbf{r}}_{nm}}{\partial \mathbf{r}} \tag{3-9}$$

And the full details of the above expression can be found in Cunningham (1970).

Earth gravity models (EGM) of different degrees and orders are available from NOAA, among which IGS92, EGM96 and EGM08 have been used in GNSS orbital determination (Herring et al., 2010).

The selection of EGM and the degrees and orders, decides the accuracy of calculated acceleration due to the Earth. In the ephemeris extension, usually gravity model of degree and order 8 ~ 10 is adopted, depending on specific requirements on accuracy, run time and memory consumption.

3.3.3 Non-gravitational Forces

The non-gravitational forces that act on a satellite usually consist of:

(1) Solar radiation
(2) Earth radiation
(3) Air drag
(4) Thrust
(5) Satellite transmit radiation pressure

For air drag at the GNSS orbital altitude, usually it is negligible. For thrust, it only occur when orbital maneuvers are conducted, which is once or twice a years, as notified by NANU. For satellite transmit radiation pressure, the impact is just 1 ~ 27 mm per orbit cycle (Steigenberger et al., 2018), and therefore ignored in the research of this thesis. According to the study in Knocke (1989) and Rodriguez et al. (2011), the influence of the earth radiation on GNSS

satellites needs to be considered only for precise orbital determination. So the primary non-gravitational forces acting on the GNSS satellite is solar radiation.

3.3.3.1 Cannon Ball (CB) Model

As the simplest solar radiation model, the satellite shape is simplified as a Cannon ball and the satellite acceleration due to solar radiation pressure is modeled as:

$$\ddot{\mathbf{r}}_{SR} = -P_{SR} \cdot C_R \cdot \frac{A}{m} \cdot \frac{\mathbf{r} - \mathbf{r}_s}{\left|\mathbf{r} - \mathbf{r}_s\right|^3} \cdot a_u^2 \tag{3-10}$$

Where:

a_u is astronomical unit (IAU 1976: 149597870000.0 m)

A is cross-section area (m^2)

C_R is a Solar Radiation Parameter (SRP) – the radiation pressure coefficient, related to reflectivity of solar panel surface and to be estimated

m is mass of the satellite

$\mathbf{r}$ is satellite position, expressed in ECI

$\mathbf{r}_s$ is the Sun's position, expressed in ECI

$\ddot{\mathbf{r}}_{SR}$ is the satellite acceleration due to the solar radiation pressure

P_{SR} is solar radiation pressure on a unit area ($4.56 \cdot 10^{-6}\ N/m^2$)

And it is assumed that the solar panel surface is perpendicular to the solar flux.

For the sake of convenience to derive the partial derivatives of $\ddot{\mathbf{r}}_{SR}$ in this section and those in the subsequent sections, the derivatives for those unit vectors that are defined in Table 3-5 are prepared in the following Table:

Table 3-7 Derivatives of Defined Unit Vectors

Vector	Derivative
$\mathbf{e}_D = \frac{\mathbf{r}-\mathbf{r}_s}{\lvert\mathbf{r}-\mathbf{r}_s\rvert}$	$\frac{\partial \mathbf{e}_D}{\partial \mathbf{r}} = \frac{1}{\lvert\mathbf{r}-\mathbf{r}_s\rvert}\cdot\mathbf{I}_{3\times3} - (\mathbf{r}-\mathbf{r}_s)\cdot\frac{(\mathbf{r}-\mathbf{r}_s)^T}{\lvert\mathbf{r}-\mathbf{r}_s\rvert^3}$
$\mathbf{e}_Z = -\frac{\mathbf{r}}{\lvert\mathbf{r}\rvert}$	$\frac{\partial \mathbf{e}_Z}{\partial \mathbf{r}} = -\frac{1}{\lvert\mathbf{r}\rvert}\cdot\mathbf{I}_{3\times3} + \mathbf{r}\cdot\frac{\mathbf{r}^T}{\lvert\mathbf{r}\rvert^3}$
$\mathbf{e}_R = -\mathbf{e}_Z$	$\frac{\partial \mathbf{e}_R}{\partial \mathbf{r}} = -\frac{\partial \mathbf{e}_Z}{\partial \mathbf{r}}$
$\mathbf{e}_Y = \mathbf{e}_D \times \mathbf{e}_Z$	$\frac{\partial \mathbf{e}_Y}{\partial \mathbf{r}} = \frac{\partial \mathbf{e}_D}{\partial \mathbf{r}}\times\mathbf{e}_Z + \mathbf{e}_D \times \frac{\partial \mathbf{e}_Z}{\partial \mathbf{r}}$
$\mathbf{e}_B = \mathbf{e}_D \times \mathbf{e}_Y$	$\frac{\partial \mathbf{e}_B}{\partial \mathbf{r}} = \frac{\partial \mathbf{e}_D}{\partial \mathbf{r}}\times\mathbf{e}_Y + \mathbf{e}_D \times \frac{\partial \mathbf{e}_Y}{\partial \mathbf{r}}$
$\mathbf{e}_X = -\mathbf{e}_Z \times \mathbf{e}_Y$	$\frac{\partial \mathbf{e}_X}{\partial \mathbf{r}} = -\frac{\partial \mathbf{e}_Z}{\partial \mathbf{r}}\times\mathbf{e}_Y - \mathbf{e}_Z \times \frac{\partial \mathbf{e}_Y}{\partial \mathbf{r}}$

Defining:

$$D_0 = P_{SR}\cdot\frac{A}{m}, \tag{3-11}$$

and making use of Table 3-5, Equation (3-10) then becomes:

$$\ddot{\mathbf{r}}_{SR} = -C_R\cdot D_0\cdot\frac{a_u^2}{\lvert\mathbf{r}-\mathbf{r}_s\rvert^2}\cdot\mathbf{e}_D \tag{3-12}$$

Further define:

$$scale = \frac{a_u^2}{\left|\mathbf{r} - \mathbf{r_s}\right|^2} \tag{3-13}$$

So the partial derivatives of $\ddot{\mathbf{r}}_{SR}$ w.r.t. the satellite position **r** are:

$$\begin{aligned}\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{r}} &= C_R \cdot D_0 \cdot a_u^2 \cdot \left(\frac{\partial\left(\frac{1}{\left|\mathbf{r} - \mathbf{r_s}\right|^2}\right)}{\partial \mathbf{r}} \cdot \mathbf{e}_D + \frac{1}{\left|\mathbf{r} - \mathbf{r_s}\right|^2} \cdot \frac{\partial \mathbf{e}_D}{\partial \mathbf{r}} \right) \\ &= scale \cdot C_R \cdot D_0 \cdot \left(\frac{1}{\left|\mathbf{r} - \mathbf{r}_s\right|} \cdot \mathbf{I}_{3\times 3} - 3(\mathbf{r} - \mathbf{r}_s) \cdot \frac{(\mathbf{r} - \mathbf{r}_s)^T}{\left|\mathbf{r} - \mathbf{r}_s\right|^3} \right)\end{aligned} \tag{3-14}$$

Considering that solar radiation is not dependent on satellite velocity, there is:

$$\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \dot{\mathbf{r}}} = \mathbf{0}_{3\times 3} \tag{3-15}$$

Then the partial derivatives of $\ddot{\mathbf{r}}_{SR}$ w.r.t. SRP is:

$$\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{p}} = \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial C_R} = -scale \cdot D_0 \cdot \mathbf{e}_D \tag{3-16}$$

3.3.3.2 Cannon Ball + Y Bias (CBY) Model

Because the mass distribution of the satellite may be asymmetrical w.r.t. the axes of the satellite body frame, torques may be incurred due to solar radiation pressure. Some solar radiation models take into account this factor, and attempt to model the influences along different axes as

biases. The Y bias modeled here is along Y axis due to the torque incurred by solar radiation pressure. Based on the CB model, satellite acceleration due to solar radiation pressure is modeled as:

$$\ddot{\mathbf{r}}_{SR} = -C_R \cdot D_0 \cdot \frac{a_u^2}{\left|\mathbf{r} - \mathbf{r}_s\right|^2} \cdot \mathbf{e}_D + C_{YB} \cdot \mathbf{e}_Y \tag{3-17}$$

where C_{YB} is a Y bias coefficient that has to be estimated. So the partial derivatives $\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{r}}$ and $\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \dot{\mathbf{r}}}$ are the same to Equations (3-14) and (3-15), with additional partial derivatives w.r.t. to SRPs derived as:

$$\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{p}} = \left[\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial C_R} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial C_{YB}} \right] = \left[-scale \cdot D_0 \cdot \mathbf{e}_D \quad \mathbf{e}_Y \right] \tag{3-18}$$

3.3.3.3 SPHRC Model

According to the SPHRC model (Fliege et al., 1985), satellite acceleration due to solar radiation pressure is modeled as:

$$\ddot{\mathbf{r}}_{SR} = \frac{a_u^2}{\left|\mathbf{r} - \mathbf{r}_s\right|^2} \cdot D_0 \cdot \left(\lambda \cdot p_D \cdot \mathbf{e}_D + p_Y \cdot \mathbf{e}_Y + p_Z \cdot \mathbf{e}_Z \right) \tag{3-19}$$

Where

λ	is shadow factor (0: in shadow, 1: out of shadow)
D_0	is a satellite dependent constant, a function of solar panel surface area, satellite mass, and etc.
p_D	is solar radiation parameter along D axis, to be estimated

p_Y is solar radiation parameter along Y axis, to be estimated

p_Z is solar radiation parameter along Z axis, to be estimated

Reorganize Equation (3-19):

$$\ddot{\mathbf{r}}_{SR} = a_u^2 \cdot D_0 \cdot \left(\lambda \cdot p_D \cdot \frac{\mathbf{e}_D}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} + p_Y \cdot \frac{\mathbf{e}_Y}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} + p_Z \cdot \frac{\mathbf{e}_Z}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \right) \tag{3-20}$$

And take the partial derivatives w.r.t. the satellite position $\mathbf{r}$:

$$\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{r}} = (part_D + part_Y + part_Z) \tag{3-21}$$

where:

$$\begin{aligned} part_D &= a_u^2 \cdot D_0 \cdot \lambda \cdot p_D \cdot \left(\frac{1}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \cdot \frac{\partial \mathbf{e}_D}{\partial \mathbf{r}} - 2 \cdot \frac{\left(\mathbf{r}-\mathbf{r}_\mathbf{s}\right)^T}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^4} \cdot \mathbf{e}_D \right) \\ &= \frac{a_u^2}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \cdot D_0 \cdot \lambda \cdot p_D \cdot \left[\frac{\partial \mathbf{e}_D}{\partial \mathbf{r}} - 2 \cdot \frac{\left(\mathbf{r}-\mathbf{r}_\mathbf{s}\right)^T}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \cdot \mathbf{e}_D \right] \\ &= scale \cdot D_0 \cdot \lambda \cdot p_D \cdot \left[\frac{\partial \mathbf{e}_D}{\partial \mathbf{r}} - 2 \cdot \frac{\left(\mathbf{r}-\mathbf{r}_\mathbf{s}\right)^T}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \cdot \mathbf{e}_D \right] \end{aligned} \tag{3-22}$$

$$part_Y = scale \cdot D_0 \cdot p_Y \cdot \left[\frac{\partial \mathbf{e}_Y}{\partial \mathbf{r}} - 2 \cdot \frac{\left(\mathbf{r}-\mathbf{r}_\mathbf{s}\right)^T}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \cdot \mathbf{e}_Y \right] \tag{3-23}$$

$$part_Z = scale \cdot D_0 \cdot p_Z \cdot \left[\frac{\partial \mathbf{e}_Z}{\partial \mathbf{r}} - 2 \cdot \frac{\left(\mathbf{r}-\mathbf{r}_\mathbf{s}\right)^T}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \cdot \mathbf{e}_Z \right] \tag{3-24}$$

Also take the partial derivatives w.r.t. the SRPs:

$$\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{p}} = \left[\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_D} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_Y} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_Z} \right]$$
$$= \frac{a_u^2}{\left|\mathbf{r}-\mathbf{s}\right|^2} \cdot D_0 \cdot \left[\lambda \cdot \mathbf{e}_D \quad \mathbf{e}_Y \quad \mathbf{e}_Z \right] \tag{3-25}$$
$$= scale \cdot D_0 \cdot \left[\lambda \cdot \mathbf{e}_D \quad \mathbf{e}_Y \quad \mathbf{e}_Z \right]$$

3.3.3.4 SRDYZ Model

Satellite acceleration due to solar radiation pressure is modeled as:

$$\ddot{\mathbf{r}}_{SR} = \frac{a_u^2}{\left|\mathbf{r}-\mathbf{s}\right|^2} \cdot \left[D_0 \cdot \left(\lambda \cdot p_D \cdot \mathbf{e}_D + p_Y \cdot \mathbf{e}_Y + p_Z \cdot \mathbf{e}_Z \right) + \lambda \cdot \left(X(B) \cdot \mathbf{e}_X + Z(B) \cdot \mathbf{e}_Z \right) \right] \tag{3-26}$$

Where B is the angular distance between the center of the Earth and that of the Sun, as viewed from the satellite, and $X(B)$ and $Z(B)$ are the periodic terms of the SRP in the directions $\mathbf{e}_X$ and $\mathbf{e}_Z$, taking different forms for GPS Block I, Block II/II-A and block II-R (Chen, 2005).

Reorganize Equation (3-26):

$$\ddot{\mathbf{r}}_{SR} = \frac{a_u^2}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \cdot \left[D_0 \cdot \left(\lambda \cdot p_D \cdot \mathbf{e}_D + p_Y \cdot \mathbf{e}_Y + p_Z \cdot \mathbf{e}_Z \right) + \lambda \cdot \left(X(B) \cdot \mathbf{e}_X + Z(B) \cdot \mathbf{e}_Z \right) \right]$$
$$= a_u^2 \cdot D_0 \cdot \left[\lambda \cdot p_D \cdot \frac{\mathbf{e}_D}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} + p_Y \cdot \frac{\mathbf{e}_Y}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} + \left(p_Z + \frac{\lambda}{D_0} Z(B) \right) \cdot \frac{\mathbf{e}_Z}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} + \frac{\lambda}{D_0} X(B) \cdot \frac{\mathbf{e}_X}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \right]$$

(3-27)

And take the partial derivatives w.r.t. the satellite position $\mathbf{r}$:

$$\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{r}} = (part_D + part_Y + part_Z + part_X) \tag{3-28}$$

where $part_D$ and $part_Y$ are available from Equations (3-22) and (3-23), and $part_Z$ and $part_X$ are as follows:

$$part_Z = scale \cdot \left[\lambda \cdot \frac{\partial Z(B)}{\partial \mathbf{r}} \cdot \mathbf{e}_Z + \left(p_Z \cdot D_0 + \lambda \cdot Z(B) \right) \cdot \left(\frac{\partial \mathbf{e}_Z}{\partial \mathbf{r}} - 2 \cdot \frac{\left(\mathbf{r} - \mathbf{r_s} \right)^T}{\left| \mathbf{r} - \mathbf{r_s} \right|^2} \cdot \mathbf{e}_Z \right) \right] \quad (3\text{-}29)$$

$$part_X = scale \cdot \lambda \cdot \left[\frac{\partial X(B)}{\partial \mathbf{r}} \cdot \mathbf{e}_X + X(B) \cdot \left(\frac{\partial \mathbf{e}_X}{\partial \mathbf{r}} - 2 \cdot \frac{\left(\mathbf{r} - \mathbf{r_s} \right)^T}{\left| \mathbf{r} - \mathbf{r_s} \right|^2} \cdot \mathbf{e}_X \right) \right] \quad (3\text{-}30)$$

And the partial derivatives w.r.t. SRP are:

$$\begin{aligned} \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{p}} &= \left[\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_D} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_Y} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_Z} \right] \\ &= scale \cdot D_0 \cdot \left[\lambda \cdot \mathbf{e}_D \quad \mathbf{e}_Y \quad \mathbf{e}_Z \right] \end{aligned} \quad (3\text{-}31)$$

3.3.3.5 SRXYZ Model

Satellite acceleration due to solar radiation pressure:

$$\ddot{\mathbf{r}}_{SR} = \frac{a_u^2}{\left| \mathbf{r} - \mathbf{r_s} \right|^2} \cdot \left(p_X \cdot X(B) \cdot \mathbf{e}_X + D_0 \cdot p_Y \cdot \mathbf{e}_Y + \lambda \cdot p_Z \cdot Z(B) \cdot \mathbf{e}_Z \right) \quad (3\text{-}32)$$

Where $X(B)$ and $Z(B)$ are the periodic terms of the SRP along the directions $\mathbf{e}_X$ and $\mathbf{e}_Z$, taking different forms for GPS Block I, Block II/II-A and block II-R, are also referred as T10, T20 and T30 (Fliegel and Gallini, 1996) respectively.

Reorganizing Equation (3-32) into:

$$\ddot{\mathbf{r}}_{SR} = a_u^2 \cdot \left(p_X \cdot X(B) \cdot \frac{\mathbf{e}_X}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} + D_0 \cdot p_Y \cdot \frac{\mathbf{e}_Y}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} + \lambda \cdot p_Z \cdot Z(B) \cdot \frac{\mathbf{e}_Z}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \right) \tag{3-33}$$

So the partial derivatives can be derived as:

$$\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{r}} = (part_X + part_Y + part_Z) \tag{3-34}$$

where $part_Y$ is the same to Equation (3-23) in Section 3.3.3.3, and $part_X$ and $part_Z$ are as follows:

$$part_X = scale \cdot p_X \cdot X(B) \cdot \left[\frac{\partial \mathbf{e}_X}{\partial \mathbf{r}} - 2 \cdot \frac{\left(\mathbf{r}-\mathbf{r}_\mathbf{s}\right)^T}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \cdot \mathbf{e}_X \right] \tag{3-35}$$

$$part_Z = scale \cdot \lambda \cdot p_Z \cdot Z(B) \cdot \left[\frac{\partial \mathbf{e}_Z}{\partial \mathbf{r}} - 2 \cdot \frac{\left(\mathbf{r}-\mathbf{r}_\mathbf{s}\right)^T}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \cdot \mathbf{e}_Z \right] \tag{3-36}$$

And the partials w.r.t. the SRPs are:

$$\begin{aligned} \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{p}} &= \left[\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_X} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_Y} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_Z} \right] \\ &= scale \cdot \left[X(B) \cdot \mathbf{e}_X \quad D_0 \cdot \mathbf{e}_Y \quad \lambda \cdot Z(B) \cdot \mathbf{e}_Z \right] \end{aligned} \tag{3-37}$$

3.3.3.6 SRDYB Model

Satellite acceleration due to solar radiation pressure:

$$\ddot{\mathbf{r}}_{SR} = \frac{a_u^2}{\left|\mathbf{r}-\mathbf{r}_\mathbf{s}\right|^2} \cdot D_0 \cdot \left(p_D \cdot \mathbf{e}_D + p_Y \cdot \mathbf{e}_Y + p_B \cdot \mathbf{e}_B \right) \tag{3-38}$$

So the partial derivatives can be derived as:

$$\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{r}} = (part_D + part_Y + part_B) \tag{3-39}$$

Where $part_Y$ is the same to Equation (3-23) in Section 3.3.3.3 and $part_D$ and $part_B$ are as follows:

$$part_D = scale \cdot D_0 \cdot p_D \cdot \left[\frac{\partial \mathbf{e}_D}{\partial \mathbf{r}} - 2 \cdot \frac{(\mathbf{r}-\mathbf{r_s})^T}{|\mathbf{r}-\mathbf{r_s}|^2} \cdot \mathbf{e}_D \right] \tag{3-40}$$

$$part_B = scale \cdot D_0 \cdot p_B \cdot \left[\frac{\partial \mathbf{e}_B}{\partial \mathbf{r}} - 2 \cdot \frac{(\mathbf{r}-\mathbf{r_s})^T}{|\mathbf{r}-\mathbf{r_s}|^2} \cdot \mathbf{e}_B \right] \tag{3-41}$$

And the partials w.r.t. SPRs are:

$$\begin{aligned} \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{p}} &= \left[\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_D} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_Y} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_B} \right] \\ &= scale \cdot D_0 \cdot \left[\mathbf{e}_D \quad \mathbf{e}_Y \quad \mathbf{e}_B \right] \end{aligned} \tag{3-42}$$

3.3.3.7 BERNE Model

The models BERNE, BERN1 and BERN2 (Beutler et al., 1994; Springer et al., 1998) are developed by Bern University based on the data of the Center of Orbit Determination in Europe (CODE) since 1992. Only the BERNE model is discussed here, and details of BERN1 and BERN2 can be found in the references. According to BENE model, the satellite acceleration due to solar radiation pressure can be modeled as:

$$\ddot{\mathbf{r}}_{SR} = \frac{a_u^2}{|\mathbf{r}-\mathbf{s}|^2} \cdot \left(D(u) \cdot \mathbf{e}_D + Y(u) \cdot \mathbf{e}_y + B(u) \cdot \mathbf{e}_B \right) \tag{3-43}$$

Where:

$$D(u) = D_0 \cdot \left[\lambda \cdot p_D + p_{DC} \cdot \cos u + p_{DS} \cdot \sin u \right] \tag{3-44}$$

$$Y(u) = D_0 \cdot \left[p_Y + p_{YC} \cdot \cos u + p_{YS} \cdot \sin u \right] \tag{3-45}$$

$$B(u) = D_0 \cdot \left[p_B + p_{BC} \cdot \cos u + p_{BS} \cdot \sin u \right] \tag{3-46}$$

And u is the angle between the satellite and node in the orbital plane, namely the argument of latitude, with u_0 being the angle between the Sun's projection onto the orbital plane and the satellite's node. Thus $u - u_0$ is the elongation of the satellite from the Sun's projection in the orbital plane.

Let the vector of SRPs be denoted as:

$$\mathbf{p} = [\, p_D \quad p_{DC} \quad p_{DS} \quad p_Y \quad p_{YC} \quad p_{YS} \quad p_B \quad p_{BC} \quad p_{BS} \,] \tag{3-47}$$

In order to derive the partial derivatives of $\ddot{\mathbf{r}}_{SR}$, let's also have:

$$\frac{\partial D(u)}{\partial \mathbf{r}} = D_0 \cdot [- p_{DC} \cdot \sin u + p_{DS} \cdot \cos u] \cdot \frac{\partial u}{\partial \mathbf{r}} \tag{3-48}$$

$$\frac{\partial Y(u)}{\partial \mathbf{r}} = D_0 \cdot [- p_{YC} \cdot \sin u + p_{YS} \cdot \cos u] \cdot \frac{\partial u}{\partial \mathbf{r}} \tag{3-49}$$

$$\frac{\partial B(u)}{\partial \mathbf{r}} = D_0 \cdot [- p_{BC} \cdot \sin u + p_{BS} \cdot \cos u] \cdot \frac{\partial u}{\partial \mathbf{r}} \tag{3-50}$$

Where $\frac{\partial u}{\partial \mathbf{r}}$ can be derived through the expression $u = a \tan(\frac{z}{\sqrt{x^2 + y^2}})$. So the partial derivatives of $\ddot{\mathbf{r}}_{SR}$ in Equation (3-43) can be derived as:

$$\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{r}} = (part_D + part_Y + part_B) \tag{3-51}$$

Where:

$$part_D = scale \cdot \left\{ D(u) \cdot \left[\frac{\partial \mathbf{e}_D}{\partial \mathbf{r}} - 2 \cdot \frac{(\mathbf{r}-\mathbf{r}_s)^T}{|\mathbf{r}-\mathbf{r}_s|^2} \cdot \mathbf{e}_D \right] + \frac{\partial D(u)}{\partial \mathbf{r}} \cdot \mathbf{e}_D \right\} \tag{3-52}$$

$$part_Y = scale \cdot \left\{ Y(u) \cdot \left[\frac{\partial \mathbf{e}_Y}{\partial \mathbf{r}} - 2 \cdot \frac{(\mathbf{r}-\mathbf{r}_s)^T}{|\mathbf{r}-\mathbf{r}_s|^2} \cdot \mathbf{e}_Y \right] + \frac{\partial Y(u)}{\partial \mathbf{r}} \cdot \mathbf{e}_Y \right\} \tag{3-53}$$

$$part_B = scale \cdot \left\{ B(u) \cdot \left[\frac{\partial \mathbf{e}_B}{\partial \mathbf{r}} - 2 \cdot \frac{(\mathbf{r}-\mathbf{r}_s)^T}{|\mathbf{r}-\mathbf{r}_s|^2} \cdot \mathbf{e}_B \right] + \frac{\partial B(u)}{\partial \mathbf{r}} \cdot \mathbf{e}_B \right\} \tag{3-54}$$

And the partial derivatives w.r.t. the SRPs are:

$$\begin{aligned} \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial \mathbf{p}} &= \left[\frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_D} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_{DC}} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_{DS}} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_Y} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_{YC}} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_S} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_B} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_{BC}} \quad \frac{\partial \ddot{\mathbf{r}}_{SR}}{\partial p_{BS}} \right] \\ &= scale \cdot D_0 \cdot \\ &\left[\lambda \cdot \mathbf{e}_D \quad \cos u \cdot \mathbf{e}_D \quad \sin u \cdot \mathbf{e}_D \quad \mathbf{e}_Y \quad \cos u \cdot \mathbf{e}_Y \quad \sin u \cdot \mathbf{e}_Y \quad \mathbf{e}_B \quad \cos u \cdot \mathbf{e}_B \quad \sin u \cdot \mathbf{e}_B \right] \end{aligned} \tag{3-55}$$

For each GPS satellite, the parameters of the BERNE model are estimated on a daily basis by SOPAC & CSRC, and are publically available in the form of g-file from ftp://garner.ucsd.edu/pub/gfiles.

3.3.4 Eclipse Impact

Eclipse refers to the Earth shadow on the satellite orbit, in which the aforementioned solar radiation is greatly reduced or even diminished. Taking GPS for example, a GPS satellite has the chance to cross the eclipse every about 6 months. A real example is given in Figure 3-14 for a GPS satellite in the year of 2007. For each time period when the GPS satellite has chance to cross the eclipse, the period lasts for about 50 days, and the satellite cross the eclipse twice a day,

with each time staying in the Eclipse for about 40 minutes. According to Table 3-6, the magnitude of the solar radiation is around $10^{-7}\ m/s^2$, so the impact of crossing eclipse on orbit could be roughly estimated as $\frac{1}{2}\times 10^{-7}\ m/s^2 \times (40\times 60s)^2 \approx 0.28(m)$. Therefore, the eclipse impact needs to be considered only for precise orbital determination, and will be neglected in the study of ephemeris extension in this thesis.

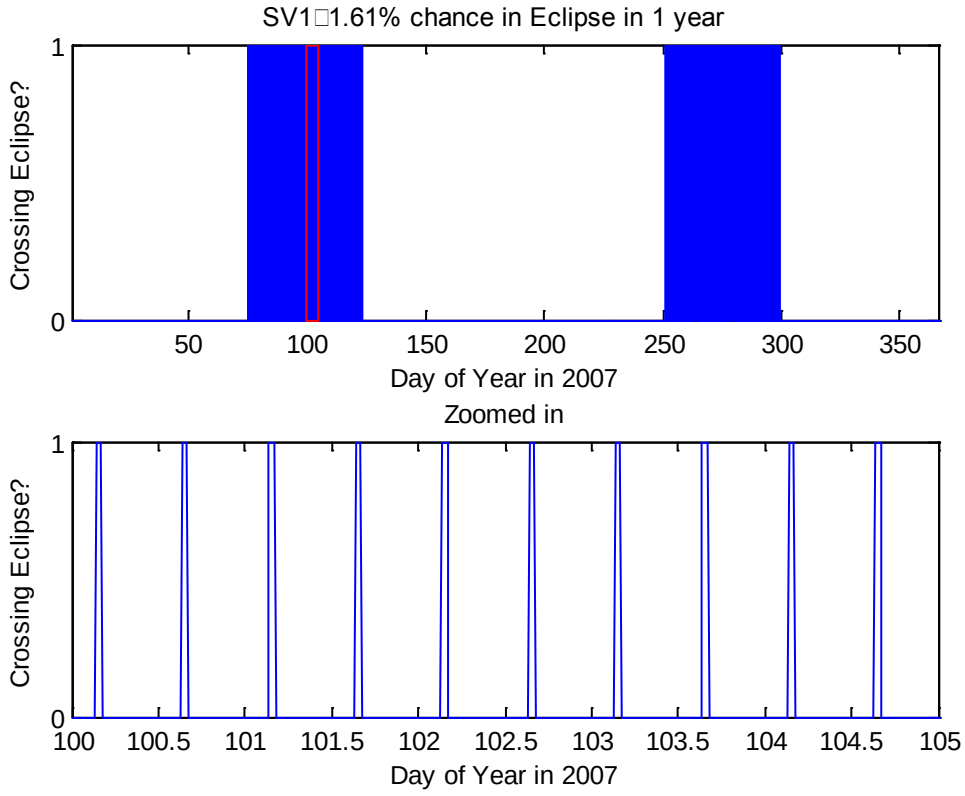


Figure 3-14 Chances for a GPS Satellite to Cross Eclipse

For the statistics from all the GPS satellites in the same year, the average chance of crossing the Eclipse is 1.22%, with the highest chance no more than 1.81%.

3.4 Numerical integration

Because of all the different perturbation forces acting on the GNSS satellites, analytical solutions to the differential equations of motion cannot suffice the GNSS position fix in accuracy, therefore we have to resort to numerical integration of GNSS satellite orbits.

In orbital determination, the desired state vector takes the form:

$$\mathbf{y} = \begin{pmatrix} \mathbf{r} \\ \dot{\mathbf{r}} \end{pmatrix} \tag{3-56}$$

With the differential equation to be solved taking the form:

$$\dot{\mathbf{y}} = f(t, \mathbf{y}) = \begin{pmatrix} \dot{\mathbf{r}} \\ \ddot{\mathbf{r}} \end{pmatrix} \tag{3-57}$$

Where

$$\ddot{\mathbf{r}} = a(t, \mathbf{r}, \dot{\mathbf{r}}) \tag{3-58}$$

Two numerical integrators that are often used in solving the above differential equations will be discussed here.

3.4.1 Runge-Kutta (RK) Integrator

Given the initial values of $\mathbf{y}$ at time $\mathbf{t_0}$,

$$\mathbf{y}(t_0)=\begin{pmatrix}\mathbf{r}_0\\ \dot{\mathbf{r}}_0\end{pmatrix} \tag{3-59}$$

the numerical solution of $\mathbf{y}$ at time $\mathbf{t_0}$+h can be obtained by numerically integrating the differential equation $\dot{\mathbf{y}}$ one step forward (h):

$$\mathbf{y}(t_0+h)\approx\mathbf{y}(t_0)+h\cdot\Phi \tag{3-60}$$

Where Φ is referred as ***increment function***. For a roughly approximate solution of $\mathbf{y}$, Φ can simply take slope of $\mathbf{y}$ at time $\mathbf{t_0}$, namely:

$$\Phi=\dot{\mathbf{y}}(t_0) \tag{3-61}$$

The RK integrator was introduced to obtain a precise solution, through which the ***increment function*** is calculated as the weighted mean of multiple slopes at different times between time $\mathbf{t_0}$ and $\mathbf{t_0}$+h. The number of slopes adopted here is referred as the integrator order, and the time step h is referred as the step size of the integrator.

The increment function of a general RK integrator can be expressed as:

$$\Phi=\sum_{i=1}^{n}b_i k_i \tag{3-62}$$

Where k_i is the slope evaluated through $\dot{\mathbf{y}}=f(t,\mathbf{y})$ at different time t_i and $\dot{\mathbf{y}}$ within the one step size h; and b_i is well selected weight in terms of certain rules (Montenbruck, 2000).

So a RK integrator consists of order *n* and a set of coefficients b_i ($i = 1,...,n$). The accuracy obtainable by a RK integrator is highly related to the order and selected step size. So numerical integration through RK method is a process of repeated calculations of increment function and one-step integration. Since the increment function is always calculated within the current integration step size, different integration steps are independent using RK method, therefore RK is often referred as a single-step method.

3.4.2 Adams-Moulton (AM) Integrator

In the calculation of increment function with RK method, each step of numerical integration involves multiple evaluations of slope through $\dot{\mathbf{y}} = f(t, \mathbf{y})$ within the integration step. A different way that is referred as multi-step method, calculates the increment function using the slopes evaluated and stored from previous steps, in an attempt to reduce the total number of function evaluations.

Adam-Bashforth (AB) and AM integrators are two of the multi-step integrators, whose increment functions are denoted with Φ_{AB_m} and Φ_{AM_m}, with m being the integrator order. Usually an AB method of order m is combined with an AM method of order m or m+1, and overall it is called a predictor-corrector (PECE) algorithm.

PECE algorithm consists of 4 steps:

(1) Predictor step – for initial estimation of solution at t_{i+1} through

$$\mathbf{y}_{i+1}^{p} = \mathbf{y}_i + h \cdot \Phi_{AB} \tag{3-63}$$

(2) Evaluation step – for a function value

$$f_{i+1}^{p} = f(t_{i+1}, \mathbf{y}_{i+1}^{p}) \tag{3-64}$$

(3) Corrector step – for an improved value

$$\mathbf{y}_{i+1} = \mathbf{y}_i + h \cdot \Phi_{AM} \cdot f_{i+1}^{p} \tag{3-65}$$

(4) Final Evaluation step – for the updated function value that is then used for start of next integration step

$$f_{i+1} = f(t_{i+1}, \mathbf{y}_{i+1}) \tag{3-66}$$

3.4.3 Numerical integration of satellite orbits

In the numerical integration of a satellite orbit, the initial conditions (ICs) are given. Taking the composition of aforementioned force models as the evaluation function $f(.)$, with a properly selected integration step size h, a RK integrator is usually used for the first a few steps of integration, until there are enough steps to start an AM integrator.

The selection of order and step size for the integrators directly relates to the performance of numerical integration, which is usually about the trade-off between accuracy and stability. For the PECE algorithm, smaller step size leads to better accuracy but less stability; Lower order ensures more stability even for large step-sizes, but results to less accuracy.

For the integrators used in ephemeris extension, a typical integration step size is 75 seconds, and RK and AM orders can be around 6 - 12, depending on different requirements on accuracy, run time and memory consumption.

3.5 Estimation of Initial Conditions (ICs)

3.5.1 Observations

The observations used in the study of this report are the satellite position $\mathbf{r}_k$ and/or velocity $\mathbf{v}_k$ at selected times $\mathbf{t_k}$, that are calculated from broadcast ephemeris (BE).

3.5.2 Derivatives at t_k

$$\dot{\mathbf{Y}}_k = \begin{bmatrix} \dot{\mathbf{r}}_k & \ddot{\mathbf{r}}_k & \mathbf{P}_{kk} \end{bmatrix} \tag{3-67}$$

$$\mathbf{P}_{kk} = \begin{bmatrix} \dfrac{d\dot{\mathbf{r}}_k}{d\mathbf{r}_k} & \dfrac{d\dot{\mathbf{r}}_k}{d\dot{\mathbf{r}}_k} & \dfrac{d\dot{\mathbf{r}}_k}{d\mathbf{p}_k} & \dfrac{d\ddot{\mathbf{r}}_k}{d\mathbf{r}_k} & \dfrac{d\ddot{\mathbf{r}}_k}{d\dot{\mathbf{r}}_k} & \dfrac{d\ddot{\mathbf{r}}_k}{d\mathbf{p}_k} \end{bmatrix} \tag{3-68}$$

Where the subscripts ‘k’ represents times $\mathbf{t_k}$, $\mathbf{P}_{kk}$ is often referred as partials in literatures (Ash, 1972; Montenbruck, 2000), $\mathbf{p_k}$ represents the parameters, such as solar radiation parameters, and $\mathbf{a_k}$ is the superposition of different acceleration components, including:

- Earth gravity
- Solar gravity
- Lunar gravity
- Solar Radiation
- Earth Radiation

- Tides
- And so on.

At the orbital altitude of GNSS satellite, air drag is negligible. Except for precise orbital determination, the influences of earth radiation and tides are usually not handled.

3.5.3 Time transition of Partials

For the partials $\mathbf{P}_{kk}$ that are directly available from the force models, they are w.r.t. the states at time $\mathbf{t_k}$. However, what is desired are the partials that are w.r.t. the states at time $\mathbf{t_0}$ – the time of initial conditions. In order to obtain the desired partials, let's take a look at the partials w.r.t. the neighboring time:

$$\mathbf{P}_{k,k-1} = \begin{bmatrix} \frac{d\dot{\mathbf{r}}_k}{d\mathbf{r}_{k-1}} & \frac{d\dot{\mathbf{r}}_k}{d\dot{\mathbf{r}}_{k-1}} & \frac{d\dot{\mathbf{r}}_k}{d\mathbf{p}_{k-1}} & \frac{d\ddot{\mathbf{r}}_k}{d\mathbf{r}_{k-1}} & \frac{d\ddot{\mathbf{r}}_k}{d\dot{\mathbf{r}}_{k-1}} & \frac{d\ddot{\mathbf{r}}_k}{d\mathbf{p}_{k-1}} \end{bmatrix} \tag{3-69}$$

Please note that Pk,k-1 is actually the single time integration of Pk-1,k-1.

Using concise expressions, the desired partials can be derived as below:

$$\begin{aligned} \mathbf{P}_{k,0} &= \frac{d(\dot{\mathbf{r}},\ddot{\mathbf{r}})_k}{d(\mathbf{r},\dot{\mathbf{r}},\mathbf{p})_0} \\ &= \frac{d(\dot{\mathbf{r}},\ddot{\mathbf{r}})_k}{d(\mathbf{r},\dot{\mathbf{r}},\mathbf{p})_k} \cdot \frac{d(\dot{\mathbf{r}},\ddot{\mathbf{r}},\mathbf{p})_k}{d(\mathbf{r},\dot{\mathbf{r}},\mathbf{p})_{k-1}} \cdot (\cdots) \cdot \frac{d(\dot{\mathbf{r}},\ddot{\mathbf{r}},\mathbf{p})_1}{d(\mathbf{r},\dot{\mathbf{r}},\mathbf{p})_0} \cdot \frac{d(\dot{\mathbf{r}},\ddot{\mathbf{r}},\mathbf{p})_0}{d(\mathbf{r},\dot{\mathbf{r}},\mathbf{p})_0} \end{aligned} \tag{3-70}$$

In detailed expressions again,

$$\mathbf{P}_{k,0} = \begin{bmatrix} \frac{d\dot{\mathbf{r}}_k}{d\mathbf{r}_0} & \frac{d\dot{\mathbf{r}}_k}{d\dot{\mathbf{r}}_0} & \frac{d\dot{\mathbf{r}}_k}{d\mathbf{p}_0} & \frac{d\ddot{\mathbf{r}}_k}{d\mathbf{r}_0} & \frac{d\ddot{\mathbf{r}}_k}{d\dot{\mathbf{r}}_0} & \frac{d\ddot{\mathbf{r}}_k}{d\mathbf{p}_0} \end{bmatrix} \tag{3-71}$$

$$\mathbf{P}_{0,0} = \begin{bmatrix} 0 & \mathbf{I} & 0 & \frac{d\ddot{\mathbf{r}}_0}{d\mathbf{r}_0} & \frac{d\ddot{\mathbf{r}}_0}{d\dot{\mathbf{r}}_0} & \frac{d\ddot{\mathbf{r}}_0}{d\mathbf{p}_0} \end{bmatrix} \tag{3-72}$$

In this way, the derivatives are transformed to:

$$\dot{\mathbf{Y}}_{k,0} = \begin{bmatrix} \dot{\mathbf{r}}_k & \ddot{\mathbf{r}}_k & \mathbf{P}_{k,0} \end{bmatrix} \tag{3-73}$$

3.5.4 State Vector

Taking the time integration on $\dot{\mathbf{Y}}_{k,0}$, the following equation can be obtained:

$$\mathbf{Y}_{k,0} = \begin{bmatrix} \mathbf{r}_k & \dot{\mathbf{r}}_k & \mathbf{A}_{k,0} \end{bmatrix} \tag{3-74}$$

Where

$$\mathbf{A}_{k,0} = \begin{bmatrix} \frac{d\mathbf{r}_k}{d\mathbf{r}_0} & \frac{d\mathbf{r}_k}{d\dot{\mathbf{r}}_0} & \frac{d\mathbf{r}_k}{d\mathbf{p}_0} & \frac{d\dot{\mathbf{r}}_k}{d\mathbf{r}_0} & \frac{d\dot{\mathbf{r}}_k}{d\dot{\mathbf{r}}_0} & \frac{d\dot{\mathbf{r}}_k}{d\mathbf{p}_0} \end{bmatrix} \tag{3-75}$$

Re-formatting $\mathbf{A_{k,0}}$, the following matrix $\mathbf{H_{k,0}}$ can be obtained, which builds the relationship between satellite position/velocity at time $\mathbf{t_k}$ with initial conditions.

$$\mathbf{H}_{k,0} = \begin{bmatrix} \dfrac{d\mathbf{r}_k}{d\mathbf{r}_0} & \dfrac{d\mathbf{r}_k}{d\dot{\mathbf{r}}_0} & \dfrac{d\mathbf{r}_k}{d\mathbf{p}_0} \\ \dfrac{d\dot{\mathbf{r}}_k}{d\mathbf{r}_0} & \dfrac{d\dot{\mathbf{r}}_k}{d\dot{\mathbf{r}}_0} & \dfrac{d\dot{\mathbf{r}}_k}{d\mathbf{p}_0} \end{bmatrix} \tag{3-76}$$

$$\mathbf{H}_{0,0} = \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times m} \\ \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times m} \end{bmatrix} \tag{3-77}$$

The dimension of the state vector in Equation (3-74) is 6 + 6 * (6 + m), where m is the number of solar radiation parameters represented by $\mathbf{p_0}$.

3.5.5 Observations and Observation Equation

$$\mathbf{l} = \mathbf{H} \cdot \mathbf{x} \tag{3-78}$$

The above equation can be further expanded as:

$$\begin{bmatrix} \mathbf{l}_1 \\ \mathbf{l}_2 \\ \cdots \\ \mathbf{l}_k \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{1,0} \\ \mathbf{H}_{2,0} \\ \cdots \\ \mathbf{H}_{k,0} \end{bmatrix} \cdot [\mathbf{x}] \tag{3-79}$$

Where $\mathbf{l}$ stands for the differences between observed orbits and numerically integrated orbits with some rough ICs, and $\mathbf{x}$ is the correction to the rough ICs.

3.5.6 Normal equation and WLS estimation of ICs

When integrating the orbit from $\mathbf{t_1}$ (namely reference time of ICs, the $\mathbf{t_{oe}}$) to time $\mathbf{t_k}$, the normal equation can be sequentially obtained through the following:

$$\mathbf{N} = \left[\mathbf{H}_{1,0}^{T} \cdot \mathbf{P}_{l1} \cdot \mathbf{H}_{1,0} + ... + \mathbf{H}_{k,0}^{T} \cdot \mathbf{P}_{lk} \cdot \mathbf{H}_{k,0}\right]^{-1} \tag{3-80}$$

And the same to the misclosure vector $\mathbf{L}$ at each observed point of orbit:

$$\mathbf{L} = \left[\mathbf{H}_{1,0}^{T} \cdot \mathbf{P}_{l1} \cdot \mathbf{l}_{1} + ... + \mathbf{H}_{k,0}^{T} \cdot \mathbf{P}_{lk} \cdot \mathbf{l}_{k}\right] \tag{3-81}$$

Once arriving at the last orbital observation, the correction to the rough initial condition can be obtained as follows:

$$\Delta\mathbf{x} = \mathbf{N} \cdot \mathbf{L} \tag{3-82}$$

Usually 2-3 iterations are needed for the correction to converge to selected level of error tolerance, say $1e^{-3}$. The following two figures illustrate the iteration processes in two cases, with the observation misclosure (namely the difference between observed satellite positions and numerically integrated positions) plotted with observation time for each iteration. Please note that in the given cases, the reference time is noted as '0', and the observations in the past 2.5-5 days are used in estimating the ICs at the reference time and therefore in each iteration backward integrations are adopted. As illustrated, the magnitude of the misclosure is quickly decreased in each new iteration, until the convergence happens.

Please note that, in ephemeris extension, the parameters that are estimated together with the ICs are usually some solar radiation parameters, and are always treated constant during the observation and extension periods.

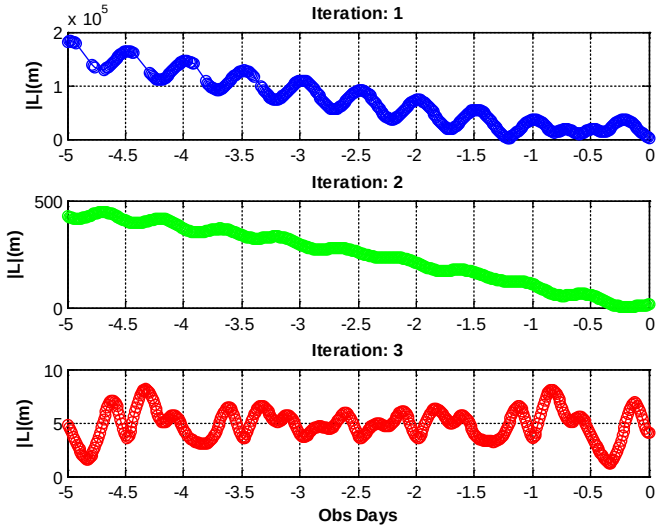


Figure 3-15 Illustration of Estimation Iterations - Case 1

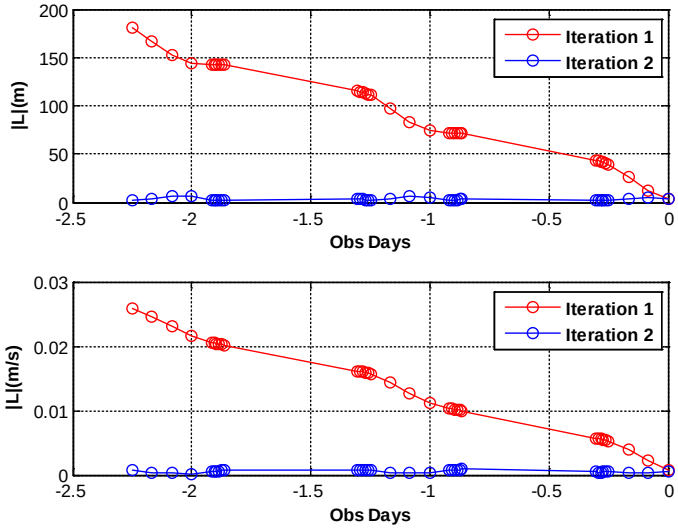


Figure 3-16 Illustration of Estimation Iterations - Case 2

3.6 Satellite Clock Modeling

Satellite clock modeling is actually the bottleneck of ephemeris extension for longer validity duration beyond 1 or 2 weeks. As the clocks are subject to stochastic processes, which can usually be categorized through Allan Variance analysis (Allan, 1987). When continuous and abundant clock measurements are available, Kalman Filter based clock modeling has been studied in lots of search, such as Barnes &Allan (1985), Breakiron (2001) and etc., which adopts different random process models to take care of the atomic clocks. However, in the study of this report, as the clock measurements can be neither continuous nor abundant, therefore only Least-Square based fitting is adopted to get the satellite clock coefficients (f_0, f_1, f_2), assuming a quadrant form of clock as the function of delta time w.r.t. clock reference time t_{oc}.

$$T_s(t) = f_0 + f_1 \cdot (t - t_{oc}) + f_2 \cdot (t - t_{oc})^2 \tag{3-83}$$

One thing important is that, the reference time t_{oc} of the clock modeling should be chosen to the same moment as the initial condition that has been estimated in Section 3.2, so that they can be used to obtain the extended ephemeris (including orbital and clock parts).

Given ***n*** clock observations $T_s(t_i)(i = 1,...,n)$, a design matrix can be derived according to above clock model:

$$\mathbf{H} = \begin{bmatrix} 1 & \Delta t_1 & \Delta t_1^2 \\ 1 & \Delta t_2 & \Delta t_2^2 \\ ... & ... & ... \\ 1 & \Delta t_n & \Delta t_n^2 \end{bmatrix} \tag{3-84}$$

where $\Delta t_i = t_i - t_{oc}$. And the misclosure vector can be expressed as:

$$\mathbf{l} = \begin{bmatrix} \Delta T_s^1 \\ \Delta T_s^2 \\ \cdots \\ \Delta T_s^n \end{bmatrix} \tag{3-85}$$

where $\Delta T_s^i = T_s(t_i) - \tilde{\tilde{T}}_s(t_i)$, with $\tilde{\tilde{T}}_s(.)$ derived from some initial guess of the coefficients (f_0, f_1, f_2). Therefore, the correction to the coefficients can be obtained through:

$$\mathbf{x} = \left[\mathbf{H}^T \cdot \mathbf{H}\right]^{-1} \cdot \mathbf{H}^T \cdot \mathbf{H} \cdot \mathbf{l} \tag{3-86}$$

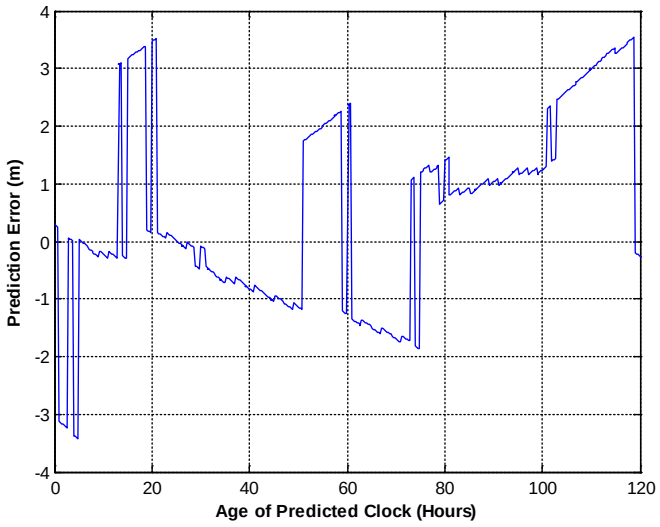


Figure 3-17 Illustration of Satellite Clock Fitting

3.7 Ephemeris Extension with Estimated Parameters

The parameters estimated in Section 3.5 and 3.3 can then be used to extend the ephemeris. For the orbital part, the higher accuracy obtained in the estimated ICs, the longer orbit can be extended to the future that still ensures good accuracy. However, for the part of satellite clock, even if the modeled clock fits well with the actual satellite clock during the observation period, it

is still not for sure that days after the clock extension the accuracy is still as good as during the fitted period because of the underlying random processes. The good thing is that, with the technology advancement, the stability of atomic clocks has been improved 2 ~ 3 orders in the past decades, and it is expected that in the near future, with higher order of atomic clocks deployed in navigation satellites (GPS World staff, 2010; Science Codex, 2011), extension of satellite clock to the same length as orbit to the future will still ensure excellent accuracy.

3.7.1 Basic Steps

The following figure gives the basic steps that are followed in ephemeris extension, among which the preprocessing of observations is to filter unqualified observations and unify the coordinate and time frames, the estimator module is to estimate the orbital parameters (ICs and solar radiation parameters) and satellite clock fitting coefficients, the predictor is to extend the orbit and satellite clock with the estimated parameters, and EE packaging is to pack the extended ephemeris in the form that is used in position fix, which could be in Keplerian or Cartesian form.

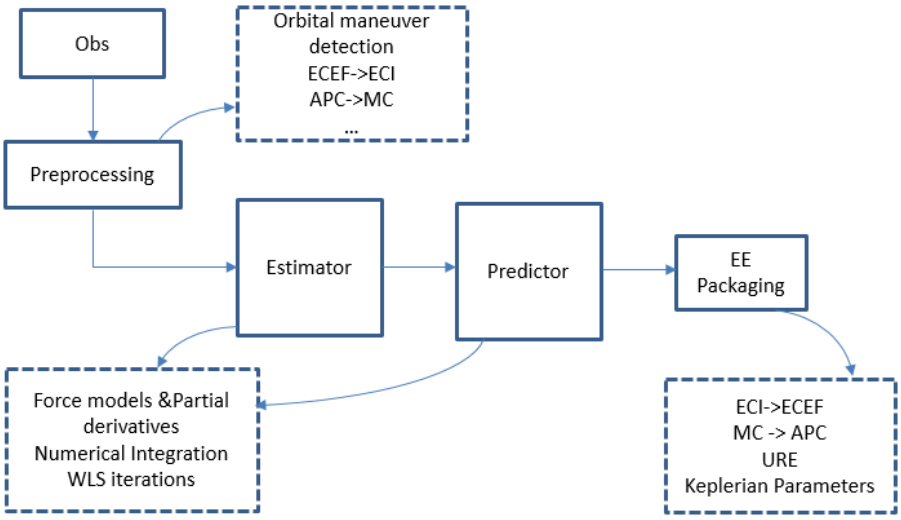


Figure 3-18 Basic Steps in Ephemeris Extension

It should be noted that, for the satellite velocity and acceleration input to numerical integrator and the output position and velocity from numerical integrator, they are all w.r.t. satellite MC. Whereas, the satellite position and velocity that are used in fixing of a GNSS receiver position are w.r.t. the satellite APC. The APC associated to the GPS and GLONASS broadcast orbits are available from the National Geospatial-Intelligence Agency (NGA) at http://earth-info.nga.mil/grandG/sathtml/gpsdoc2013_05a.html.

3.7.2 Flow chart

The core modules in ephemeris extension are the parameter estimation and orbital prediction (extension). The following figure attempts to give more details in the two modules.

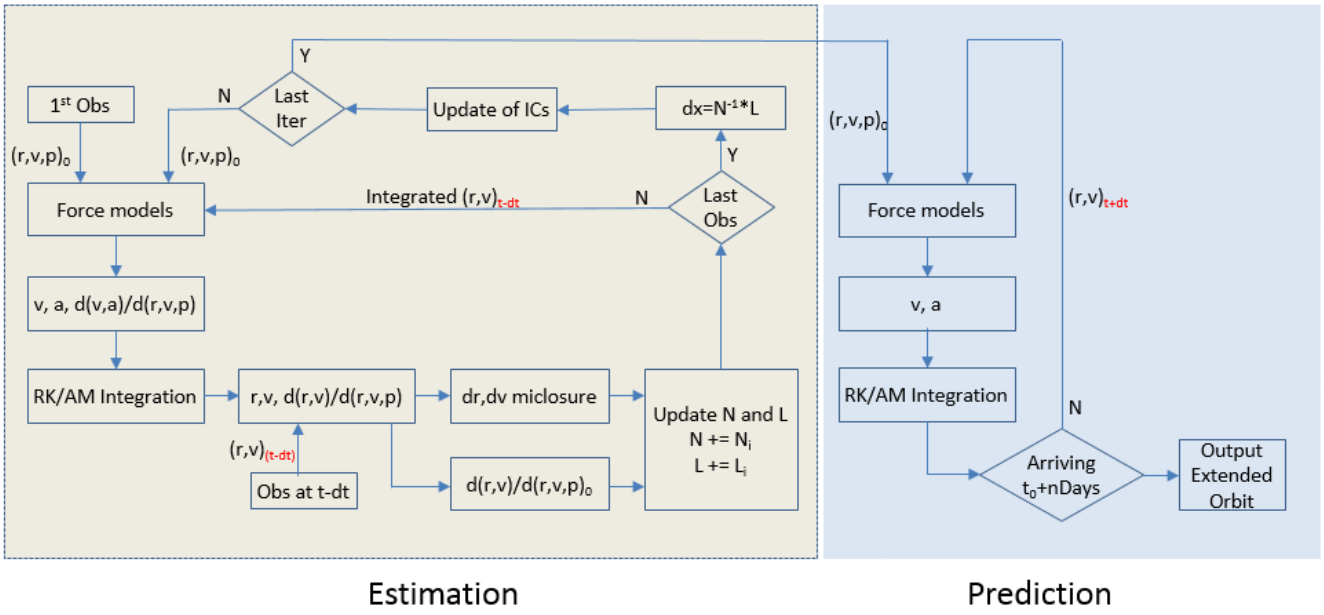


Figure 3-19 Flowchart of Parameter Estimation and Orbital Prediction

As shown in the above flowchart, the force models and numerical integrators are commonly used in both estimation and prediction. The difference is that, in estimation, partials are additionally calculated and the numerical integrations are iterated until converging to predetermined error tolerance level.

3.7.3 Evaluation of Extended Ephemeris

Real Broadcast Ephemeris (BE) or SP3 orbits can be used to calculated reference orbits and clock to evaluate the extended ephemeris (EE).

3.7.3.1 Orbital error in ECEF

The satellite position and velocity directly obtained from BE or SP3 are expressed in ECEF frame, so once the extended orbits are transformed to ECEF from ECI, they can be evaluated by comparing against the reference orbits.

3.7.3.2 Orbital error in RTN

The orbital errors calculated in ECEF are usually further transformed to RTN frame, in which the error components are easier to understand. It should be noted that RTN could be expressed in ECEF or ECI, but the radial components are common no matter in ECEF or ECI.

An example of orbital errors is given below, which shows that the orbital errors expressed in RTN give better idea of the influence of each error component – errors in radial direction are usually the smallest, errors in both radial and normal show obvious periodicity and zero-centered, but errors in tangent direction is non-zero-centered.

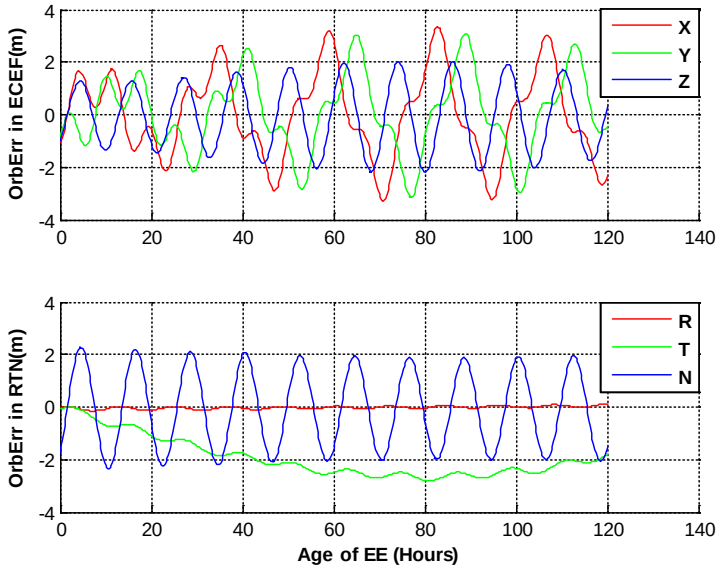


Figure 3-20 Orbital Errors Expressed in ECEF and RTN

CHAPTER 4 FAST POSITION FIX WITH LONG-VALIDITY EPHEMERIS

This chapter will focus on the algorithm study for position fix at different aforementioned working stages, under the condition that GNSS ephemeris is valid for long.

4.1 Fast Acquisition

The task of acquisition is to detect the signal and locate the start of PRN is the intermediate frequency (IF) samples. The way to speed up the acquisition process is to narrow down the search range by prediction of code phase and Doppler with the best available means.

4.1.1 Prediction of Code Phase and Doppler

The frequency of a satellite signal received at UE antenna is different from the nominal frequency at the time when the signal is transmitted from satellite antenna. Such frequency offset is resulted by the Doppler effect that is caused by the relative dynamics of satellite with respect to receivers on the Earth surface in the following form (Misra and Enge 2001):

$$D = f_s \cdot \frac{\Delta v}{c} \tag{4-2}$$

Where f_s is the nominal carrier frequency; c represents the speed of light; Δv is the relative velocity between the satellite vehicle (SV) and the receiver along the Line-Of-Sight (LOS) direction, which can be expressed as:

$$\Delta v = \frac{\mathbf{r}_s - \mathbf{r}_r}{|\mathbf{r}_s - \mathbf{r}_r|} \cdot (\dot{\mathbf{r}}_s - \dot{\mathbf{r}}_r) \tag{4-3}$$

Where $\mathbf{r}_s$ and $\mathbf{r}_r$ represent the SV and receiver positions, and $\dot{\mathbf{r}}_s$ and $\dot{\mathbf{r}}_r$ represent the SV and receiver velocities. In Figure 4-21, the maximum LOS velocity due to SV dynamics is illustrated, where R_E represents the Earth radius and H represents the SV orbital altitude. As the $|\dot{\mathbf{r}}_s|$ is determined by orbital altitude, for GPS SVs, the maximum $|\Delta v|$ is ~ 824 m/s.

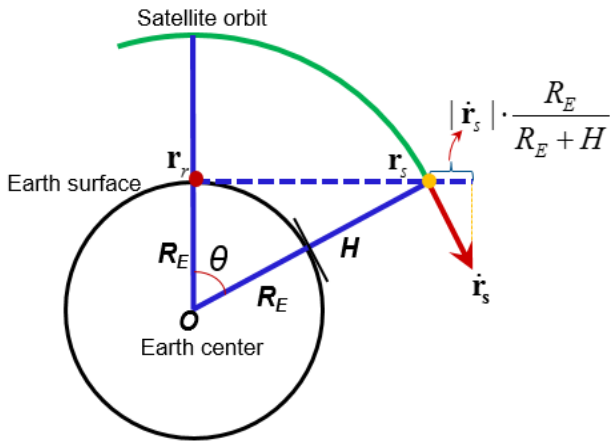


Figure 4-21 Maximum LOS Velocity due to SV Dynamics

To predict the code phase for a signal arriving at the UE antenna at the sampling moment, the time of signal reception (ToR), with uncertainty of σ_{ToR}, has to be precisely known. If the receiver position $\mathbf{r}_r$ and SV position $\mathbf{r}_s$ are also known with uncertainties of σ_{Rx} and σ_{sv}, the distance that the signal travels in space since the time of signal transmission (ToT) would be:

$$R = |\mathbf{r}_s - \mathbf{r}_r| \tag{4-4}$$

Therefore the error in R can be expressed as:

$$\Delta R = \frac{\mathbf{r}_s - \mathbf{r}_r}{\left|\mathbf{r}_s - \mathbf{r}_r\right|} \cdot (\Delta \mathbf{r}_s - \Delta \mathbf{r}_r) \tag{4-4}$$

So:

$$\left|\Delta R\right| <= \left|\Delta \mathbf{r}_s\right| + \left|\Delta \mathbf{r}_r\right| \tag{4-5}$$

Where the ΔR is at maximal when the vectors $\Delta \mathbf{r}_s$ and $\Delta \mathbf{r}_r$ are coincident with the direction of LOS, as the scenario illustrated in Figure 4-21, with the maximum uncertainty of σ_R being around $\sqrt{\sigma_{sv}^2 + \sigma_{rx}^2}$. Before precise ToT is obtained, $\mathbf{r}_s$ is computed with ephemeris at time $ToR - \frac{nominal\ travel\ time}{c}$, in which the nominal travel time for GPS is ~75 ms. The distance the signal travels in space (R) before arriving UE antenna can be directly translated to time the signal travels in space since ToT, considering the SV clock bias dt_s, then we get:

$$ToT = ToR - \frac{R}{c} - dt_s \tag{4-6}$$

with overall time uncertainty of σ_{ToT} depending on the σ_{ToR}, σ_{Rx}, σ_{sv} and etc. in the following form:

$$\sigma_{ToT} = \sqrt{\sigma_{ToR}^2 + \sigma_{R/c}^2 + \sigma_{ts}^2} \tag{4-7}$$

The σ_{ToT} expressed in Equation (4-7) represents the uncertainty of predicted code phase for a SV signal when arriving the UE antenna. In Figure 4-22, the relation among ToR, R and ToT is further illustrated along with the uncertainties.

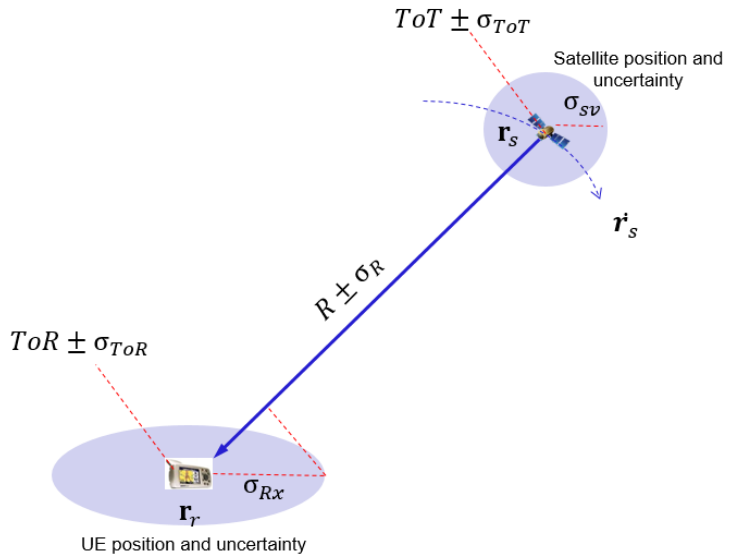


Figure 4-22 Prediction of Code Phase and Doppler of a SV Signal

4.1.2 Components Needed for Acquisition Prediction

It is now clear that, in order to predict the code phase and Doppler of a signal, the receiver time (ToR), receiver position and velocity, satellite position and velocity, satellite clock must be known at reasonable uncertainties.

The ToR may be obtained from a locally maintained clock, or time assistance from cellular network via different protocols, such as Precise Time Protocol (PTP), Network Time Protocol (NTP) and etc. (NTP, 2014). The time, with accuracy ranging from microseconds (us), milliseconds (msec) to seconds (s), can usually be referred as precise, fine and coarse time respectively.

For $\mathbf{r}_r$, it can be obtained from history position fixes on device, the fixes by alternative positioning systems on device, or network assistance. As alternative sources, position assistance may be from WiFi signal footprints or signal triangulations, Bluetooth, FM/AM radio, TV signals, cellular tower, Iridium system and etc., providing the position assistance at accuracy level of meters to tens of kilometers. When projected into LOS of UE to SV, such level of errors in R_r may introduce nanoseconds to tens of microseconds in predicted signal travel time.

For $\mathbf{r}_s$, $\dot{\mathbf{r}}_s$ and dt_s, they can be computed from ephemeris, almanac, or extended ephemeris, depending on whichever with the best accuracy is available. It should be noted that, the GPS ephemeris is fitted for 4 hours, and the accuracy will drastically increase beyond 4 hours. The almanac, updated no more than every 6 days (GPS IS, 2004), intends to provide approximate satellite orbit for a longer time, with accuracy mush worse than that of ephemeris but degrading with the age at a slower pace than ephemeris. The extended ephemeris, usually good for a few days, provides accuracy close to ephemeris. No matter which is used to derive $\mathbf{r}_s$, $\dot{\mathbf{r}}_s$ and dt_s, the uncertainties should be properly assumed.

4.1.3 Uncertainties in Satellite Position and Velocity

GNSS signal acquisition involves intensive search in both time and frequency domains, with the search time highly dependent on the sizes of the search windows. The frequency search window can be determined by ephemeris or almanac, with ephemeris more preferable for better precision. However, for the traditional GPS ephemeris, the accuracy degrades rapidly when it is used

beyond 2 hours, and the long validity ephemeris would demonstrate incomparable advantages in this case. Therefore, almanac, traditional ephemeris and the long validity ephemeris will be compared in terms of accuracy change with time, which is then translated into the accuracy of determined frequency search window and leads to difference in acquisition time.

In Figure 4-23, examples of typical orbital errors versus age are given for almanac (Alm), ephemeris (Eph) and long validity ephemeris (extended ephemeris - EE in this case).

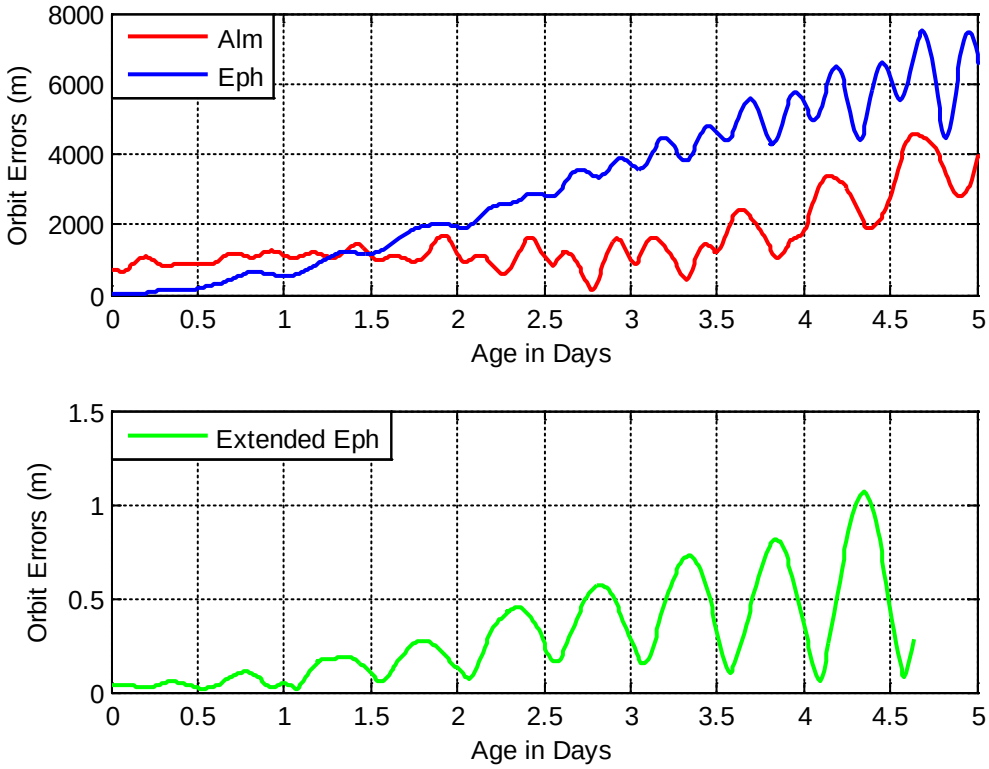


Figure 4-23 Typical Orbital Errors vs. Age for Alm, Eph and Extended Eph

As seen in the above Figure, the growth of ephemeris errors with the age is much faster than almanac. In the given example, the errors from ephemeris exceed almanac at the age of ~1.5 days

and reach almost 8 km at the age of 5 days, whereas the errors of extended ephemeris remain within 1.5 m for up to 5 days.

4.2 Fast Time Synchronization

Different time scales is illustrated in Figure 2-6 in Chapter 2, which explains the steps to obtain different time accuracies by synchronizing to signals at different stages. Depending on the uncertainty of available time at UE start-up, it is possible to skip some of the steps to achieve precise time, therefore the whole process of Time Synchronization (TimeSync) would be much faster than usual. The details of regular time synchronization, including Bit Synchronization (BitSync) and Frame Synchronization (FrameSync), are available in lots of literatures, such as Kaplan (2016) and will not be given here. Instead, this Section will focus on some techniques for fast time synchronization.

4.2.1 BitSync

Still taking GPS as example, once the start of C/A code is acquired, it is desired to locate the start of modulated bits (or bit edge) in the C/A sequence. Once the bit edge is identified (namely synchronized), time resolution is immediately improved to millisecond level, but with the part of time above millisecond still being uncertain. In other words, at the end of BitSync, precise time is not fully available yet, and only the part under 20 milliseconds (length of a bit) is known.

In case that the ToT predicted in Equation (4-6) has uncertainty (σ_{ToT}) beyond 20 milliseconds, there are 20 possible locations of bit edge to be tested and verified one by one in regular BitSync. However, if the uncertainty of the predicted ToT is better than 20 milliseconds, the process of BitSync can be sped up, as there are fewer possible locations of bit edge to be tested and verified. If ToT is predicted at the accuracy of better than a millisecond, bit edge can be directly obtained and therefore the time consuming process for BitSync can be skipped.

Moreover, obtaining the bit edge in one SV can benefit the BitSync in the rest of the visible SVs. As illustrated in Figure 4-24, assuming that bit edge of Signal in SV1 is identified, namely the portion of ToT_1 under 20 milliseconds is precisely known. Let's denote it as:

$$\sigma_{mod(ToT_1,20ms)} = 0 \tag{4-6}$$

Considering that,

$$\begin{aligned} ToT_2 - ToT_1 &= (ToR - \frac{R_2}{c} - dt_{s_2}) - (ToR - \frac{R_1}{c} - dt_{s_1}) \\ &= \frac{R_2 - R_1}{c} + (dt_{s_1} - dt_{s_2}) \\ &= \frac{\Delta R_{2,1}}{c} + \Delta dt_s \end{aligned} \tag{4-7}$$

Therefore:

$$ToT_2 = ToT_1 + \frac{\Delta R_{2,1}}{c} + \Delta dt_s \tag{4-8}$$

Taking modular on both sides in the above Equation (4-8), the following uncertainty can then be derived:

$$\sigma_{\text{mod}(\text{ToT}_2,20\text{ms})} = \sqrt{\sigma^2_{\text{mod}(\text{ToT}_1,20\text{ms})} + \sigma^2_{\Delta R_{2,1}/c} + \sigma^2_{\Delta t_s}} = \sqrt{\sigma^2_{\Delta R_{2,1}/c} + \sigma^2_{\Delta t_s}} \quad (4\text{-}9)$$

The uncertainty of bit edge in the signal of SV2 is no longer directly subject to σ_{ToR}, and is significantly narrowed down comparing to Equation (4-8). Therefore, the process of BitSync on SV2 is sped up, and similarly in the rest of the visible SVs.

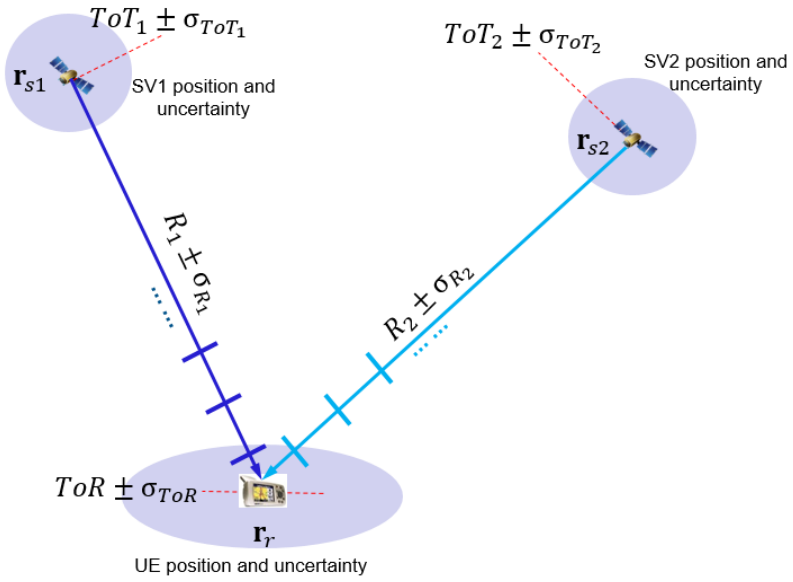


Figure 4-24 Relative Bit Edge between SVs

4.2.2 FrameSync

Once bit edge is determined, the modulated NAV data bits can be decoded and collected one by one. In the collected bit stream, locating the start of a subframe is the critical step to obtaining the full precise time.

As introduced in GPS IS (2004), each subframe starts with an 8-bit preamble (10001011) in a telemetry (TLM) word, as the indicator of the start of a subframe. Locating the preamble in the bit stream is a primary means for Frame Synchronization (FrameSync). As the preamble occurs every 6 seconds in the data frames, locating the preamble takes at least 6 seconds.

In each subframe, immediately following the TLM word is the Handover Word (HOW), which includes the time-of-week (TOW) count corresponding to the start of the next following subframe, and the subframe ID. The full precise time is obtained when a TOW is successfully decoded, together with the further decoded subframe ID, the subsequent NAV data can be decoded with corresponding data structures.

Once the FrameSync is obtained in one SV, the full ToT corresponding to selected sampling time is available, the $\mathbf{r}_s$, R and dt_s in Figure 4-22, that are originally based on a rough ToT, can now be updated with a precise ToT. The uncertainty σ_{ToT} of the precise ToT after FrameSync, depending on the signal environment and design of code tracking loop, is often achievable at nanoseconds level. So the ToR can also be immediately updated as:

$$ToR = ToT + \frac{\Delta R}{c} + dt_s \tag{4-10}$$

with the uncertainty σ_{ToR}, that is initially at seconds level, now being improved to milliseconds or even microseconds level.

Given the FrameSync in one SV, it is possible to skip the same FrameSync process in the rest of the SVs, as the start of subframes in those SVs can be determined using Equation (4-8). As illustrated in Figure 4-25, assume that the FrameSync is achieved in SV1, so the full ToT_1 corresponding to the sampling moment can be calculated by counting the number of words, bits, epoch numbers and code phases since the start of current subframe (TOW).

Using Equation (4-8), the ToT_2 corresponding to the same sample moment can be derived, with uncertainty σ_{ToT_2}. Considering the improvement in σ_{ToR}, $\sigma_{\Delta d}$, σ_{dt_s} incurred by the improvement in σ_{ToT_1}, the uncertainty σ_{ToT_2} is also significantly improved. The uncertainties $\sigma_{\Delta R}$ and σ_{dt_s} are further dependent on the σ_{R_r} and the source of orbit and clock (ephemeris or almanac). In the context that ephemerides are available for both SV1 and SV2, and σ_{R_r} is less than 300 km, it is feasible to derive ToT_2 with uncertainty σ_{ToT_2} less than 1 millisecond. Given ToT_2 at such precision, it is easy to determine which bit the sample moment is located since the start of current subframe, and equivalently, it is easy to determine which bit is the start of current subframe in the bit stream of SV2.

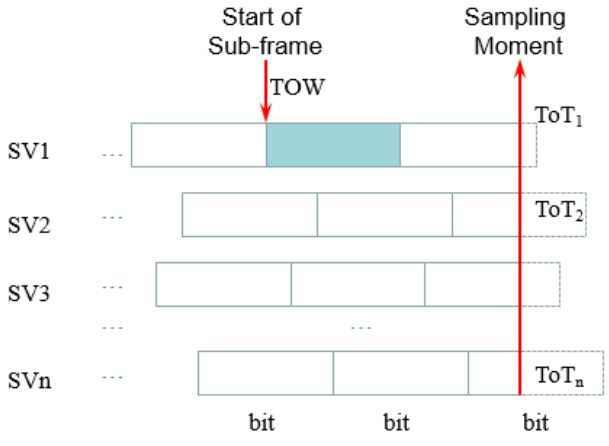


Figure 4-25 FrameSync and Relative Bit Edges between Different SVs

4.2.3 Soft TimeSync

Instead of obtaining time from the above FrameSync, it is possible to obtain the time synchronization through parameter estimation when previously collected ephemerides are still valid, however under some prerequisite conditions on σ_{r_r} and σ_{ToR}. Comparing to traditional FrameSync, the process of obtaining precise time through parameter estimation is referred as Soft Time Synchronization in this thesis, which obviously has the advantage of pushing a position fix prior to FrameSync or even BitSync. As such parameter estimation is always conducted together with position estimation, it is only briefly mentioned in this section for completeness of methods for time synchronization, leaving the algorithm details to be further explored in Section 4.4.

4.2.4 Significance of long-validity ephemeris on Fast TimeSync

For all the techniques discussed for obtaining fast TimeSync, the prerequisite condition is that the ephemeris is valid. Therefore, the long-validity ephemeris is be of great significance for obtaining fast TimeSync as it is not subject to frequent expirations.

4.3 Obtaining Pseudorange Measurements

After obtaining full TimeSync on a SV, the precise ToT for the signal at the instant when arriving at UE antenna can be derived as:

$$ToT = TOW + Words \times 0.6 + Bits \times 0.02 + EN \times 0.001 + CodePhase \times 10^{-6}\,(\text{s}) \qquad (4\text{-}11)$$

Where TOW is stands for the time of the start of current subframe that is decoded from the HOW, $Words$ stands for the number of words counted from the start of current subframe, $Bits$ stands for the number of bits counted from current word, EN stands for the epoch number – the location of current C/A code in current bit (0 ~ 19 msec), and $CodePhase$ stands for submsec code phase of current C/A that is taken from code tracking loop. For all of these items, each represents a different time scale that has been illustrated in Figure 2-6 in Chapter 2, when aggregated together, they constitute the most precise measure of the ToT for the signal.

Then the full pseudorange (PR) measurement of the signal can be taken as:

$$Full_PR = ToR - ToT \qquad (4\text{-}12)$$

Where the ToR represents the UE time for the moment when the SV signal arrives at the UE antenna, which contains time bias w.r.t. the GNSS time system. For multiple SV signals arriving at the UE antenna at the same instant, ToR is in common for taking the PR measurements and the time bias is estimated in position fix.

In the stage before BitSync is obtained, $CodePhase$ is the only available part in Equation (4-13), where the parts in the square bracket are still unknown. $CodePhase$ is actually available as soon as the signal is acquired and then the accuracy is further improved in stable tracking stage.

$$ToT = [TOW + Word \times 0.6 + Bits \times 0.02 + EN \times 0.001] + CodePhase \times 10^{-6}(\text{s}) \quad (4\text{-}13)$$

Equation (4-14) illustrates the stage after BitSync but before FrameSync is obtained, where EN and $CodePhase$ are available, with the parts in the square bracket remained unknown.

$$ToT = [TOW + Word \times 0.6 + Bits \times 0.02] + EN \times 0.001 + CodePhase \times 10^{-6}(s) \quad (4\text{-}14)$$

4.4 Fast First Position Fix

Regular first position fix is after full ephemeris is downloaded, which is often too slow especially under weak signal condition. When the long validity ephemeris is available, fast position fix becomes possible immediately after signal acquisition. Therefore, different algorithms are studied to push for the first position fix in the stage as early as possible on a standalone GNSS UE (Figure 2-6).

4.4.1 Position Fix before Ephemeris is Decoded Over The Air (OTA)

When ephemeris with long validity is available, the first position fix is possible right after the FrameSync is obtained, and the OTA ephemeris decoding can be skipped in each new start-up until the existing ephemeris is expiring. Therefore, the full TTFF for position fix of a standalone GNSS UE that is depicted in Figure 2-6 in Chapter 2 is immediately reduced from >=18 seconds to >=6 seconds.

4.4.2 Position Fix before Full TimeSync

Even if long validity ephemeris is available, regular position fix is possible only after full time synchronization, in order to get accurate signal transmission time. Algorithms for fast position fix are available to skip the full TimeSync, which significantly improve TTFF (McBurney and

Sanders, 2001; Diggelen, 2002) from >=6 seconds to <= 1second as long as a few stringent prerequisite conditions are satisfied:

- Measurements of a few msec for >=5 SVs
- Valid ephemeris
- Rough time (<1 minute)
- Rough location (<30 km)

When full TimeSync is obtained, the time offset in the initialized UE time is removed and UE time accuracy is immediately improved to msec level, leaving the remaining time bias to be estimated in regular position fix. However, when full TimeSync is skipped, in addition to the part of time bias, the time offset introduced to the UE time from initialization is remained.

Positioning algorithms to skip the TimeSync, no matter FrameSync, or even BitSync, are of little difference as both introduce an additional parameter to estimate for the UE time offset due to the lack of TimeSync. As a part of the research of this thesis, a new method is developed that is able to further improve the TTFF with much relaxed prerequisite conditions. Therefore, the algorithms are not introduced separately for the 2 cases, but will be explained together in next section.

4.4.3 Position Fix with Snapshot Measurements

Based on the discussions in previous sections, there are cases to push the ability of position fix on a GNSS UE to extreme by using a few snapshots of measurements for sub-millisecond (submsec) code phase and Doppler. Certainly, there must be some preconditions on $\sigma_{\mathbf{r}_r}$ and σ_{ToR}.

In Table 4-8, the conditions for position fix with snapshot measurements are specified. As seen, comparing to the literatures (Liu et al., 2010), with the methods proposed by this thesis, the conditions can be further relaxed.

Table 4-8 Preconditions on Position Fix with Snapshot Measurements

Literatures	This thesis
• Measurements of a few msec for >=5 SVs • Valid ephemeris • Rough time (<1 minute) • Rough location (<30 km)	• Meas of a few msec for >=4 SVs • Valid ephemeris • Rough time • No rough location is needed

The method developed by this thesis consists of 3 steps:

(1) Determine approximate user location
(2) Construct full PRs that consists of multiple integer msec and submsec parts
(3) Estimate time offset, and accurate user location

4.4.4.1 Determine approximate user location

In this section, the approximation of user location is categorized into 3 levels from very rough (a few thousand km), to rough (< 300 km) and approximate (a few tens of km).

In order to determine the very rough user location, the sub-points $\bar{r}_s$ in Equation (4-15) for the visible satellites on the Earth surface at the rough time are calculated, then converted into (Latitude, Longitude, Altitude) and averaged.

$$\bar{\mathbf{r}}_s = \frac{\mathbf{r}_s}{|\mathbf{r}_s|} \cdot R_E \tag{4-15}$$

The approximation with the average usually guarantees accuracy of a few thousand km. If signal strength of each SV is available as the weight when averaging the sub-points, accuracy of the approximation of user location is achievable under 1000 km.

The very rough user position can be further improved to rough (under 300 km) through the following Location Search Algorithm (LSA). Firstly, let's define a metrics for LSA – the standard deviation of Doppler residuals:

$$\sigma_{k,p} = std(D_{k,c} - D_{p,o}) \tag{4-16}$$

where $D_{p,o}$ denotes the Doppler observations of all acquired SVs at location p (to be estimated), and $D_{k,c}$ denotes the calculated Doppler at selected candidate location k for those SVs. Considering that for one SV, the difference between the calculated and observed Dopplers (residual) is a reflection of user clock drift and user location offset, for multiple SVs, the consistency of the above residuals is a measure of user location offset. So the smaller $\sigma_{k,p}$, the closer the selected location k is to the actual position p.

Next, let's get a list of candidate locations covering the uncertainty area of the actual position, so that the above metrics can be applied. Figure 4-26 illustrates how the list of candidate locations are prepared that cover the whole uncertain area. The list consists of a 'Start' point, that is selected at the aforementioned average of the sub-points, and additional points that are selected

evenly distributed on the edges of the surrounding hexagons. The total number of the candidate locations depends on the selected search step size and uncertainty of the initial position. By configuring different search step sizes, positions at different approximation levels can be obtained by the LSA.

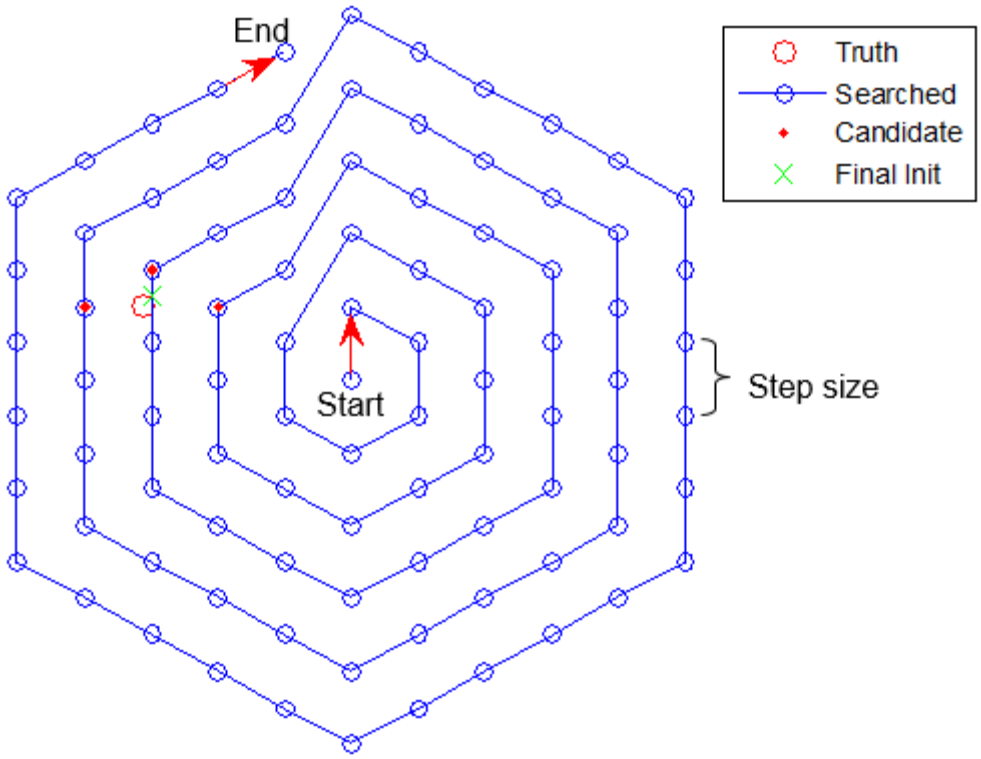


Figure 4-26 List of Candidate Locations for LSA

Lastly, determine the final position by looping through the list of candidates. At each candidate location k, calculate the metrics $\sigma_{k,p}$, and record the 3 candidate locations with the smallest 3 $\sigma_{k,p}$. The final location $\mathbf{r}_{LSA}$ can be derived from the weighted average as:

$$\mathbf{r}_{LSA} = \sum_{k=1}^{3} \mathbf{r}_k \cdot \frac{1}{\sigma_{k,p}} \tag{4-17}$$

Where $\mathbf{r}_k$ represents the recorded 3 candidate locations.

4.4.4.2 Construct full pseudoranges

Firstly let's take a look at signal travel distance based on accurate ToT and ToR:

$$\hat{\rho} = \hat{t}_{ToR} - \hat{t}_{ToT} \tag{4-18}$$

Where

$\hat{t}_{ToR}$ represents accurate time of signal reception in msec

$\hat{t}_{ToT}$ represents accurate time of signal transmission in msec

$\hat{\rho}$ represents accurate signal travel distance in unit of msec

Also consider the observation equation, that relates the signal travel distance with the geometry range between the satellite and UE, neglecting atmospheric and other factors for simplicity:

$$\hat{\rho} = \hat{R} - dt_s + ... \tag{4-19}$$

where

$\hat{R}$ the geometry range between UE at $\hat{t}_{tor}$ and satellite at $\hat{t}_{tot}$, in unit of msec

dt_s the satellite clock bias at $\hat{t}_{ToT}$ in unit of msec, which is maintained within range of [0, 1) msec

As accurate ToR is usually not directly available, we could only derive PR based on accurate ToT and approximate ToR, and then Equation (4-19) becomes:

$$\begin{aligned}\tilde{\rho} &= \tilde{t}_{ToR} - \hat{t}_{ToT} \\ &= \tilde{t}_{ToR} - (\hat{t}_{Int_ToT} + subms) \\ &= (\tilde{t}_{ToR} - \hat{t}_{Int_ToT} - 1) + (1 - subms) \\ &= Int1_ms + (1 - subms)\end{aligned} \tag{4-20}$$

Where

$\tilde{t}_{ToR}$	represents ToR in msec, selected to be the closet integer msec to $\hat{t}_{ToR}$
$Int1_ms$	represents an integer number of msec
$subms$	represents sub-msec code phase, in the range of [0, 1)
$\tilde{\rho}$	represents the PR in unit of msec, with bias from the inaccuracy in $\tilde{t}_{ToR}$

In regular handling, the $\hat{t}_{Int_ToT}$ in the Equation (4-20) is obtained from FrameSync and count of words and bits, and $subms$ is the code phase measurements taken from tracking loop. Then $\tilde{t}_{ToR}$ is assumed to be $\hat{t}_{ToT}$ plus an approximate signal traveling time, leaving the remaining time error to be estimated along with the position fix. However, in the context of the discussion in this section, the full process of time synchronization is not finished yet, therefore $\hat{t}_{ToT}$ is actually not available.

Given an approximate UE location $\tilde{\mathbf{r}}_r$ and considering the observation (4-19), we get:

$$\begin{aligned}\tilde{\rho} &= \tilde{R} - dt_s + dt_r + ... \\ &= \tilde{R} - dt_s + dt_r + ... \\ &= (Int2_ms + \Delta\tilde{R}) - dt_s + dt_r \\ &= Int2_ms + (\Delta\tilde{R} - dt_s + dt_r)\end{aligned} \tag{4-21}$$

Where

dt_r is the UE clock bias in $\tilde{t}_{ToR}$, in range of [0, 1) msec, as $\tilde{t}_{ToR}$ is selected to be the closet integer msec to $\hat{t}_{ToR}$

$\tilde{R}$ is the calculated geometry range from rough UE location $\tilde{\mathbf{r}}_r$ to satellite at $\tilde{t}_{ToT}$, in unit of msec

$Int2_ms$ is the integer msec part of $\tilde{R}$

$\Delta\tilde{R}$ is a sub-msec part of $\tilde{R}$

Please note, in the calculation of above $\tilde{R}$, as $\hat{t}_{ToT}$ is not yet available, the $\tilde{t}_{ToT}$, approximated by the $\tilde{t}_{ToR}$ minus signal travel time (also approximate), is used to calculated satellite position. The approximation error in GPS signal travel time can be controlled within 20 msec, and the impact on $\Delta\tilde{R}$ is negligible comparing to the error in $\tilde{t}_{ToR}$, because a GPS satellite travels less than 60 m within the 20 msec. The impact of error in $\tilde{t}_{ToR}$ will be further discussed later on.

The $Int1_ms$ is needed in Equation (4-20) to construct a full pseudo-range. Combining Equation (4-20) and (4-21), we can get the following:

$$Int1_ms = Int2_ms + \left[(\Delta\tilde{R} - dt_s + dt_r) - (1 - subms)\right] \quad (4\text{-}22)$$

If the selected initial UE location and initial time are accurate enough to ensure that uncertainty of $\Delta\tilde{R}$ to be within 1 msec, and considering the following uncertainty ranges:

$\Delta\widetilde{R}$	[-0.5, 0.5]
dt_s	[-0.5, 0.5]
dt_r	[0, 1]
subms	[0, 1]

approximating $Int1_ms$ with $Int2_ms$ (the integer msec part of $\widetilde{R}$) could result to ambiguity of a few integer msec in the range of [-2, 2], because of the part in the square bracket.

So the **full pseudo-range** can be constructed as:

$$\widetilde{\rho} = Int2_ms + (1 - subms) \tag{4-23}$$

Which could have ambiguity of a few integer msec in the range of [-2, 2], and needs to be resolved SV by SV through a Viterbi-alike Msec Search Algorithm (MSA) (Zhang, 2012) that is described in Figure 4-27:

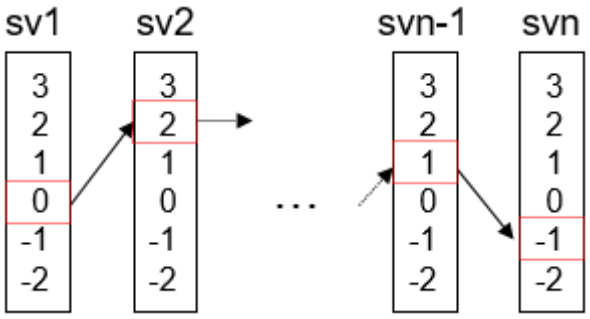


Figure 4-27 Resolving Msec Ambiguities in obtained PRs

The above MSA takes the following steps:

- For the 1st SV, select the ambiguity to be 0 msec, and calculate the range residual

$\tilde{L}(1) = \tilde{\rho}^1 - \tilde{R}^1 + \Delta T_{s1}$

- For the 2nd SV, calculate the range residual for each ambiguity msec, and choose the msec that gives the minimum $std(\tilde{L}(1), \tilde{L}(2))$
- …
- For the i-th SV, calculate the range residual for each ambiguity msec, and choose the msec that gives the minimum $std(\tilde{L}(1),...,\tilde{L}(i))$

The rationale behind the above MSA is similar to the aforementioned LSA algorithm. In the MSA description, when the right msec ambiguity is determined for a SV, the calculated range residual $\tilde{L}(i)$ primarily reflects the magnitude of the time offset in UE clock, which is a common part for all SVs. When the right msec ambiguities are determined for all SVs, the corresponding range residuals should appear the most consistent, or equivalently have the smallest stand deviation. It should be noted that, $\tilde{L}(i)$ actually also contains the impact of $\Delta\tilde{R}$ that incurred by the orbit offset when inaccurate ToT is used, but the impact is so insignificant comparing to the ambiguities of the integer msec. However, smaller uncertainty of ToR can ensure better reliability of the above MSA.

4.4.4.3 Estimate time offset and accurate user location

Considering that precise ToT is not available, the following linearization can be obtained from Equation (4-4):

$$
\begin{aligned}
R &= \left|\mathbf{r}_s(t) - \mathbf{r}_r\right| \\
&\approx R_0 - \frac{\mathbf{r}_s(t) - \mathbf{r}_r}{R_0} \cdot \Delta\mathbf{r}_r - \frac{\mathbf{r}_s(t) - \mathbf{r}_r}{R_0} \cdot \dot{\mathbf{r}}_s(t) \cdot \Delta t_{ToT} \\
&= R_0 - \frac{\mathbf{r}_s(t) - \mathbf{r}_r}{R_0} \cdot \Delta\mathbf{r}_r - \dot{R}_0 \cdot \Delta t_{ToT}
\end{aligned}
\tag{4-24}
$$

Where

R_0 the calculated geometry range with given UE location and ToT t

$\dot{R}_0$ the change rate of above geometry range, equivalent to Δv in Equation (4-2), with upper bound of 824 m/s for GPS satellites

$\Delta\mathbf{r}_r$ the position offset in UE location $\mathbf{r}_r$

Δt_{ToT} the time offset in the ToT

Please note that, before FrameSync, $\tilde{t}_{ToT}$ that is approximated by $\tilde{t}_{ToR}$ minus signal travel time, is used instead of $\hat{t}_{ToT}$. The Δt_{ToT} in the above equation, denoting the offset in $\tilde{t}_{ToT}$, is actually one part of time offset in $\tilde{t}_{ToR}$, as $\tilde{t}_{ToR}$ is where $\tilde{t}_{ToT}$ is derived from.

In Equation (4-24), Let

$$
\Delta R = \frac{\mathbf{r}_s(t) - \mathbf{r}_r}{R_0} \cdot \Delta\mathbf{r}_r + \dot{R}_0 \cdot \Delta t_{ToT} \tag{4-25}
$$

So we get

$$
\left|\Delta R\right| \le \left|\Delta\mathbf{r}_r\right| + \left|\dot{R}_0\right| \cdot \left|\Delta t_{ToT}\right| \tag{4-26}
$$

Which gives the upper bound of the error in the computed geometry range incurred by the inaccuracy in the UE location and the *ToR*. If to limit $|\Delta R|$ within certain uncertainty range, proper combination of uncertainty requirements must be exerted on initial UE location and UE time. For example, considering the maximum $|\dot{R}_0|$ that is explained in Figure 4-21, as a rule of thumb combination, the uncertainty of $\Delta\mathbf{r}_r$ within 30 km and uncertainty of Δt within 100 sec, gives confidence in the uncertainty of $|\Delta R|$ within 150 km (or equivalently 0.5 msec).

Based on Equation (4-25), the observation Equation can be further derived as:

$$\begin{aligned}\tilde{\rho} &= \tilde{R} - dt_s + dt_r \\ &\approx (R_0 - \frac{\mathbf{r}_s(t) - \mathbf{r}_r}{R_0} \cdot \Delta\mathbf{r} - \dot{R}_0 \cdot \Delta t_{ToT}) - dt_s + dt_r\end{aligned} \tag{4-27}$$

Where $\Delta\mathbf{r}$, Δt_{ToT} and dt_r are unknown parameters, with $\mathbf{r}_s(t)$ and dt_s dependent on Δt_{ToT}. In Equation (4-27), both Δt_{ToT} and dt_r are the time offsets in the initial UE time (ToR), however, Δt_{ToT} only represents the part that could affect *ToT*, or consequently the calculation of $\mathbf{r}_s(t)$ and dt_s, and dt_r represents the part that affect the measure of the PR.

As in most applications, the UE is on the earth surface, it is possible to apply such a constraint as an additional observation:

$$\begin{aligned}\tilde{\rho} &= \tilde{R} \\ &\approx (R_0 - \frac{\mathbf{0} - \mathbf{r}}{R_0} \cdot \Delta\mathbf{r}) \\ &= R_0 + \frac{\mathbf{r}}{R_0} \cdot \Delta\mathbf{r}\end{aligned} \tag{4-28}$$

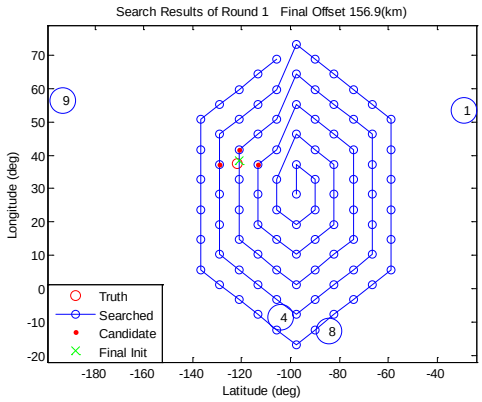
Making use of observation Equation (4-27) alone, or together with Equation (4-28), the position fix can be made without FrameSync. So the state vector is:

$$\mathbf{x} = \begin{bmatrix} \Delta x & \Delta y & \Delta z & dt_{r2} & dt_{r1} \end{bmatrix}^T \tag{4-29}$$

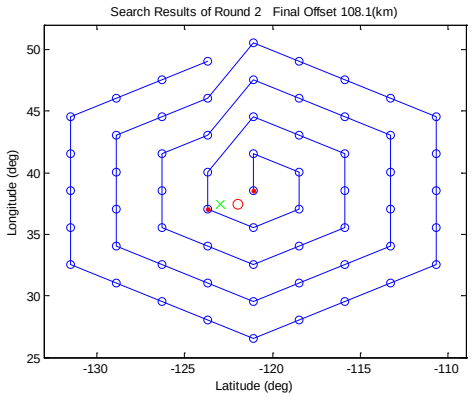
And design matrix is in the form of:

$$\mathbf{A} = \begin{bmatrix} \cos\alpha^1 & \cos\beta^1 & \cos\gamma^1 & \dot{R}_0^1 & -1 \\ \cos\alpha^2 & \cos\beta^2 & \cos\gamma^2 & \dot{R}_0^2 & -1 \\ \cos\alpha^3 & \cos\beta^3 & \cos\gamma^3 & \dot{R}_0^3 & -1 \\ \cos\alpha^4 & \cos\beta^4 & \cos\gamma^4 & \dot{R}_0^4 & -1 \\ .. & ... & ... & ... & -\mathbf{1} \\ \cos\alpha^0 & \cos\beta^0 & \cos\gamma^0 & 0 & 0 \end{bmatrix} \tag{4-30}$$

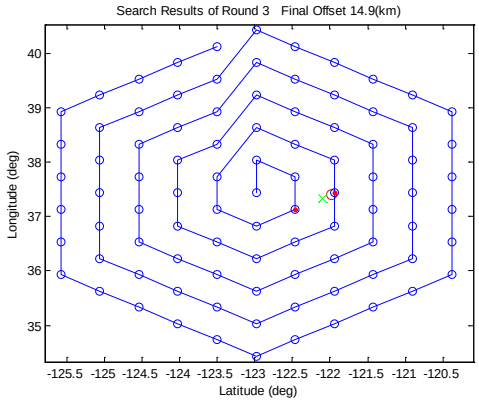
Using the new method, an example of fast position fix is given below in Figure 4-28. The user scenario includes valid ephemerides and measurements (sub-msec code phases, Doppler and CNo) from 4 satellites (SV 1, 4, 8, 9), time uncertainty of around 100 seconds, and no knowledge of user location. After applying the LSA for 3 rounds, user location is improved from unknown to ~160.2 km, ~95.6 km, and then to ~9.5 km, starting from which millisecond ambiguity in PR measurements are resolved and the final position is estimated.



(a)



(b)



(c)

Figure 4-28 Illustration of the Fast Position Fix

It should be noted that, the DOP computed in this section is different from in regular position estimation with full TimeSync, because different design matrix (Equation 4-30) is used. Full explanation of the DOP and comparison with the DOP from regular fix are available in Diggelen (2009). Also, because the estimation of the UE time offset has dependency on the satellite velocity, the resolution of the final estimation is difficult to be better than a few tens of msec (say 20-40ms). So the fix obtained using this method is just used as temporary solution to improve TTFF, and the fix is switched to regular method as soon as FrameSync is obtained for best accuracy.

4.4.4 Feed-forward the Estimated Time for Fast Time Synchronization

As has been introduced, in the regular first position fix on a UE with valid ephemeris, the acquired signals are firstly pulled in the tracking loop, then bit edge must be detected (namely BitSync), and then the data bits are collected one by one until FrameSync is obtained. The

position fix can be initiated only after FrameSync. For such regular position fix, the state vector is:

$$\mathbf{x} = \begin{bmatrix} \Delta x & \Delta y & \Delta z & \Delta t_{r1} \end{bmatrix}^T \tag{4-31}$$

Where Δx, Δy and Δz represent the coordinate offset in the initial UE location, and Δt_{r1} represents the time bias in the UE clock, which is directly translated into a common ranging offset in the PR measurements for all SVs and is typically maintained within sub-msec.

When the algorithms introduced in last section is applied for a position fix before FrameSync is obtained, the state vector becomes:

$$\mathbf{x} = \begin{bmatrix} \Delta x & \Delta y & \Delta z & \Delta t_{r2} & \Delta t_{r1} \end{bmatrix}^T \tag{4-32}$$

Where Δt_{r2} represents the time offset in the initial UE time, which disappears immediately after precise ToT is obtained from FrameSync. To avoid confusion, Δt_{r1} is referred to as UE Clock Bias, and Δt_{r2} is referred to as UE Time Offset in this thesis. So before FrameSync, even if full PR are obtained with the algorithms introduced in last section, the precise ToT is still not available for each SV, but has to be derived by ToR, with the accuracy improved gradually through a few iterations.

Each time the state vector in Equation (4-32) is estimated, the improved ToR is obtained as follows:

$$\tilde{\tilde{t}}_r = \tilde{t}_r + \Delta t_{r2} + \Delta t_{r1} \tag{4-33}$$

Then the corresponding ToT for each SV can be obtained as:

$$\tilde{t}_s = \tilde{\tilde{t}}_r - \Delta t_{travel} + \Delta t_s \tag{4-34}$$

Where Δt_{travel} is the signal travel time in space from satellite antenna to UE antenna, and it is initially approximated with ~75 ms, and then replaced with R/c. Then $\tilde{t}_s$ can be used to calculate the satellite position, in order to continue the estimation in next iteration. Figure 4-29 depicts the full iteration process for the estimating the state vector in Equation (4-32).

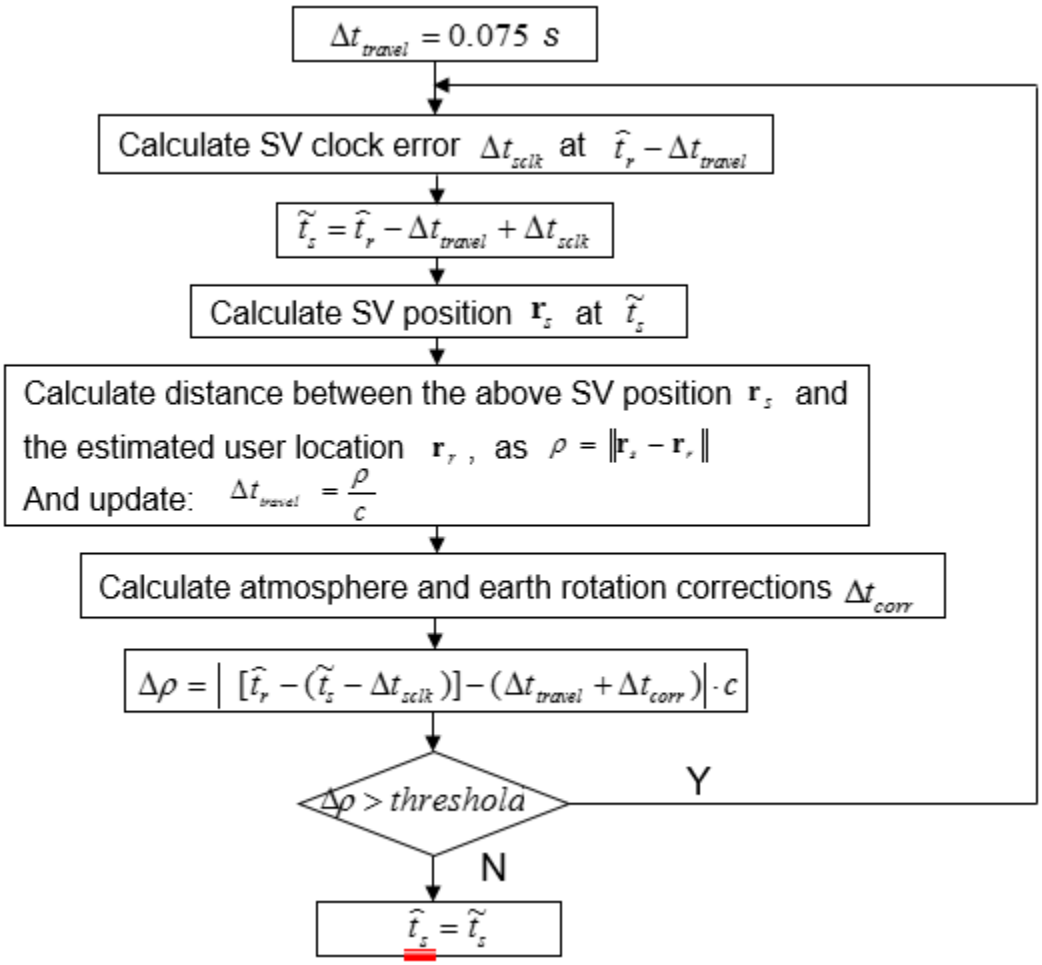


Figure 4-29 Iterative Estimation Process without FrameSync

The position fix process without FrameSync is usually carried out in parallel to regular BitSync and FrameSync, as depicted in Figure 4-31. Such position fixes have to be made again and again, until BitSync and FrameSync are obtained, so that it can be then transitioned to regular position fix algorithm.

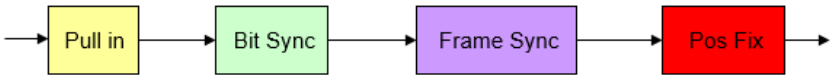


Figure 4-30 Process of Regular Position Fix

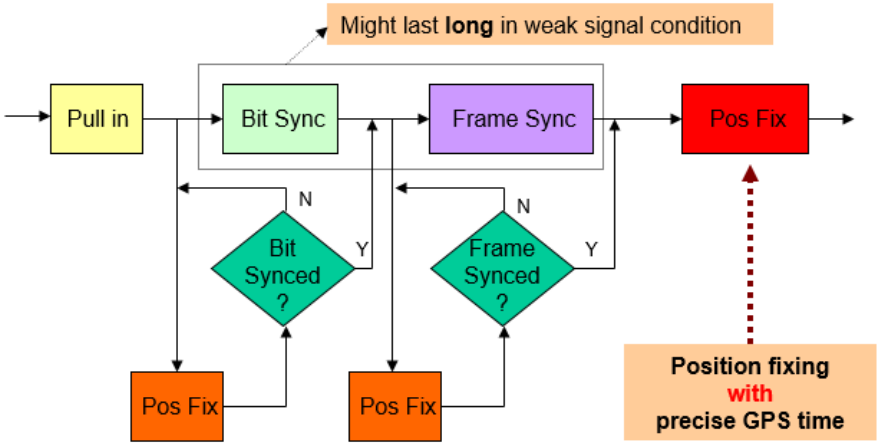


Figure 4-31 Process of Position Fix without FrameSync

Considering that in certain cases it may take long time to obtain BitSync and FrameSync, Zhang (2010) proposed a method to feed forward the estimated UE Time Offset into BitSync and FrameSync, as illustrated in Figure 4-32, so that the process can be accelerated.

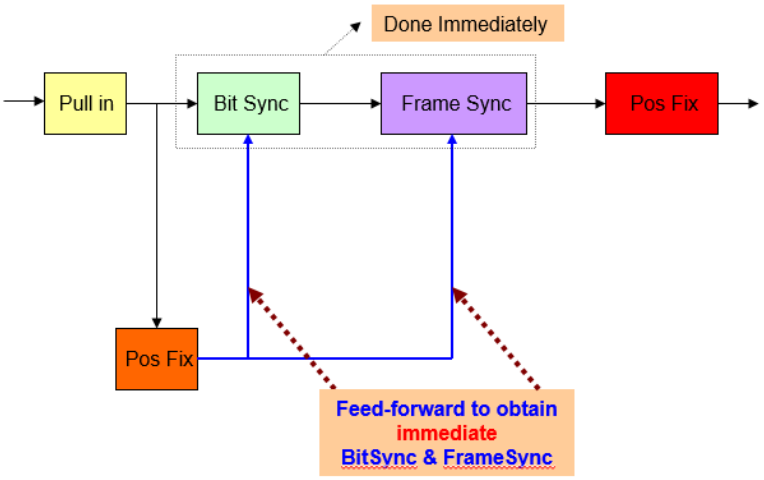


Figure 4-32 Process of Position Fix with Feed-forward of Estimated Time

As soon as the fix is obtained, precise $\hat{t}_r$ and $\hat{t}_s$ can be immediately derived through Equation (4-33) and Equation (4-34), with the accuracies depending on geometry and the quality of sub-msec code phases. Then based on $\hat{t}_s$, the measurements at different time scales can be derived according to Table 2-1.

Considering the range of accuracy in $\hat{t}_s$, it may help speed up FrameSync but may not always help BitSync. To help BitSync, $\hat{t}_s$ has to be as accurate as a few msec, so that *iEpochNum* can be derived; given typical accuracy of a few tens of msec in $\hat{t}_s$, the derived *iBit* is expected with very small uncertainty, say 1-2 bits, so the location of sub-frame boundary can be immediately narrowed down to range of 1-2 bits. If the NAV data bits have been previously collected, it is feasible to immediately get FrameSync by verifying the located Frame boundary within 1-2 bit uncertainty.

Table 4-9 Deriving BitSync and FrameSync from Estimated ToT

ToT Components	Range
$iPage = ceil(mod(\hat{t}_s, 750)/30)$	1 ~ 25, integer
$iSubframe = ceil(mod(\hat{t}_s, 30)/6)$	1 ~ 5, integer
$iWord = ceil(mod(\hat{t}_s, 6)/0.6)$	1 ~ 10, integer
$iBit = ceil(mod(\hat{t}_s, 0.6)/0.02)$	1 ~ 30, integer
$iEpochNum = floor(mod(\hat{t}_s, 0.02)/0.001)$	0 ~ 19, integer
$iChip = floor(mod(\hat{t}_s, 0.001) \times 1023000)$	0 ~ 1022, integer
$iChipPhase = mod(\hat{t}_s, 0.001) \times 1023000 - iChip$	0.0 ~ 1.0, float
$\mathrm{mod}(y, x) = y - floor(\frac{y}{x}) \cdot x$	

4.5 Position Fix with 3 SVs

In order to speed up the TTFF, it is a common practice that 3 SVs are used in the estimation of the first fix to save the time waiting for the measurements from the 4th and additional SVs. However, a 3-SV fix often comes with the risk of an image solution, under some extreme case. As one of the outcome of this research, a new method for the detection of 3-SV image solution will be introduced here, serving the fast and reliable first position fix.

When only 3 SVs are used in a position fix, an altitude constraint is always applied based on the assumed altitude at the user location. The design matrix therefore is in the following form:

$$A = \begin{bmatrix} cos\alpha^1 & cos\beta^1 & cos\gamma^1 & -1 \\ cos\alpha^2 & cos\beta^2 & cos\gamma^2 & -1 \\ cos\alpha^3 & cos\beta^3 & cos\gamma^3 & -1 \\ cos\alpha_0 & cos\beta_0 & cos\gamma_0 & 0 \end{bmatrix} \tag{4-35}$$

Where $[cos\alpha^i \quad cos\beta^i \quad cos\gamma^i]$ represents the directional cosine from the i-th SV w.r.t. the given initial position, and $[cos\alpha_0 \quad cos\beta_0 \quad cos\gamma_0]$ represents the directional cosine of the given initial position w.r.t. the Earth center. In pure geometry, for the 3 spheres centered at the 3 SVs, the intersection on the Earth surface consists of 2 points. When extending the issue to pseudoranging from 3 SVs, the intersection on the Earth surface also consists of 2 points. So when the above design matrix is used, the fix could end up with an image solution – namely converging to the other point. In case of 3-SV fix, how to ensure the fix is not from an image solution? This section is to present an innovated method.

Let's start with some 3-SV scenarios. At GPS Time 1846, 247263 and truth location (Lat/Lon/Alt) [37.79736621, -122.398865, 4.226], the full set of visible GPS SVs includes 5, 7, 11, 13, 15, 17, 19, 28 and 30. Multiple 3-SV scenarios of different geometry are created by selecting different 3 SVs from the list of visible SVs. Given an initial user position, the residuals of both the pseudorange (PR) and PR rate (PRR) are computed for the selected 3 SVs. Assuming there are no outliers in the PR and PRR measurements, when a fix is obtained, the PR residuals of the 3 SVs should be well consistent, so should the PRR residuals.

In order to visualize the consistency of the computed residuals, a metrics for PR residuals is selected as $log_{10}\{std(PR\ Residuals)\}$, and a metrics for PRR residuals is selected as $log_{10}\{std(PRR\ Residuals)\}$. In Figure 4-33 and Figure 4-34, the metrics of PR/PRR residuals

are computed on grid points from longitude [-180 180] degrees and latitude [-90 90] degrees and are plotted for a 3-SV scenario with good geometry, in which SVs 4, 6, 9 are selected and the PDOP is 2.2. It should be noted that the PDOP is calculated with the design matrix given in Equation (4-35) based on the above truth position.

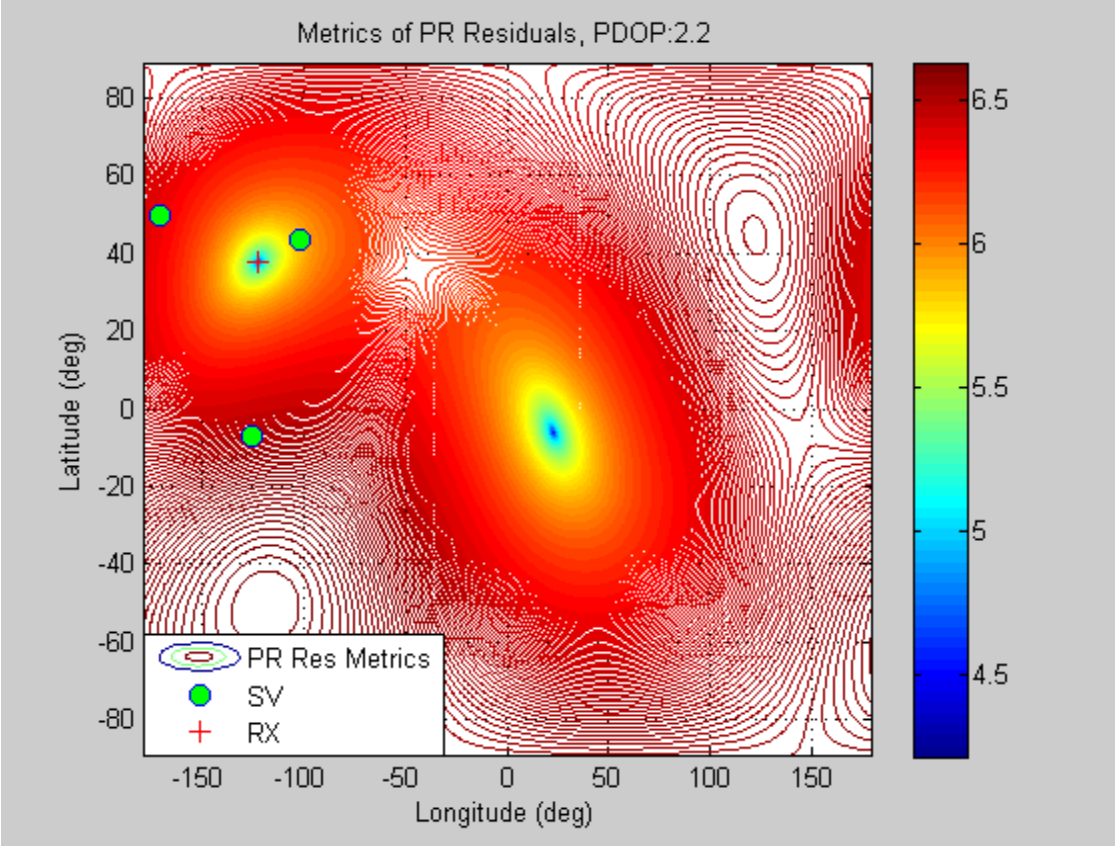


Figure 4-33 Metrics of PR Residuals for a 3-SV scenario with PDOP 2.2

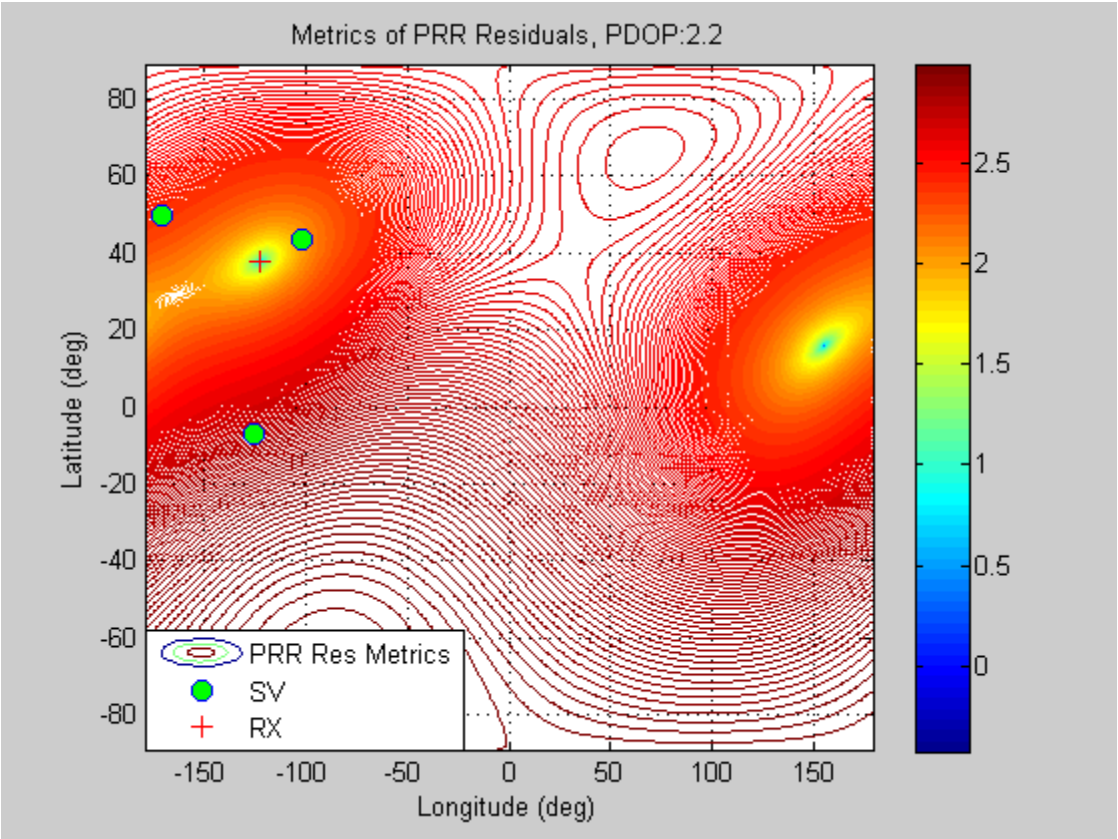


Figure 4-34 Metrics of PRR Residuals for a 3-SV scenario with PDOP 2.2

In the above Figures, it can be seen that, the lowest metrics of both PR and PRR residuals happen at two positions, with one being the truth position and marked in red '+', and the other obviously being the so-called image position. Also, at the image position where the metrics of PR residual is the lowest, the metrics of PRR residual is NOT the lowest. The image position in Figure 4-33 is different from the image position in Figure 4-34.

In Figure 4-35 and Figure 4-36, the metrics of PR/PRR residuals are plot for a 3-SV scenario with extremely poor geometry, in which SVs 2, 5, 9 are selected and the PDOP is 53.4. In addition to what has been observed in Figure 4-33 and Figure 4-34, there is something new in

these 2 Figures. Around both the truth and image positions, the change of metrics is not as sharp as seen when PDOP is small, which indicates that under poor geometry condition, it would be much more difficult for a fix to converge.

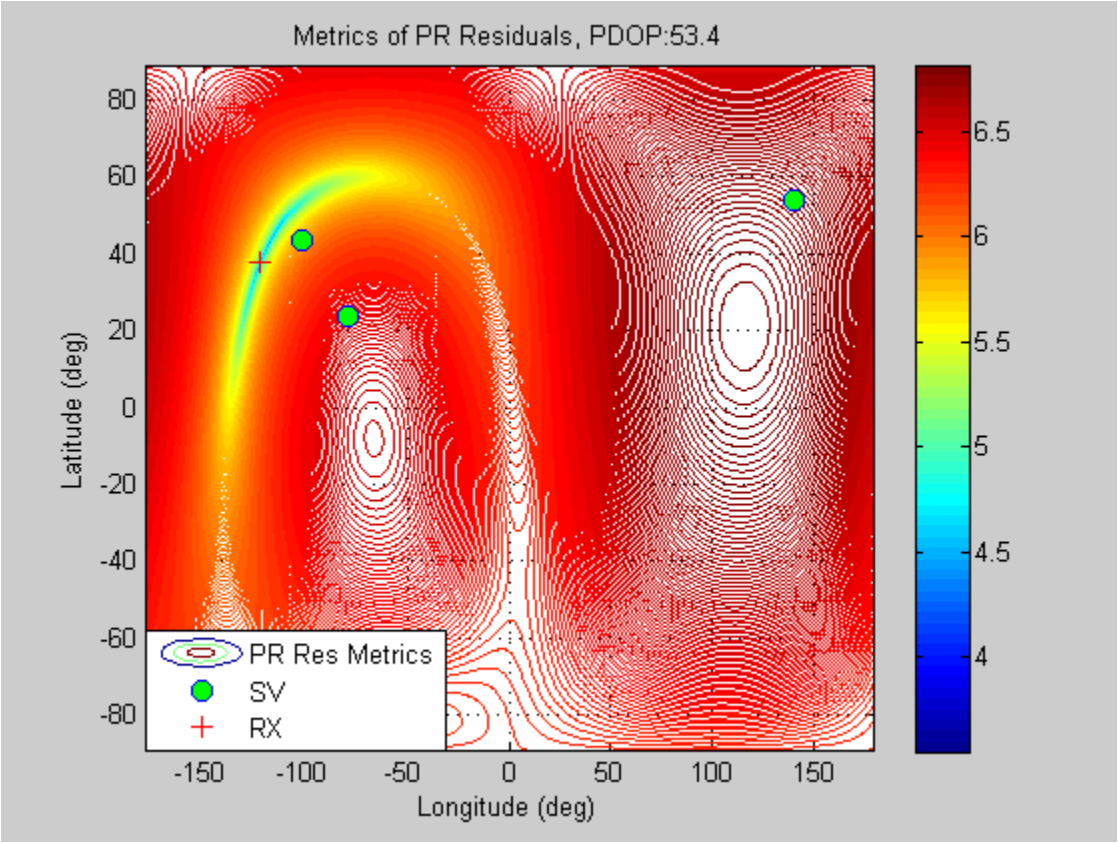


Figure 4-35 Metrics of PR Residuals for a 3-SV scenario with PDOP 53.4

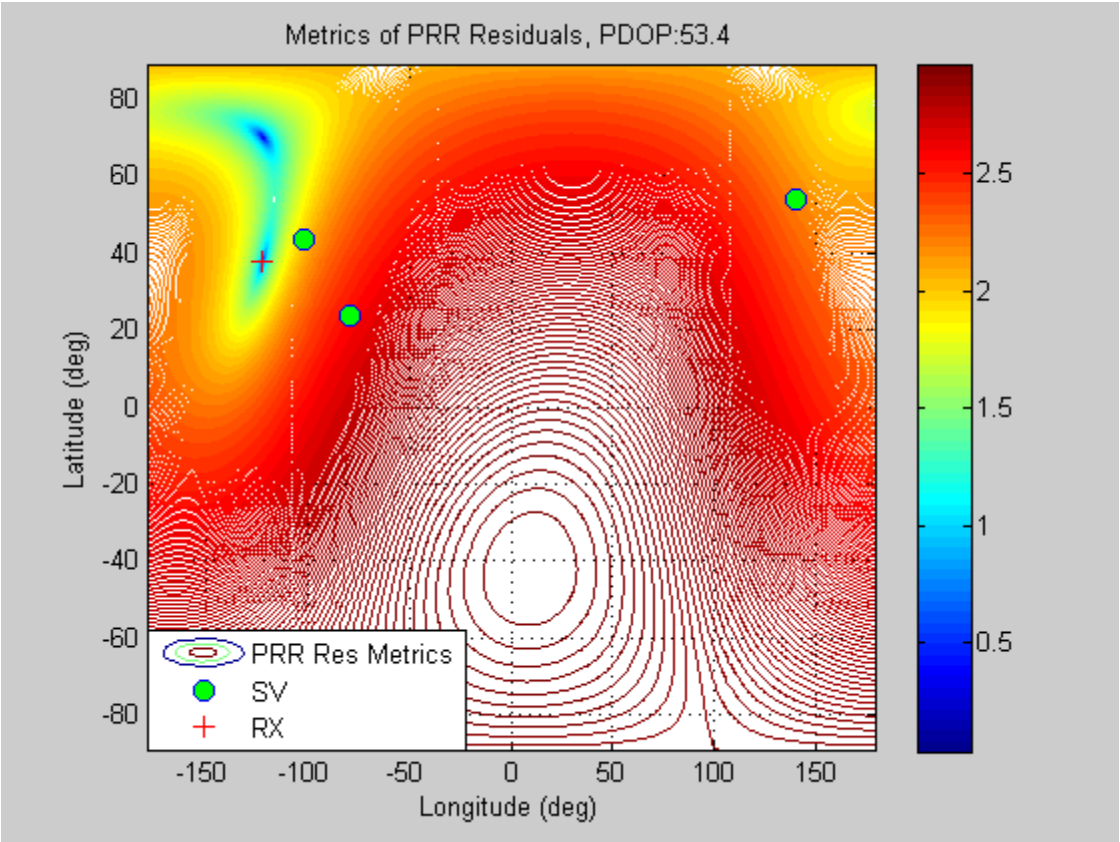


Figure 4-36 Metrics of PRR Residuals for a 3-SV scenario with PDOP 53.4

Analysis of the metrics reminds us that, if the given initial position is close to the image position, the position fix is at high risk of converging to a wrong position – the image position. The observation that the image position associated to PR is not overlapped with the image position associated with PRR, enables us to discriminate whether a 3-SV fix is converged to the image position or not. Here is a summary of the method:

(1) After a 3-SV fix is estimated, computed the metrics of PRR residuals $log_{10}\{std(PRR\ Residuals)\}$ at the estimated position;

(2) Select one position ~300 km apart from the above fix and computed the corresponding metrics of PRR residuals;

(3) Select one more position on the other side of the above fix at same distance and computed the metrics of PRR residuals;

(4) Compare the above 3 sets of metrics. If the one at the above fixed position is the lowest, then the fix is converged to the right position, otherwise it is converged to the image position.

4.6 Concluding remarks

Conventional startup modes on a GNSS UE usually includes cold start, warm start and hot start. Different startup modes indicate different user location uncertainties, different time uncertainties and ephemeris availabilities. Assuming that ephemerides with long validity period are available, it is potentially feasible to turn the conventional warm start into a hot start, a cold start into a partial warm start, which would result to significantly shortened TTFF. With the long validity ephemeris, the TTFF improvements are expected primarily lie in two aspects:

(1) Accelerate the signal acquisition

(2) Enable position fixes with different algorithms as soon as measurements become available

CHAPTER 5 HIGH SENSITIVITY WITH LONG-VALIDITY NAV MESSAGES

This chapter will study the techniques that are commonly used in obtaining higher acquisition and tracking sensitivity in existing GNSS UEs, identify the challenges, and investigate how the sensitivity on GNSS UEs can be facilitated with long-validity NAV messages.

5.1 Overview

As the primary content in NAV messages, ephemeris with long validity period is eventually turned into long-validity NAV messages.

The study will look into the details of techniques for high sensitivity – coherent and non-coherent integration with long dwell time. After quick comparison between coherent and non-coherent integration, the study will then focus on coherent integration, for which data bits of NAV messages must be known and fortunately could be known during the integration period, given the long-validity NAV messages.

Using existing GNSS, the barriers for long coherent integration could be either the data bits are unavailable or some are wrong due to erroneous decoding or data expiration. In such cases, the deterioration on sensitivity will be studied, and compared with when long-validity NAV messages are available.

Then different techniques related to data bit aiding will be fully studied, including different methods for fast bit edge determination and bit decoding, different methods for data bit alignment, reconstruction of full navigation frames with sliced navigation messages that are previously collected, etc. Although the purpose of the study in this part is sensitivity improvement, sensitivity and TTFF are often discussed together, as the improvements in the two aspects are often coupled together.

5.1.1 Sensitivity of GNSS Receiver

Taking GPS as the example, the minimum signal power transmitted from the satellite antenna is 26.8 dBW, and after experiencing free-space loss and atmospheric attenuation, the minimum signal power (P_S) arriving at the UE antenna would be -127.6 dBm (GPS IS, 2004). For signal transmitted in bandwidth B (Hz), the noise power can be approximated as:

$$P_N = k \cdot T_N \cdot B \tag{5-1}$$

Where k is Boltzmann's constant ($1.38 \times 10^{-23} Joule/^\circ K$), and T_N is the effective noise temperature in degree Kelvin, which is typically 513 degree for a GPS receiver (Braasch et al., 1999). After the intermediate frequency (IF) sampling, the signal bandwidth is at least 2 MHz, which corresponds to the noise floor of -108.5 dBm according to Equation (5-1). Therefore, the signal-to-noise ratio (SNR) at this stage is:

$$\begin{aligned} SNR &= P_S(dB) - P_N(dB) \\ &= -19.1dB \end{aligned} \tag{5-2}$$

A common goal in the different stages of baseband processing is to obtain sufficient gain in SNR, so that the signal can be successfully detected and measurements of reasonable accuracies can be taken for position fixes. Given the level of signal power (P_S), improving SNR is actually about lowering the noise power (P_N), which is equivalent to lowering the noise bandwidth (B_n).

For signals received in open sky environment, the ideal SNRs at some typical B_n are summarized in Table 5-10, neglecting degradations due to finite-bit quantization in A/D conversion.

Table 5-10 Ideal SNR at Different Noise Bandwidth

Signal Power P_s (dBm)	Noise Bandwidth B_n	Noise Power P_n (dBm)	SNR (dB)
-127.6	2 MHz	-108.5	-19.1
	1 kHz	-141.5	13.9
	50 Hz	-154.5	26.9
	1 Hz	-167.5	43.9

The probability of successful signal detection (p_d) is a function of SNR and the false alarming rate (p_{fa}) (see Equation 5-4), and the SNR has to be improved to a reasonable level to ensure successful signal detection.

Sensitivity in a receiver is usually referred to the lowest level of input signal power that is required to keep the receiver operational. Therefore, high sensitivity is equivalent to low signal power that can be processed. For the position fix procedures in a GNSS receivers as described in Figure 2-6 in Chapter 2, to keep different stages operational, the required minimum input signal

power may be different, which enable the categorization of the receiver sensitivity in terms of the following metrics:

(1) Acquisition sensitivity

It specifies the lowest level of received signal power at which the signal can be detected with reasonable p_d and acceptable p_{fa}.

(2) Tracking sensitivity

It specifies the lowest level of signal power at which the signal can remain steady locking status. Usually, tracking sensitivity of a receiver is much higher than the acquisition sensitivity. In other words, after a signal is acquired, it can be tracked even if the signal level is lowered.

(3) Position fix sensitivity

Very often, although the signals can be tracked, the measurements cannot be used to deliver a valid position fix. This metrics refers to the lowest level of signal power that a receiver is able to provide fixes at reasonable accuracy.

Regular sensitivity usually indicates that, a GNSS UE is operational only when the received signal power is close to the minimum received signal power that is specified in GPS ICD 200C (2000) (also see Table 5-10). Receivers of regular sensitivity are mostly standalone devices, which are able to work through each stage as described in Figure 2-6.

High sensitivity is now a standard feature in the GNSS products in the market, which refers to the operational signal level on a receiver, which, comparing to the regular sensitivity, typically could be a much as 30 dB lower. However, high sensitivity receivers may need external assistance to work through the stages described in Figure 2-6, such as NAV data bits, which are needed but may not be decoded over the air (OTA) by the high sensitivity receivers.

In the GNSS markets, there are ultra-high sensitivity products (MediaTek, 2014), which, comparing to the high sensitivity receivers, claim to work at even lower signal power and however, are believed under heavy assistance and stringent prerequisite conditions.

5.1.2 Obtaining High Sensitivity

The sensitivity improvement in a GNSS UE involves the efforts from antenna, RF front-end to baseband signal processing. In the RF front-end processing, the received signals are down converted to an intermediate frequency (IF), and the IF signals are further converted to baseband signals, in which the frequency content will be concentrated around 0 Hz, rather than on the original carrier frequency or the IF. For illustration purpose, the simplified diagram of baseband processing in a single channel is given in Figure 5-37 (Navipedia, 2017).

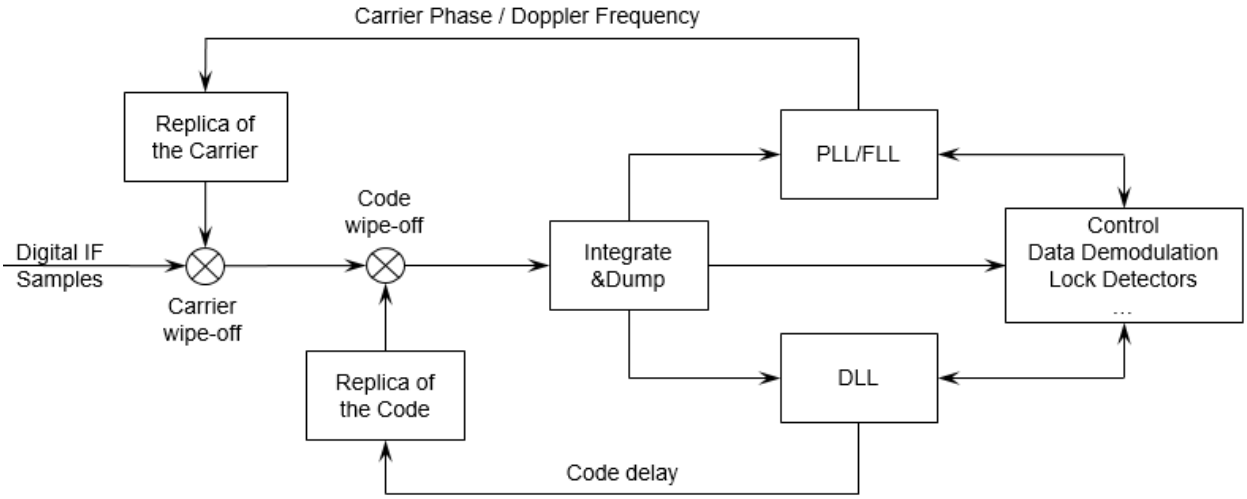


Figure 5-37 Simplified Block Diagram of a Single Channel for Baseband Processing

As depicted in the Figure 5-37, the carrier of the IF samples is removed through the multiplication with a replica of local carrier, then the results are further multiplied with a replica of local code for code wiping off. Usually IF samples are passed millisecond (msec) by msec to the carrier and code wiping off, and the results are further integrated msec by msec over a configured time period before being dumped to carrier/code tracking loops, and data demodulation. The carrier tracking may consist of Frequency Locked Loop (FLL) and/or a Phase Locked Loop (PLL), and the code tracking is through a Delay Locked Loop (DLL). The outputs of the carrier and code tracking loops are used to adjust the local carrier and code (replica of the carrier/code), to better align with the received carrier and code.

In regular signal condition, signal can be usually detected in the integration of 1 msec or only a few msec's results. However, in weak signal condition, adequate processing gain in Signal-to-Noise Ratio (SNR) needs to be obtained to have sufficient success rate in signal detection. For this purpose, usually the combination of coherent and non-coherent integrations of samples are

used in the baseband signal processing. Equation (5-3) gives an approximation to the processing gain combining coherent and non-coherent integrations. For a fixed total integration period (T_I), increasing the coherent period (T_c) is more efficient to get higher processing gain than non-coherent, because the non-coherent integration is subject to squaring loss (SL). Therefore, if possible, coherent integration is usually more preferable. However, without knowing the navigation bits, the coherent integration is limited within 1 bit period, namely 20 ms for GPS given the data rate of 50 bps.

$$Gain(dB) \approx 20 \cdot \log(\sqrt{T_c}) + 10 \cdot \log(\sqrt{\frac{T_I}{T_c}}) \tag{5-3}$$

To improve the sensitivity to -160 dBm, coherent integration over multiple bits is desired. Therefore, valid navigation bits as well as the bit boundaries are needed for data wipe-off. For this purpose, the previously collected navigation bits can be directly used if still valid; or the fresh navigation messages from different sources, including ephemeris and almanac, can be used to recover the navigation bits. Therefore, with the long validity NAV bits, it is foreseeable that the obtaining high sensitivity in signal acquisition and tracking would be greatly facilitated.

Once the weak signals are acquired, tracked, and the measurements are taken, even if the NAV bits could not be decoded from the tracked signals, position fixes are still possible with the long validity ephemeris. Certainly, lots of new challenges arise in the position computation, such as the dealing with false acquisition, false tracking and weighing the signals beyond regular strength range.

5.2 Signal Acquisition

The first goal of signal acquisition is to determine whether a signal is present, if present, then find the carrier frequency of the input signal and the beginning of the C/A code. In open sky environment, signal acquisition involves search in time and frequency domains in order to locate the carrier frequency of received signal and the start of the C/A code. However, when signal becomes weak, things become much more complicated, especially when there is no knowledge about whether a signal is present or not. Under such condition, the signal search has to be extended to the 3rd domain, namely the level of signal strength, as illustrated in Figure 5-38. Detection of weak signal requires the integration of samples in dwell time longer than usual, however the paradox is that signal strength is unavailable until the signal is acquired. A signal is not detected probably because the signal is not present or it is too weak. Therefore, on one side, before declaring the absence of a signal, samples must be integrated in dwell time that is long enough for designed limit in acquisition sensitivity; and on the other side, with the increase of dwell time, before declaring the presence of a signal, the detection threshold must be properly selected to avoid false acquisition.

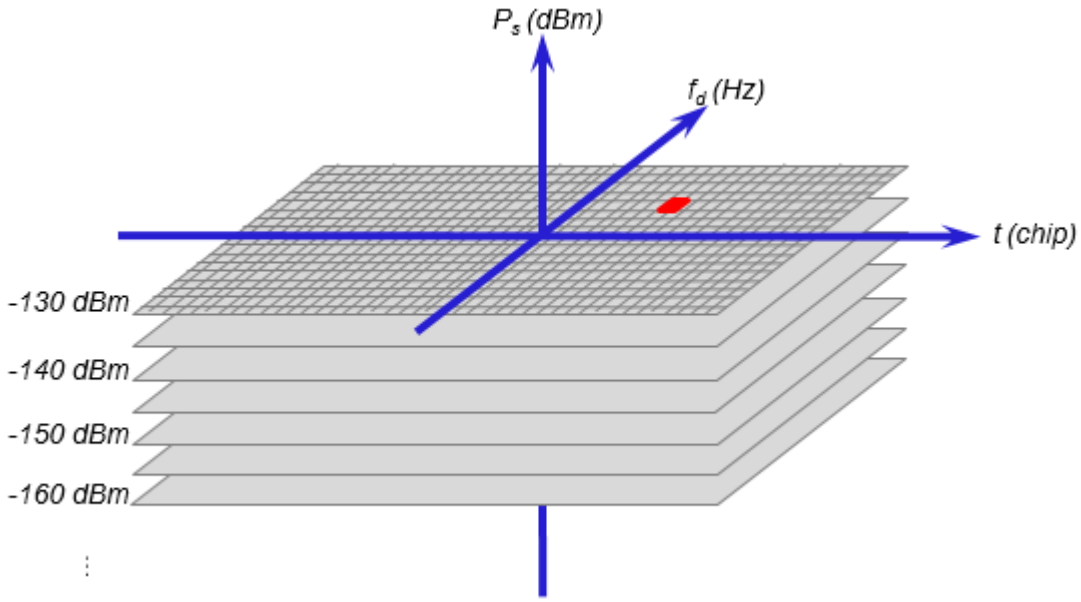


Figure 5-38 Search Space in Weak Signal Acquisition

The acquisition usually starts from a regular signal level, say -130 dBm, and then continues on a lower level if needed. Given a signal level, the dwell time can be determined. So for each assumed signal level, the integration results on the corresponding dwell time are sent for signal detection. If the signal is not detected, the integration is continued on a longer dwell time for the next signal level before being passed to a new round of signal detection. For different dwell time, different detection thresholds have to be determined in advance.

5.2.1 Theory of Signal Detection

The signal detection is based on the signal envelope ($\sqrt{I^2+Q^2}$), namely the magnitude of the in phase (I) and quadrature phase (Q) components, which is computed after the I and Q components

are integrated respectively over a preselected dwell time (T). The computed envelope is then compared to a predetermined threshold to decide whether a signal is present or not.

The I and Q components are assumed following Gaussian distribution, therefore the envelope $\sqrt{I^2+Q^2}$ follows Ricean distribution. When a signal is present, the probability density function (pdf) of the envelope, p_s, is a Ricean distribution and denoted as; and when a signal is not present, namely the signal amplitude is zero, the pdf of the envelope, p_n, then becomes a Rayleigh distribution.

In the detection of signal, an envelope threshold is selected in advance based on the desired p_{fd} and the measured 1-sigma noise power. For a single trial detection, the envelope above the threshold is declared containing the signal, otherwise is declared containing only noise. The probability of signal detection in a single trial, p_d, can therefore be determined by the integration of the p_s when the envelope is above the threshold; and the probability of signal false alarm in a single trial (namely a signal is declared present but actually it is absent), p_{fd}, can be determined by the integration of the p_n when the envelope is the threshold.

For a single trial, the approximate relationship between p_d and p_{fd} can be obtained through Neyman-Pearson test, and is given by North (2007):

$$p_d = 0.5 - erfc(\sqrt{-\ln p_{fa}} - \sqrt{SNR + 0.5}) \tag{5-4}$$

where $erfc(z) = 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-x} dx$. Usually in the receiver design, p_d and p_{fd} are selected in advance, so the required SNR can be derived by this Equation, and the strategy of combination of coherent and non-coherent integrations can be determined according to Equation (5-3). As an example, SNR of 12.85 is necessary to ensure a p_d of 99.5% with a p_{fd} of 0.1% (Table 4.1 in Chapman, 2000). Given the signal power in Table 5-10, integration of 1 msec samples is marginally sufficient to achieve similar p_d and p_{fd}. In some extremely signal environment, such as indoor and deep indoor, the received P_S could be further attenuated by more than -30 dB comparing to that in Table 5-10. In such case, the B_n has to be reduced to 1 Hz or even below, corresponding to coherent integration of the IF samples for 1000 msec or longer to ensure a p_d of 99.5% with a p_{fd} of 0.1%.

The signal detection usually consists of multiple trials, and the overall probability of successful signal detection will be an aggregation of the above single trial p_d, and the overall probability of false alarm will be an aggregation of the above single trial p_{fd}.

5.2.2 Signal Acquisition

In the search of a signal, the full range of time uncertainty is 1 msec, corresponding code phase of 0 to 1022 chips; the full range of frequency uncertainty is +/- 10 KHz. At the assumed signal level, the signal search is in time and Doppler domains, with a cell referring to the combination

of a code bin and a Doppler bin. The proper selection of code bin and Doppler bin is crucial, not only for fast TTFF, but also for acquisition sensitivity. For the length of integration of the correlation samples used for signal acquisition, it is referred to as dwell time. The Doppler bin is dependent on the dwell time in unit of seconds, with longer dwell time resulting to smaller Doppler bin size, empirically selected as ~2/3T Hz. During each dwell time T, the I &Q components are integrated and dumped, then the envelope $\sqrt{I^2 + Q^2}$ is computed and compared with the selected threshold (Section 5.1.1) to determine the presence of signal.

Keep performing acquisition through non-coherent integration on successive 1ms of data and summing the power until the results exceed a certain threshold, which enables to declare signal detected, or the data length reaches a maximum length, which enables to declare signal absence. Usually, strong signals can be detected in 1-ms sample, and weak signals have to be detected in much longer data length. In case of phase shift by NAV bits, only the 1ms data with the phase shift will be affected.

Usually, after coherent integration of every N-ms correlation samples, the power is integrated (non-coherent). The start and stop boundaries of integrate and dump should not straddle the data bit boundary because each time the data bits change signs, the signs of subsequent integrated I and Q samples may change. If the boundary is straddled and there is a data bit transition, the result of integrate/dump for the interval will be degraded, and in the worst case, the signal will be cancelled for the interval if the data transition occurs at the middle point. A short PDI can ensure

that most integrate/dump do not contain a bit transition boundary. Performance degradation has to be accepted until the bit sync is obtained.

5.2.3 Prediction of Present Signals

Under weak signal condition, the used code bin and Doppler bin for signal search are usually much smaller than under normal signal condition, therefore the full range search of a signal is time consuming even given the most advanced hardware.

In order to speed up the signal acquisition, it is desired to narrow down the signal search ranges. Given the knowledge of UE position, time, and SV orbit/clock of certain accuracy level, it is possible to predict the code phase and Doppler of received signal at corresponding accuracy. The method to predict ToT (code phase) has been given in Figure 4-29, in which the UE is assumed known at $\mathbf{r}_r$, and satellite position and clock can be derived as $\mathbf{r}_s$ and Δt_{sclk}.

The achievable accuracy (uncertainty) in ToT depends on the accuracy of given ToR, r_r, and the accuracy in the r_s and Δt_{sclk}. The above process is actually the reverse process of receiver position fixing, in which r_r and ToR (including receiver clock bias Δt_{rclk}) are to be estimated based on given ToT, r_s and Δt_{sclk}. Once the ToT is obtained, then the Doppler can be predicted in terms of the following equation:

$$D = \frac{\mathbf{r}_s - \mathbf{r}_r}{\left|\mathbf{r}_s - \mathbf{r}_r\right|} \cdot (\dot{\mathbf{r}}_s - \dot{\mathbf{r}}_r) + \Delta \dot{t}_{rclk} \tag{5-5}$$

The sub-millisecond portion of predicted ToT corresponds to the code phase of received signal, which is in unit of chips once it is multiplied by 1023; and predicted Doppler is in unit of m/s and can be converted to Hz as soon as it is divided by the carrier wavelength (0.19 m).

In the above prediction, the satellite orbit and clock can be usually derived from almanac or ephemeris. However, the errors grow rapidly with the increase of age in almanac, and when ephemeris is used beyond the validity time window, the error also grows rapidly with time. Some examples have been given in Figure 4-23 to illustrate how the errors grow with the age of almanac and ephemeris:

With the long validity ephemeris available, it is obviously an advantage in the prediction of above ToT and Doppler regarding the achievable accuracy.

Therefore, based on the accuracies in the ToR, r_r, r_s and Δt_{sclk}, the uncertainty ranges of predicted code phase and Doppler can be determined, which are then used to define the search window in signal acquisition. Narrower search window requires smaller uncertainty ranges in the prediction, which eventually turns into the requirement on better accuracies in the ToR, r_r, r_s and Δt_{sclk}.

5.2.4 Coherent and Non-coherent Integration

Consider complex signal sample with constant signal component s[n] and additive white Gaussian noise w[n] of variance σ_w^2:

$$x[n] = s[n] + w[n] = A \cdot e^{j\theta} + w[n] \tag{5-6}$$

Then the SNR of the signal is:

$$SNR_1 = \frac{A^2}{\sigma_w^2} \tag{5-7}$$

In coherent integration, multiple samples are directly summed up, so both the signal amplitude and the phase of the data are utilized:

$$\sum_1^N x[n] = \sum_1^N (s[n] + w[n]) = N \cdot A \cdot e^{j\theta} + \sum_1^N w[n] \tag{5-8}$$

Therefore, the SNR after integration is:

$$SNR_c = \frac{(N \cdot A)^2}{N \cdot \sigma_w^2} = N \cdot SNR_1 \tag{5-9}$$

And the resultant coherent integration gain is:

$$G_c = \frac{SNR_C}{SNR_1} = N \tag{5-10}$$

It should be noted that, obtaining this gain requires the signal samples to be added in phase so that the signal component power increases by a factor of N^2, while the power of the integrated noise increases only by a factor of N.

However, in non-coherent integration, instead of the samples, the envelopes of samples ($|x[n]|$) are summed up, and thus the phase information is discarded before integration:

$$\sum_{1}^{M}|x[n]| = \sum_{1}^{M}|A \cdot e^{j\theta} + w[n]| \tag{5-11}$$

Directly computing the SNR of non-coherent integrated signal is impossible because the above summation cannot be decomposed into signal only and noise only components. Approximation of the non-coherent integration gain has been studied in lots of literatures (Richards, 2010;). Without going into details, the non-coherent integration gain is expressed in the form below:

$$G_{NC} = N^{\alpha} \tag{5-12}$$

Where the α ranges from 0.5 to 1.0, and it is often taken as 0.5 in the SNR gain analysis by many literatures.

Comparing to Equation (5-10), non-coherent integration is not as efficient as coherent integration in the sense that it takes a larger number of samples N to achieve a given integration gain than is required for coherent integration.

Comparison of coherent integration and non-coherent integration can be further extended to noise performance and computation complexity, but beyond the scope of this Chapter. Based on the analysis so far, the following conclusion can be made:

(1) Coherent integration is subject to the phase change in the samples, whereas non-coherent is not

(2) Coherent integration is more efficient in improving SNR when extending the length of samples for integration.

5.3 Signal Tracking with Long Validity NAV Messages

At the end of signal acquisition, the beginning of the present C/A code is found and the carrier frequency of the input signal is known. In order to keep the local replica code and local replica carrier synchronized with the received C/A code and carrier, it is necessary to transition to next stage – signal tracking. The tracking of code corresponds to a code tracking loop; and the tracking of the carrier corresponds to a carrier tracking loop. Carrier tracking loop is further implemented through carrier frequency tracking or phase tracking.

5.3.1 Tracking of Weak signals

As depicted in Figure 5-37, after the signal acquisition, the carrier (including the carrier Doppler) is stripped off the digital IF by the replica carrier signals to produce in-phase (I) and quadrature-phase (Q) sampled data, right before the code is wiped off by the local replica of code, which is actually a process of de-spreading the spectrum.

The signal processing after the IF signal has been converted to baseband by the carrier and code stripping processes, but prior to being passed through signal discriminators, is referred to as pre-detection. The pre-detection stage consists of integration of correlation results over multiple milliseconds and dump to signal discriminators, with the time interval referred as integrate and

dump interval (PDI) and the bandwidth denoted by B_{IF}. The pre-detection bandwidth B_{IF} is reversely proportional to the length of PDI, an example of which is, a PDI of 1 ms (length of a C/A code) corresponds to the B_{IF} of 1 kHz.

The PDI must be as long as possible under weak signal condition, with the SNR gain expressed in Equation (5-13).

$$gain(dB) = 10 \times \log_{10}(N) \tag{5-13}$$

Although extending the length of PDI is able to improve the signal tracking sensitivity, it is subject to a lot of challenges, such as:

(1) Bit synchronization.

Before bit edge is located, PDI is limited to 1 ms.

(2) Unknown bits

After Bit synchronization, however, if the data bits are unknown, pre-detection integration cannot go beyond a bit boundary, and therefore PDI is limited to 20 ms.

(3) Dynamics bandwidth (B_d)

With data bits are known, pre-detection integration can be extended across multiple bit boundaries to obtain much higher gain in SNR, however, it is limited by the dynamics bandwidth

(B_d) which is the composition of UE dynamics, SV dynamics, ionosphere scintillations, oscillator oscillation with temperature.

The objective of a code tracking loop is to keep the code phase between replica code and the incoming SV code phase at zero. A code tracking loop is realized through a delay lock loop (DLL), usually consisting of Early(E), Prompt(P) and Late(L) correlators, integrate &dump, code phase discriminator, loop filter and code Numerically Controlled Oscillator (NCO). With the integrated correlation results of E/P/L channels dumped to the discriminator, the error in replica code phase is roughly estimated and then passed to a loop filter for further noise reduction. The filtered error is then fed back to code NCO to generate the local replica of the code, so that the discriminator output is driven to zero and then the correlators produce maximum correlation.

The objective of the carrier tracking loop is to keep the phase error between the replica carrier and the incoming SV carrier signals at zero. Any misalignment of the phase is detected and corrected by the carrier tracking loop. In a phase lock loop (PLL), when it is phase locked, the I signals are maximum (primarily signal) and the Q signals are minimum (primarily noise).

In coherent integration, the correlation components I and Q are separately integrated over multiple code lengths (N) and are then dumped. For each dump of I and Q, the signal power can be computed. The SNR gain using coherent integration has been given in Equation (5-13). However, although coherent integration is able to efficiently reduce noise, its performance is limited by the bit boundaries. If the integration it straddles bit boundaries, the integrated correlation results may lose the overall power.

In non-coherent integration, multiple signal powers (M) are summed up and then used for signal detection. When going from coherent integration to non-coherent integration, the change in SNR is often referred to as squaring loss (Strassle, 2007; Diggelen, 2009). Based on Equation (5-13), the SNR gain in dB from non-coherent integration is given below. This technique is less or not subject to change in navigation bits, but subject to squaring loss, therefore less effective in reducing noise than coherent integration.

$$gain(dB) = 10 \times log_{10}(M^{\alpha}) \tag{5-14}$$

In a real system, the use of different combinations of Coherent integration (N) and Non-coherent Integration (M) are the primary ways to improve SNR. An empirical Equation of overall SNR gain for such a system is given below:

$$gain(dB) \approx 10 \times log_{10}(N) + 10 \times log_{10}(\sqrt{M}) \tag{5-15}$$

5.3.2 Prediction and Maintenance of Alignment with Bit Boundaries

In order to enable long coherent integration across bit boundaries, the NAV bits in the received signal have to be wiped off in advance. When the NAV bits are known, however, to wipe off the NAV bits in the received signal requires that the boundaries of the known NAV bits must be aligned with the ms samples (as illustrated in Figure 5-39), so that given a series of ms samples, the right NAV bits can be found to multiply.

It should be noted that, the alignment discussed here is different from bit synchronization discussed in Chapter 4. Although bit boundaries are determined in bit synchronization, the wipe-off of NAV bits is still impossible because it is unknown yet which bits the boundaries correspond to.

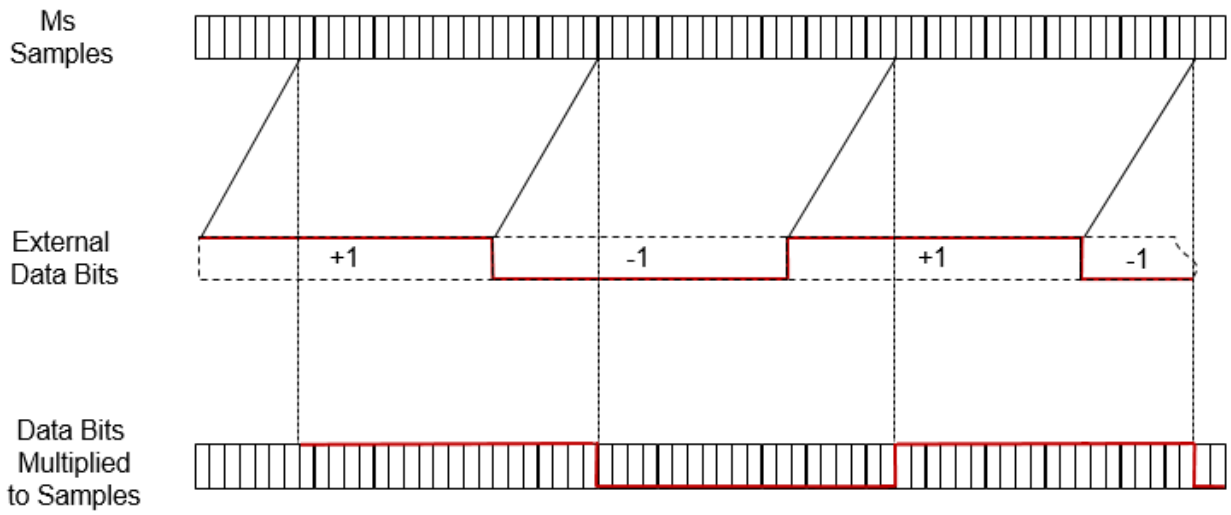


Figure 5-39 Alignment between Aided NAV Bits and Msec Samples during Data Wipe-Off

Then how to align the boundaries of the NAV bits with the ms samples? Depending on the availability of time, position and SV ephemerides in the UE, the methods for such alignment that are studied in this section are:

(1) Prediction of ToT based on precise ToR

In the context that long validity NAV messages are available, SV ephemerides are certainly always available. So when UE location is known within certain uncertainty range (say <150km), and UE time is known within uncertainty level (say <0.5 ms), for any signal arriving at UE at precise ToR, it is possible to use the method depicted in Figure 4-29 to derive the precise ToT.

For the correlation samples collected millisecond by millisecond, they are always tagged with millisecond counters from local clock in a UE, and can be referred to as MsCount. Considering that the boundaries of NAV bits correspond to multiple of 20-ms in ToT (GPS IS, 2004), for a msec sample time-tagged with 'MsCount', if the corresponding ToT is precisely determined, then the location of current ms sample in current NAV bit (referred as Epoch Number) can be determined by taking the modular of 20 ms on the ToT (expressed in milliseconds), as illustrated in Equation (5-16):

$$EpochNum = mod(ToT_msec, 20) \tag{5-16}$$

With the Epoch Number determined, it is straightforward to get the right MsCounts that correspond to bit boundaries. By using the part of multiple 20 ms in the ToT, it can be determined that the boundaries correspond to the start of which NAV bits.

(2) Fast time synchronization through bit pattern matching

In the scenario that precise ToT is not available, however, some NAV bits have been decoded from the received signals, it is possible to align the ms samples with the boundaries of the right NAV bits through fast frame synchronization. The method is describe in terms of example given in Figure 5-40, where the UE time is known at the uncertainty of 1 second and 10 NAV bits have been decoded.

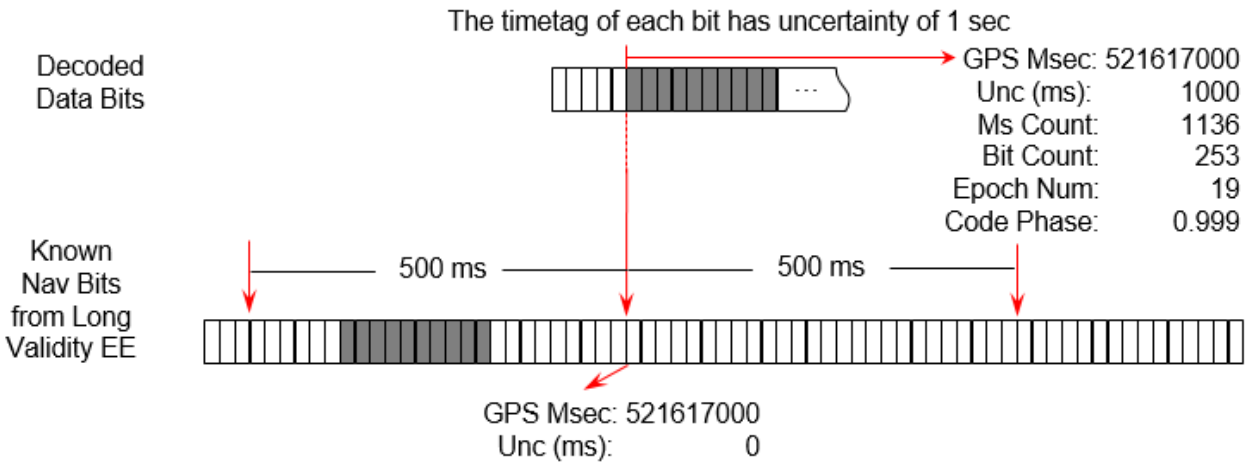


Figure 5-40 Fast Time Synchronization Through Pattern Matching of A Few Decoded Bits in the Known NAV bits

Then fast time synchronization can be obtained through the following steps:

- Derive the GPS Msec of the starting edge of the decoded data bits, and determine the uncertainty. In the example given in the above figure, the first starting edge of first decoded bit correspond to GPS Msec 521617000, with 1000 msec uncertainty.
- Locate the bit in the known NAV bit stream that corresponds to the GPS Msec 521617000
- Determine uncertainty range of bits in the known NAV bit stream given the time uncertainty in the starting edge of the first decoded bit, considering that 1 bit corresponds to 20 msec.
- Slide bit by bit within the uncertainty range in the known NAV bit stream, and locate the same number of bits that match (or complementarily match) the pattern of the decoded data bits
- Once the matched bits are located in the known NAV bit stream, derive the GPS Msec for the starting edge of the first bit, which is 521616620 in the above example as illustrated by the shadowed bits.
- So the starting edge of the first decoded bit corresponds to GPS Msec 521616620.

- In the above search, there could be occurrences of multiple matches. When it happens, just add more decoded bits in the pattern matching. The longer of decoded bits are used in the pattern matching, the less chance for multiple matches.

With the above time synchronization obtained, the right NAV bits corresponding to the boundaries can be retrieved from the long validity NAV messages to wipe off the NAV bits in the received signal.

(3) Regular frame synchronization

For the data bits, the boundary of each bit corresponds to a specific GPS time (expressed in GPS Msecs). To be able to use those bits, the frame sync must have been obtained in the received signal, so that the aiding bits can be aligned with the received signals to wipe off the navigation bits.

(4) Maintaining the previously aligned bit boundaries

Once the alignment between NAV bits and the ms samples has established through the above methods, it is no longer necessary to predict such alignment for every subsequent NAV bits unless the tracking loop is reset. So to be able to perform the wipe-off of NAV bits in received signal, it is only needed to properly maintain the counter for ms samples and the counter for NAV bits (Figure 5-41).

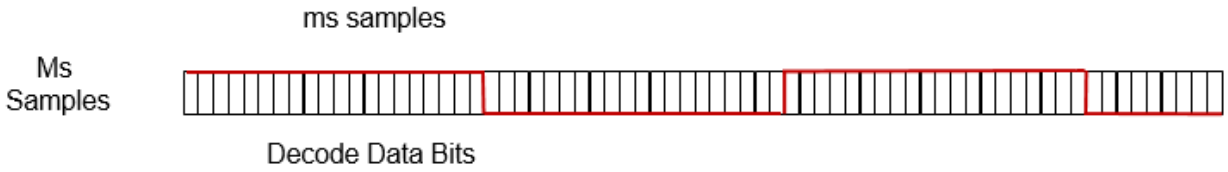


Figure 5-41 Maintaining the Alignment of NAV Bits and Ms Samples

However, the established the alignment between NAV bits and the Ms samples is subject to change due to the drift in the bit boundary, as illustrated in Figure 5-42. In the case the alignment is not properly handled for such drift, the consequence would be serious. The tracking loop could experience loss of integration power, and incur errors in computed pseudorange measurements at magnitude of msec level.

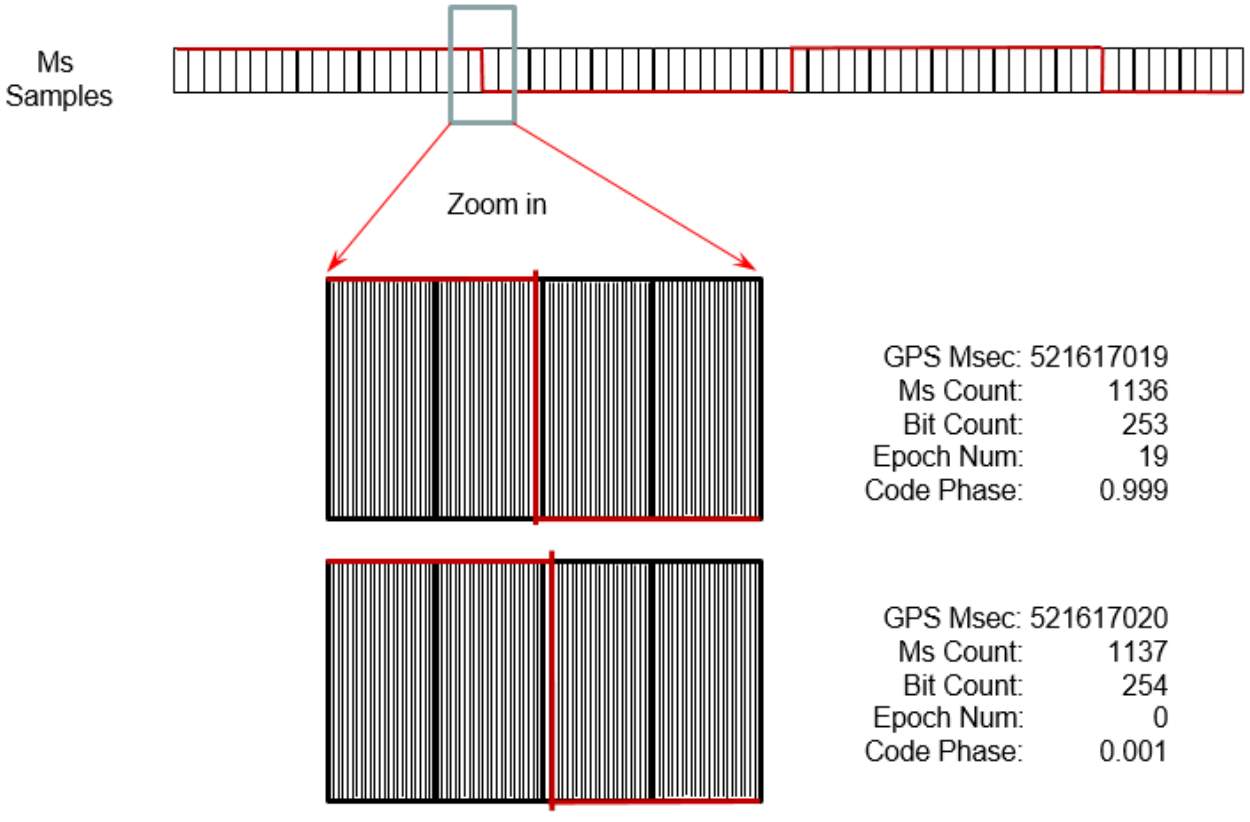


Figure 5-42 Taking Care of the Drift of Bit Boundary

5.3.3 Bit Decoding in Weak Signal Condition

As mentioned in Section 5.2.1, once in stable tracking state, the I components are maximum (primarily signal) and the Q components are minimum (primarily noise). Then the accumulated I_{PS} over a bit interval can be taken to decode the corresponding NAV data bit. The errors in decoded NAV bits is quantified in terms of the bit error rate (BER), which is the ratio of bits in error to total number of decoded bits. According to the theory of CDMA BPSK modulation, the probability of decoding error in a single bit, p_e, is of the form:

$$p_e = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{N_0}}\right) = \frac{1}{2} erfc\left(\sqrt{(C/N_0)/R_b}\right) \tag{5-17}$$

where $erfc(\cdot)$ is the complementary error function introduced earlier; $\frac{E_b}{N_0}$ is a normalized signal-to-noise ratio (SNR), also known as SNR per bit, and R_b is the data rate in bits per second.

The BER for the C/A code signal has been plotted in Figure 2-7. So as seen in the above figure, with the drop of signal strength, the BER increases to a level that NAV bit decoding is unreliable. Per Braasch et al. (1999), usually when CN0 is below 27 dB-Hz, reliable NAV bit decoding is regarded not possible.

In the implemented NAV messages, each NAV word consists of 30 bits, then the overall BER is binomially distributed and given by:

$$BER = \frac{32!}{4! \cdot 28!} \cdot p_e^4 \cdot (1 - p_e)^{28} \tag{5-18}$$

The NAV bits separately decoded are further passed to parity algorithm, which claims to be able to detect up to 3 simultaneous bit errors and can correct 1-bit error (DFC, 2007).

In the plots below, the signal conditions are given from a realistic scenario, where a vehicle was driving through the downtown San Francisco and experiencing typical weak signals and dynamic change in signal conditions:

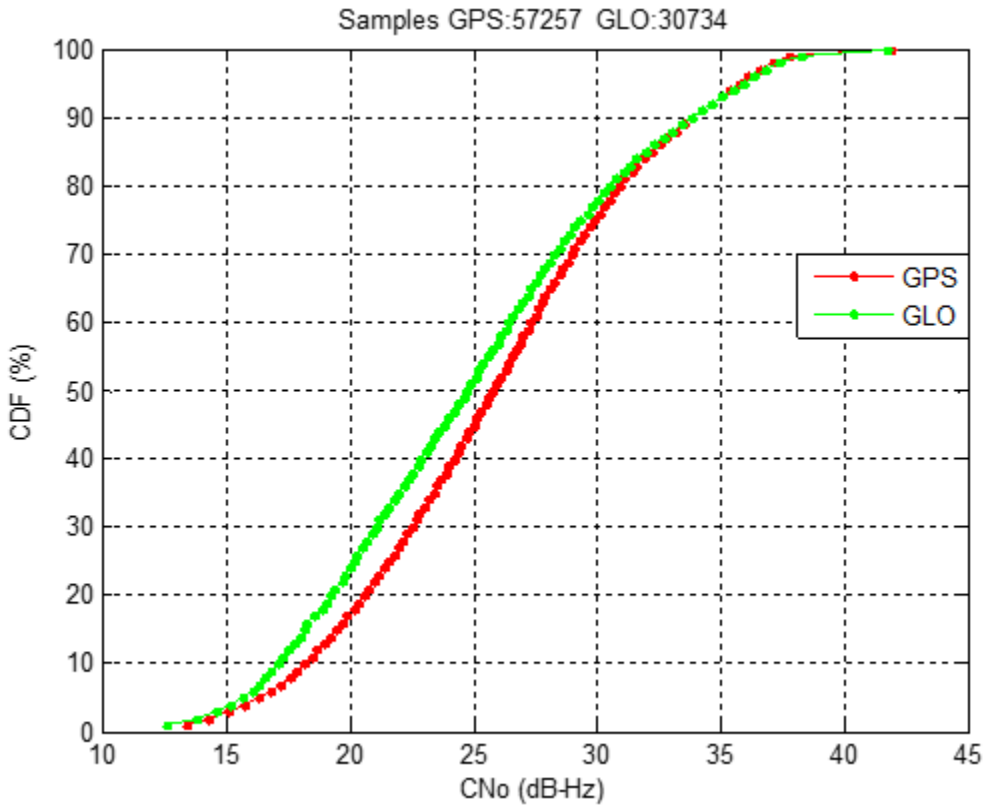


Figure 5-43 CN0 Distribution in Urban Canynon from a Realistic Driving Test

As seen in the above Figure, 50% of GPS signals and 45% of GLONASS were below 25 dB-Hz when driving through the urban canyon. To be able to successfully decode a full subframe of NAV message from the received signal, it is required to have the signal continuously stronger than a certain threshold (say 27 dB-Hz) so that the BER is sufficiently low. In Figure 5-44, the statistics of time is given for CN0 to stay continuously stay stronger than 27 dB-Hz.

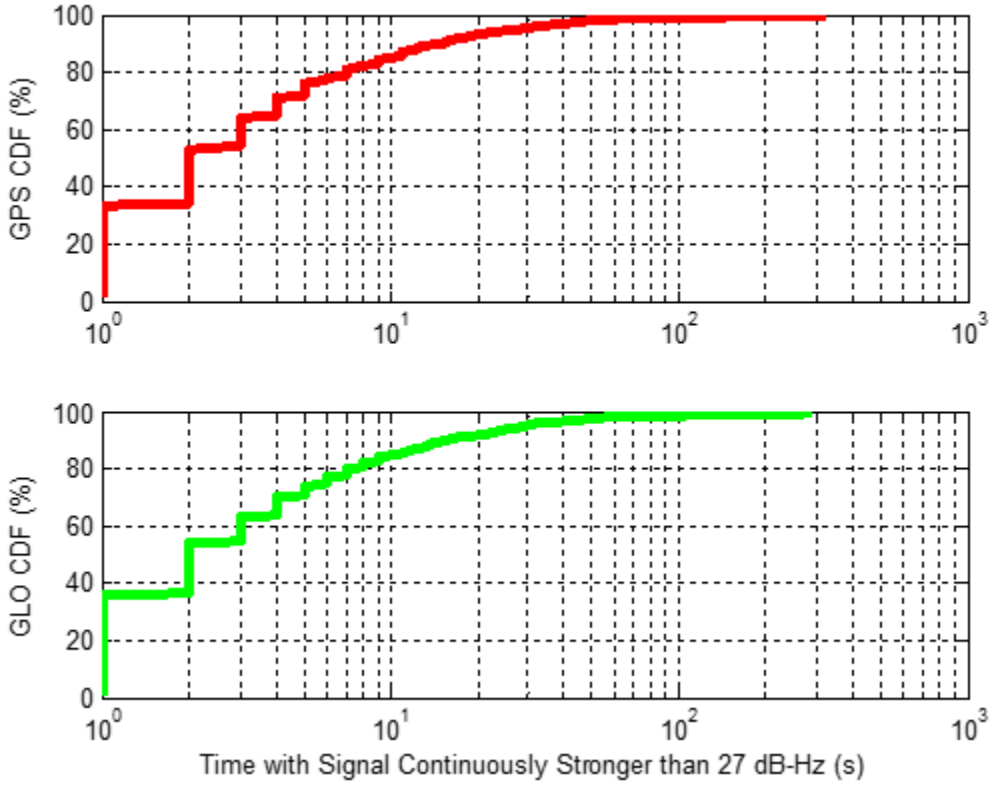


Figure 5-44 Statistics of Time for CNo being Continuously Stronger than 27 dB-Hz

As seen, 80% of the time, the signals cannot stay stronger than 27 dB-Hz for continuously 6 seconds, which means 80% of the time, decoding a full sub-frame of NAV message would be difficult.

With the long validity NAV messages, the need to decode data bits is relieved, which is of great significance especially in weak signal condition. In other words, there is almost no impact on the ephemeris decoding on UE by the weak signals, as there is almost no need to decode ephemeris when long validity NAV messages are available. So from this perspective, the long validity NAV messages are equivalent to improvement in ephemeris decoding sensitivity.

5.4 Position Fix with Long Validity Ephemeris under Weak Signal Condition

UE of normal sensitivity is almost not operational in tough signal environment, because of the challenges on different levels: Firstly, the GNSS signals are difficult to detect and track; secondly, even if the signals are tracked, the navigation messages are difficult to decode because the signals are too weak; lastly, even if the ephemerides are decoded or obtained through other means, position fixes are difficult to obtain because of too many outliers in the measurements.

Therefore, the improvement in the acquisition and tracking sensitivity, and the availability of long validity ephemeris together open the Pandora box for position fixes in such tough signal environment. In such context, the position fixes are made possible at the cost of extremely increased complexity wherever it is not seen in the position fix in normal sensitivity.

The weak signals are usually from attenuation in signal penetration, reflection and interference. For weak signals, it is difficult to discriminator direct signal with strong attenuation, and purely reflected signal. A challenging scenario is illustrated in Figure 5-45, where most of the GPS satellites are blocked, however the signals are still being received and tracked. In such

environment, it is very common that the received signals are either direct signals accompanied with strong multipath, or totally reflected signals.

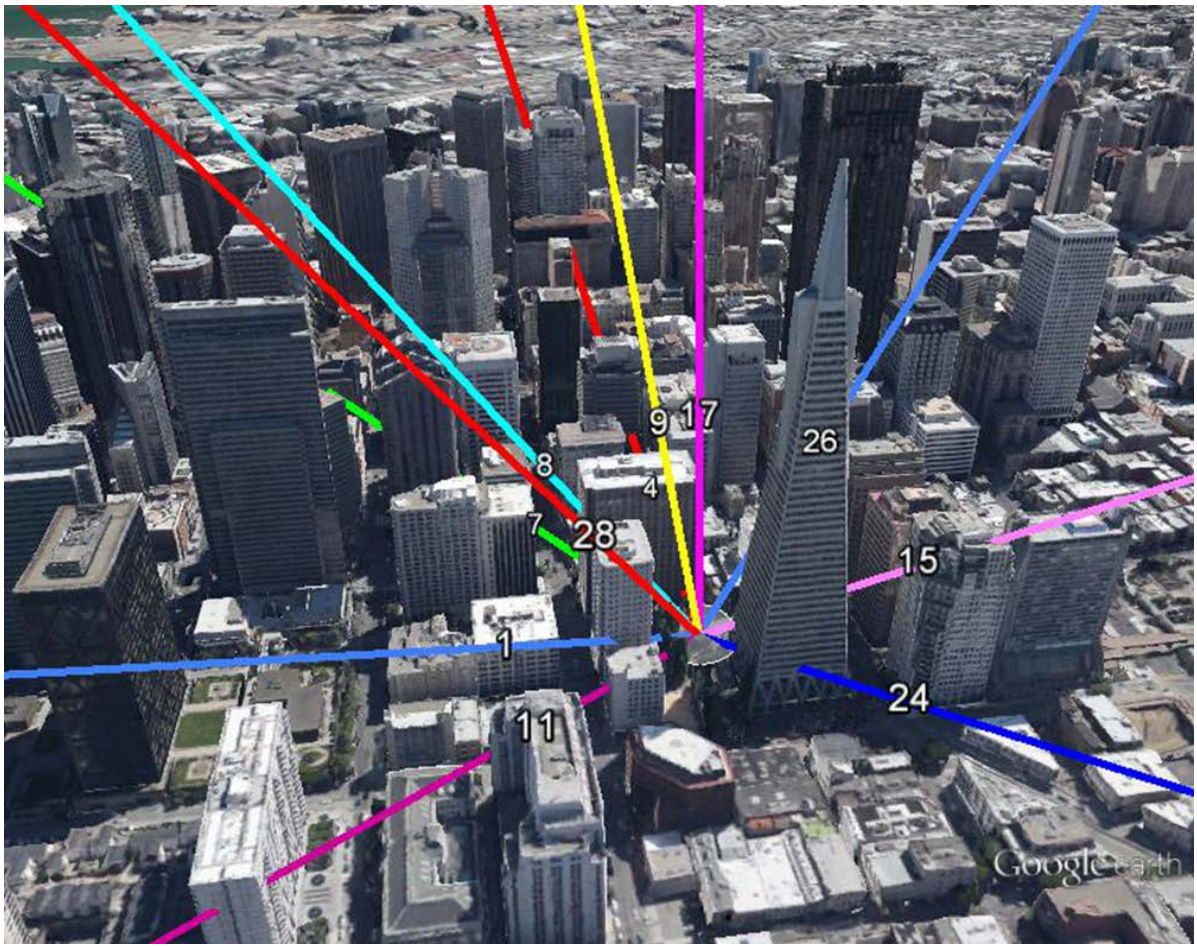


Figure 5-45 GPS Signal Blockage in Urban Caynon

Things could become even more complicated when considering the large variety of the source of measurement errors (large outliers) in weak signal environment:

Table 5-11 Sources of Large Errors in High Sensitivity Scenario

Source	Symptoms / Causes	Consequences

False acquisition	- Signal is acquired in the name of a wrong SV due to cross correlation; - A wrong time-frequency cell is picked.	- Failure in position fix; - Large measurement errors.
False tracking	- Signal is tracked in side-lobes	- Large measurement errors.
False bit sync	- The wrong bit edge is detected	- Multiple of 300 km errors in pseudorange.
False frame sync	- The start of frame is incorrectly determined because of improper time transferring	- Error of 6000 km in pseudorange.
False bit decoding	- False parity check	- Wrong ephemeris parameters; - Loss of signal lock, if falsely decoded bits are further used bit aiding in tracking loop.
Multipath signals	- Purely reflected signals; - Composition of direct and reflected signals.	- Large measurement errors; - Failure in position fix.

5.4.1 Challenge in Discriminating Reflected and Direct Signals

The successful acquisition and tracking of weak signals and the availability of long validity ephemeris together make it possible for position fixes under weak signal condition. However, the complexity in position fixes drastically increases under weak signal condition comparing to with regular signals.

The complexity comes from the difficulty in discriminating direct signals and reflected signals, heavily attenuated direct signals and purely reflected signals. The signal reflection is dependent on signal frequency (carrier wavelength), the reflection coefficients and smoothness of the material surface, and the incident angle. The reflection coefficients of typical exterior wall

surfaces are available in lots of earlier research. According to Landron et al. (1996), at the incident angle of 30 degree, the reflection coefficient of glass for RF signal of 1.9 GHz could be 0.6 (corresponding to -2.2 dB). And, per Hein et al., 2008, when GPS L1 signals propagate through tinted glass of 4-mm thickness, the attenuation could amount to -24.44 dB. Given such facts, in tough signal environments, the direct signals could be even weaker than the purely reflected ones.

5.4.2 Challenge in Outlier Detection Algorithms

Conventional RAIM algorithms assesse consistency among pseudorange estimates and discards those that are significantly inconsistent from the majority (Parkinson and Axelrad 1988; Brown 1992; Walter and Enge 1995). Traditional outlier or blunder detection methods are based on the assumption that the measurement errors follow Gaussian distribution and majority of the measurements are good. Here is a brief description of the fundamental theory for the traditional methods, that consists of steps:

(1) Global statistical test

This is to check whether any outlier exists among the measurements through a statistical test on the square root of the sum of squares:

$$Q = \left(\frac{v_1}{r_1}\right)^2 + \left(\frac{v_2}{r_2}\right)^2 + \ldots + \left(\frac{v_n}{r_n}\right)^2 \tag{5-19}$$

Where v_i is the i-th measurement residual, and r_i is the square root of the corresponding variance, and n is the number of measurements.

Without a blunder, ideally the following equation exists:

$$Q \sim \chi^2(n-m) \tag{5-20}$$

Where m is the number of the unknowns.

By selecting empirical thresholds for the above statistic tests according to the degree of freedom (n-m), conclusion can be made whether any blunder exists in the measurements.

(2) Local statistical test

If the above statistic test concludes that blunder(s) may exist, it is necessary to locate the blunder(s) among the measurements through a local statistical test. For good measurements, ideally the following equation exists:

$$\frac{|v_i|}{r_i} \sim N(0,1) \tag{5-21}$$

An outlier is declared if the above statistics falls beyond 3 sigma or an empirical threshold, and then it is de-weighted accordingly or even completely removed.

However, in weak signal conditions, especially in the challenging environments like urban canyon, the above assumption of measurement errors may not be true. See an example of pseudorange errors in downtown area (Figure 5-46), where actually the majority are outliers.

After some outliers being used and biased being introduced in position estimation, the detection of outliers becomes even more difficult, or even ends up with false detection – namely good measurements are detected as outliers. The consequence is, the actual outliers are trusted and given high weight, but the good measurements are heavily de-weighted, and the position fix is stuck in a 'hole'. Considering the complexity of weak signal scenarios, traditional outlier or blunder detection algorithms no longer work, and new techniques, including heuristic methods are necessary to get measurements weighted as properly as possible.

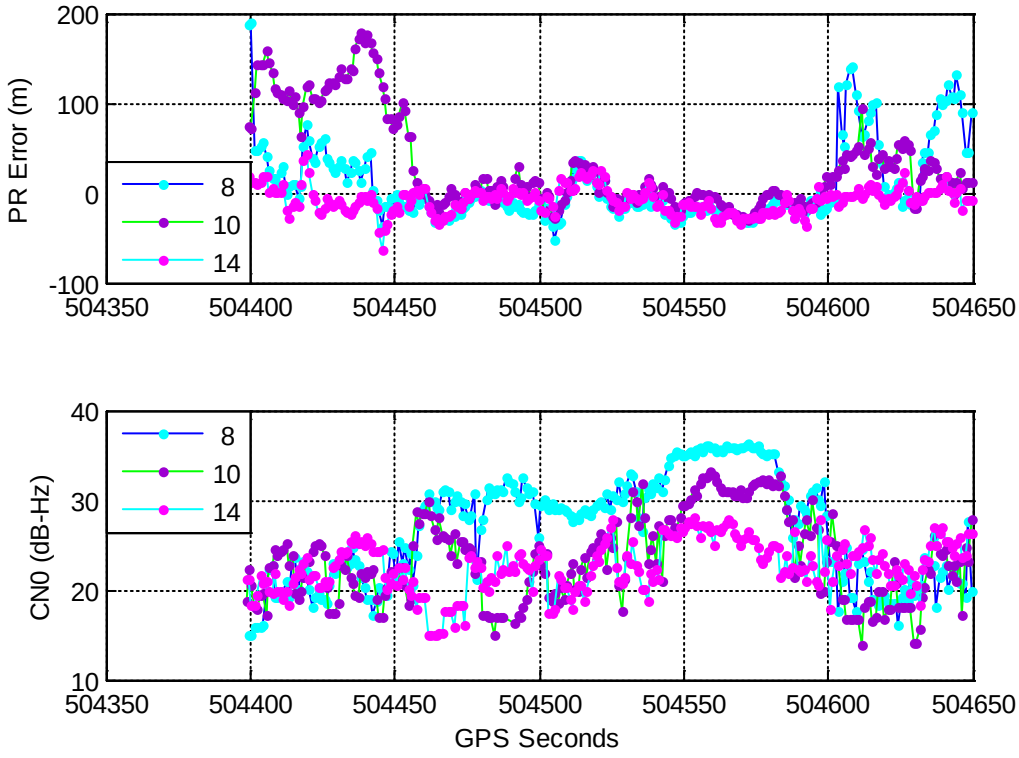


Figure 5-46 Example of Measurement Errors vs. CN0 in Urban Canyon

As mentioned in Section 5.3.2, the CN0, which is used to indicate the measurement quality, is no longer able to exactly reflect the actual measurement quality under weak signal scenarios, and

very often could even mislead the estimation. So when the measurements, with the weights assigned according to CN0 being used, the resulting position estimation could be totally biased, and the corresponding uncertainty could be no longer trustable.

Therefore, in the context of high sensitivity, it is necessary to seek new methodologies for outlier detection so that reliable position estimation could be achieved with the measurements obtained through high sensitivity techniques in signal acquisition and tracking. Herein, as part of the innovations in this thesis, 2 novel and practical methods will be brought up in the next 2 sub-sections.

5.4.3.1 Guard range for measurements to be used in positioning

In the case some a-prior information of the starting position is available, the level of uncertainty in the starting position can be translated into the capability of preventing outliers of corresponding level from being used in position estimation and deteriorating the position fix. Considering that the improper weighting might be imposed to the outliers, if detected and identified, the best option to isolate the impact is to exclude the outliers from position estimation.

In order to stop the measurement outliers from being used, a guard range for each measurement is adopted here. For each measurement, only when it falls within the guard range, it is allowed to be used in the position estimation.

To prepare the guard ranges for the measurements, the uncertainty of the starting position is projected to the Line-Of-Sight (LOS) of each measurement:

$$\left[\Delta r_1 \quad \Delta r_2 \quad ... \quad \Delta r_n\right]_{1\times n}^{T} \tag{5-22}$$

Where Δr_i represents the impact of the position offset on the calculated range along the LOS direction of the i-th measurement.

In order to remove the impact of UE clock, a reference SV is selected and denoted as '**0**'. Single difference is taken between each measurement and the reference measurement, and the following delta-measurements are derived:

$$\left[\Delta\rho_{01} \quad \Delta\rho_{02} \quad ... \quad \Delta\rho_{0n}\right]_{1\times(n-1)}^{T} \tag{5-23}$$

Similarly, the delta-range can also be calculated between each SV and the reference SV:

$$\left[\Delta R_{01} \quad \Delta R_{02} \quad ... \quad \Delta R_{0n}\right]_{1\times(n-1)}^{T} \tag{5-24}$$

Considering the impact of the uncertainty in the starting position used in the above calculation:

$$\Delta r_{0i} = \sqrt{\Delta r_0^2 + \Delta r_i^2} \tag{5-25}$$

So the guard ranges for the measurements can be:

$$\left[\Delta R_{01} \pm s\cdot\Delta r_{01} \quad \Delta R_{02} \pm s\cdot\Delta r_{02} \quad ... \quad \Delta R_{0n} \pm s\cdot\Delta r_{0n}\right]_{1\times(n-1)}^{T} \tag{5-26}$$

Where '*s*' is a selected scale factor, which by default could be selected to be 1. For each measurement, it is used in the position estimation only when its delta-measurement falls within its guard range. The success of this method is highly dependent on the qualification of the reference SV, so derived varieties of this method can be developed to make sure a SV of the best measurement quality is selected as the reference.

5.4.3.2 Classification of Residual Clusters

In some urban canyon with dense high-rise buildings, majority of the measurements could suffer from multipath signals, where errors of magnitudes amounting from tens of meters to a few kilometers are very common, but the CN0 could hardly provide an indication of the measurement quality.

In addition to the method introduced above, another new method for the selection of measurements is illustrated in this section. It should be noted that, this method focuses more on selecting the relatively good portion of the measurements in the position estimation than on detecting and excluding the outliers.

In Figure 5-47, the residuals from different SVs are illustrated. The calculated residual for each SV includes both the impact of position offset and the receiver clock bias. As the receiver clock bias is a common part in the residuals from different SVs, it is not given special attention here.

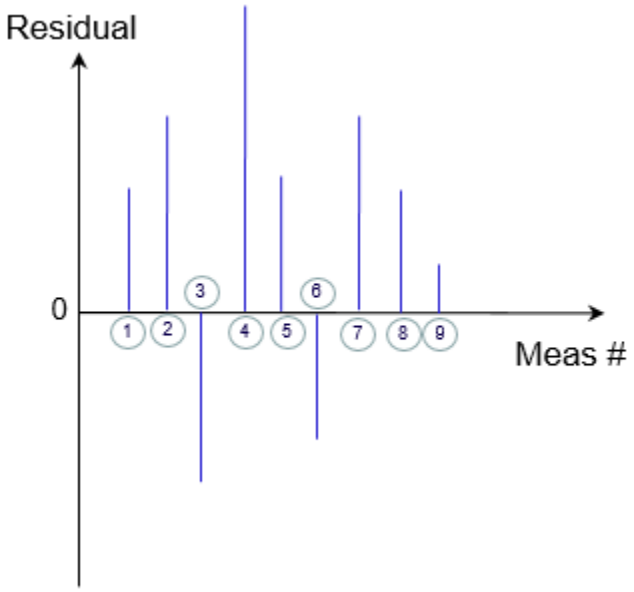


Figure 5-47 Illustration of Measurement Residuals

Given the fact that the CN0 of the signal is not a trustworthy indication of the measurement quality, the selection of the proper measurements for position fix consists of 3 steps:

(1) Sorting of residuals

As illustrated in Figure 5-48, the residuals are sorted in terms of the value of the residuals.

(2) Cluster classification of residuals

The purpose of this step is to classify the residuals into multiple clusters, according to the magnitude of the separation between the neighboring sorted residuals. In the classification of the residuals, the number of clusters needs to be determined, as well as which cluster a residual belongs to. For the residuals that are determined belonging to the same cluster, a rule-of-thumb

criteria is that, the standard deviation of the residuals should be within a predetermined threshold, indicating a certain level of consistency among the residuals in the cluster. The selection of proper threshold for the residual standard deviation depends on the uncertainty of the starting position that is used in the residual calculation and the expectation on the position fix that is to be made. Therefore, the strategy of the threshold selection consists of tuning efforts to take into account different use scenarios.

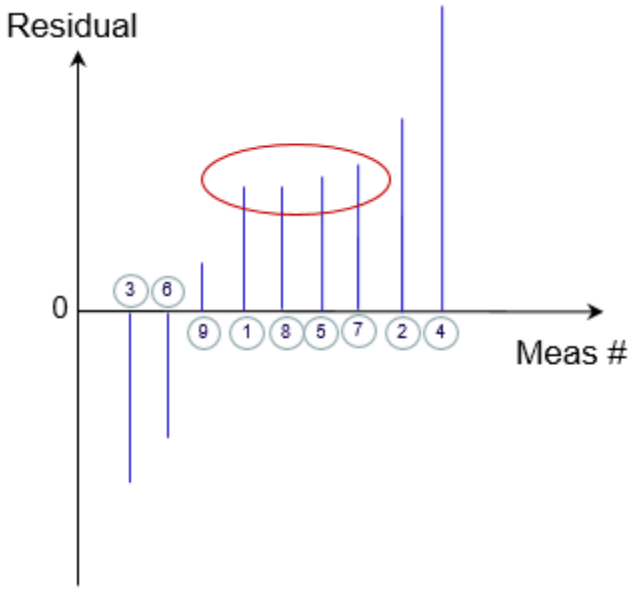


Figure 5-48 Sorting of Measurement Residuals

(3) Selection of the right cluster

The right cluster needs to be selected for the best possible position fix, once the residuals are classified into different clusters. The following procedures can be followed to select the right residual cluster:

- Select the cluster that contains the max number of residuals;

- Select the cluster with the best consistency, or equivalently the smallest standard deviation among the residuals, if multiple clusters have the same number of residuals.

In Figure 5-48, it is obvious that the circled residual cluster has the max number of residuals and should be selected for position fix. When the case happens that no cluster contains more than 1 residual, it indicates that either the starting position if off too much, or the signals are really too bad.

CHAPTER 6 PROPOSAL OF NEW GNSS NAV MESSAGES

This chapter will come up with the details of the proposal for new GNSS NAV messages with long validity period, and further study the best update rate of NAV messages, User Range Accuracy algorithms, and practical implementation issues on GNSS UEs.

6.1 Weakness in the Design of NAV Messages of Existing GNSS

The existing efforts of improving the TTFF and sensitivity on GNSS UEs can be categorized into two different approaches: (1) developing assisting systems, (2) inventing and implementing new algorithms for standalone UEs. No matter which, it is still subject to the limitations that have been fully discussed and listed in the Table 2-4.

The limitations are resulted from the design of NAV messages of current GNSSs. The fundamental cause to the problem with current NAV messages, in this thesis' view, lies in the congenital weakness of the design of the existing GNSS NAV messages.

Taking GPS as an example, the contents in GPS sub-frames 1-3 are updated every 2 hours, although the ephemeris is valid for up to 4 hours. It is challenging and questionable for standalone GPS UEs working in weak signal environments to catch up with such frequent ephemeris updates. Working properly in the past 2 hours does not mean that the UE can work properly in the next 2 hours if ephemerides are not downloaded in time. The NAV messages received 2 hours ago cannot be used for the data aiding in the next 2 hours to improve the

tracking sensitivity. For startups under normal signal conditions, the UEs, if missing the start of sub-frame 1, have to wait 30 s to get to the next sub-frame 1 to download a complete copy of the ephemeris. The successful startups 4 hours ago also do not help much reduce the TTFF in the subsequent startups, as time needs again for ephemeris downloading.

Let's take a look at the NAV messages for other GNSSs, like GLONASS, BDS and GALILEO. The GALILEO F/NAV, available on E5A-I channel at 25 bps, consists of 12 sub-frames lasting 600 s, with each sub-frame composed of 5 pages lasting 50 s. The ephemeris occupies 3 pages, lasting 30 s, taking a very similar form as GPS. The GALILEO ephemeris comprises 17 parameters, including 6 Keplerian parameters and 6 harmonic coefficients, with the content valid for up to 4 hours and updated every 3 hours (GALILEO ICD, 2008). The downloading of GALILEO ephemeris takes at least 30 s, and if missing the start of the first ephemeris page, it will take at least 50 s to get a complete copy. So from this perspective, the GALILEO TTFF for standalone devices is expected to be slower than GPS.

For BDS, the D1 NAV message is broadcast at the rate of 50 bps, with the super-frame structure very similar to the GPS super-frame. The BDS D1 super-frame lasts for 12 minutes, consisting of 5 sub-frames, with the ephemeris and clock broadcast in the first 3 sub-frames and repeated every 30 s, and with the almanacs broadcast in the last 2 sub-frames in 24 pages (BDS ICD, 2013). The BDS ephemeris is similar to the GPS ephemeris, occupying 3 sub-frames but it is updated every 1 hour and can be used for up to 2 hours. It is expected that for standalone BDS UEs, the TTFF is also similar to standalone GPS UEs.

For GLONASS, at the data rate of 50 bps, the super-frame has duration of 2.5 minutes and consists of 5 frames, with each lasting for 30 s and further consisting of 15 strings. Each string has duration of 2 seconds. The ephemeris, regarded as immediate data and repeated in every frame in the GLONASS navigation messages, occupies 5 strings and comprises the Cartesian coordinates, velocity components, and solar/lunar gravitational accelerations at the reference time, with the content valid over about 0.5 hour (GLONASS ICD, 2008). Upon receiving the ephemeris, the navigation device is to calculate the satellite orbit by numerically integrating the motion equations that include the second zonal geopotential coefficients through a fourth-order Runge-Kutta method. The downloading of the GLONASS ephemeris takes at least 10 s, and if missing the start of the first ephemeris string, it will take at least 30 s to get a complete copy of the ephemeris. Therefore, on this point, the GLONASS TTFF for standalone devices is expected to be faster than GPS.

A summary of the NAV messages for GPS, GLONASS (GLO), BDS System (BDS) and GALILEO (GAL) has been given in Table 6-13 for the purpose of comparison. It is in common that, the designed NAV messages for GPS, GLONASS, BDS and GALILEO are valid for short periods, and therefore all of them are subject to aforementioned limitations in attempts to improve TTFF and sensitivity, even with the assisting technologies and systems.

6.2 Considering New GNSS NAV Messages

Instead of providing remedy solutions to overcome the weakness in the design of NAV messages in current GNSSs, this section take a different perspective by considering new design of NAV messages for future GNSS to avoid such weakness.

6.2.1 Proposal of New GNSS NAV Messages

The common weaknesses in the NAV messages of the GPS, GLONASS, BDS and GALILEO have been addressed the last section. Those weaknesses can be overcome, and fast TTFF and high sensitivity can be facilitated through the design of new NAV messages, as long as the following 'rules' are followed:

(1) Update rate, as low as possible
(2) Repeat rate, as high as possible
(3) Length of Eph content, as short as possible
(4) Eph life expectancy, as long as possible

Let's take further analysis on the GPS NAV messages in terms of the above 4 'rules'. The primary contents in the GPS NAV messages include:

(1) Satellite clock
(2) Satellite ephemeris
(3) Ionosphere information
(4) UTC parameters
(5) Almanacs

Currently, two types of atomic clocks, Rubidium and Cesium clocks, at the accuracies of $\pm 1\times 10^{-12} \sim \pm 1\times 10^{-13}$, are used on the GPS satellites (USCG, 2014). So it is possible to have

the clock parameters updated at a longer interval, say 12 hours, without introducing significant errors in the pseudorange observations.

For the GPS ephemeris, the Keplerian parameters are from the fitting of 4-hour orbit curves. The orbit, represented by the Keplerian parameters plus perturbation corrections, gives the overall best fitting performance of the whole orbit segment. But as a longer orbit curve is used in the fitting, it is harder for the fitted orbit to agree well with each small portion of the original orbit. A set of Keplerian orbital parameters can be a good description of the satellite orbit over a short period (say 4 hours), but can hardly be a good approximation of the orbit for a long period (say 24 hours). So frequent update of the ephemeris content is indispensable in order to guarantee the orbit accuracy, and at this point, there is no much room to extend the ephemeris update interval (namely to reduce update frequency).

The ionosphere information included in the GPS NAV messages is actually the Klobuchar model; the UTC parameters are for relating the GPS time with the UTC time; and the almanacs are the rough orbits for all GPS satellites in service. According to GPS IS (2004), all these messages are updated at least once every 6 days, and they are actually observed as often as once per day although the update time is not fixed on each day.

The analysis above indicates that only the ephemeris message (the Keplerian parameters) changes frequently when compared to all other GPS NAV messages. To facilitate fast TTFF and high sensitivity, we therefore only need to find a way to minimize the frequency of the ephemeris update in the GPS NAV messages, according to the aforementioned ‘rules’.

The GLONASS ephemeris gives us a good hint, although the life of the GLONASS ephemeris is only around 30 minutes. For a satellite in space, given the initial conditions (position $\mathbf{r}$, velocity $\dot{\mathbf{r}}$, and etc) as illustrated in Figure 6-49 and also expressed in Equation (6-1) at time t_0, the succeeding orbit, $\mathbf{r}(t)$, can be obtained by integrating the accelerations $\ddot{\mathbf{r}}$ in Equation (6-2), as illustrated in Equation (6-3).

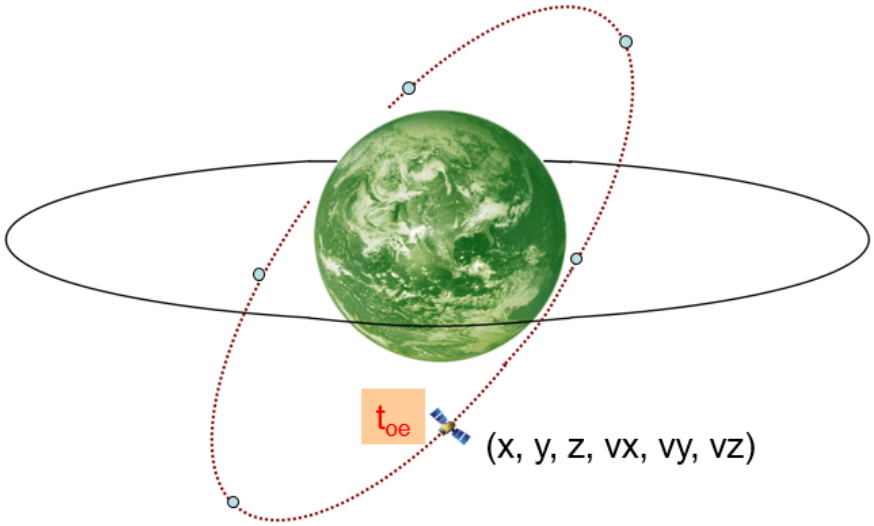


Figure 6-49 Initial Condition of a Satellite Orbit

$$\begin{cases} \mathbf{r}_0 = \mathbf{r}(t_0) \\ \dot{\mathbf{r}}_0 = \dot{\mathbf{r}}(t_0) \\ \ldots \end{cases} \tag{6-1}$$

$$\ddot{\mathbf{r}}(t) = -GM\frac{\mathbf{r}}{|\mathbf{r}|^3} + a(t, \mathbf{r}, \dot{\mathbf{r}}, p_0, \ldots) \tag{6-2}$$

$$\mathbf{r}(t) = \mathbf{r}(t_0) + \int_{t0}^{t}\left[\dot{\mathbf{r}}(t_0) + \int_{t0}^{t}\ddot{\mathbf{r}}(t)dt\right]dt \tag{6-3}$$

To ensure the accuracy of the derived orbit, $\mathbf{r}(t)$, the forces that result in the acceleration, $\ddot{\mathbf{r}}(t)$, on the satellites, should be well modeled (Montenbruck, 2000). The forces can be gravitational and non-gravitational. Zhang et al. (2008) and Han (2009) illustrate a successful ephemeris extension for up to 5 days, with only the non-gravitation force - solar radiation (p_0), and the gravitational attractions from the Earth, the Sun and the Moon modeled.

For all the above gravitational components, given that standard models are embedded on the mobile devices, they can be well modeled independently for years, with the work in Zhang et al. (2008) as the evidence. As to the solar radiation, it is related to the reflectivity and attitude of solar panel of the satellite in space, and fortunately can also be well modeled, with some slow-varying and satellite dependent parameters (Montenbruck, 2000). So, if a set of such parameter(s) along with the satellite position and velocity at certain accuracy levels can be provided once every a longer period (say 1 day), the satellite orbit can be accurately derived on a standalone mobile device with some embedded force models.

At any epoch, the satellite position and velocity expressed in Cartesian form ($\mathbf{r}$, $\dot{\mathbf{r}}$) can be also identically expressed in Keplerian form through a set of elements (*e, a, i,* Ω*,* ω*,* ν), as illustrated in Figure 6-50, where:

e Eccentricity of orbit

A Semi-major axis of orbit

i Inclination of orbital plane

Ω Longitude of the ascending node

ω Argument of perigee

ν True anomaly at epoch t

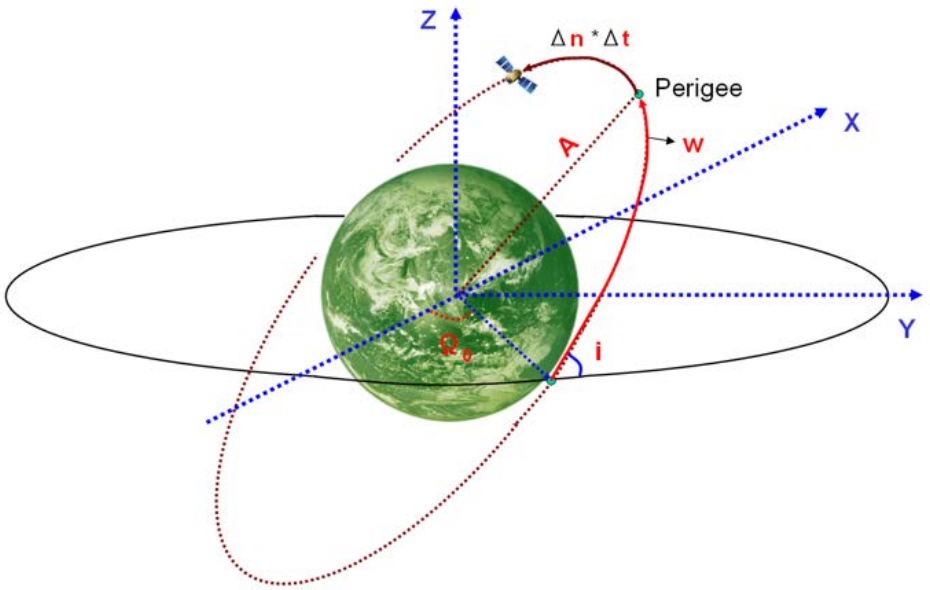


Figure 6-50 Keplerian Elements of a Satellite Orbit

The transformation back and forth between the Cartesian and the Keplerian forms is possible. However, when representing the satellite positions and velocities for different epochs in an orbital cycle through multiple sets of Keplerian elements, all the elements should be slow-varying quantities, except ν, which satisfies the needs of the new NAV messages.

Based on all the above fundamental analysis, I would propose the new GNSS NAV messages as described in Table 6-12, which can well comply with the above 'rules' and can therefore inherently support fast TTFF and high sensitivity.

Table 6-12 Proposed Contents of New GNSS Navigation Messages

No.	Items	Remarks
1	Satellite clock	Af0, af1, af2
2	Satellite ephemeris	6 Keplerian elements, 1 Solar radiation parameter
3	Ionosphere information	Klobuchar model
4	System Time Parameters	Relating this GNSS time to UTC Relating the time of this GNSS to other GNSSs
5	EOP data	Earth orientation parameters
6	Almanacs	Same form as GPS almanac
7	…	…

The proposed update interval for each part of the new NAV messages in Table 6-12 is 1 day, but for Almanac part, the update interval is possibly extended to a few days similar to GPS. In the ephemeris part, the proposal contains 6 basic Keplerian elements (*e, a, i,* Ω*,* ω*,* ν) and 1 solar radiation parameter (p_0) for selected reference time t_0 . Once the ephemeris is downloaded, the 6 Keplerian elements can be immediately transformed to Cartesian position $\mathbf{r}(t_0)$ and velocity $\dot{\mathbf{r}}(t_0)$, and then can be used as the initial condition to derive the entire orbit through Equation (6-3).

Comparing to current GPS ephemeris, Table 6-12 contains much fewer parameters, therefore it is possible to have the new GNSS ephemeris and clock packed in only 2 sub-frames, assuming the same data rate, same word structure and same sub-frame length. For the remaining parts listed in Table 6-12, they can be packed into multiple pages of 2 more sub-frames, in a similar way as the pages of sub-frames 4 and 5 in current GPS NAV messages. Therefore, the super-frame structure of the proposed new GNSS NAV messages will look like what is depicted in Figure 6-51. Considering that the contents of the first 2 sub-frames play a primary role in TTFF, the pages of sub-frames 3 and 4 are not further discussed in this paper.

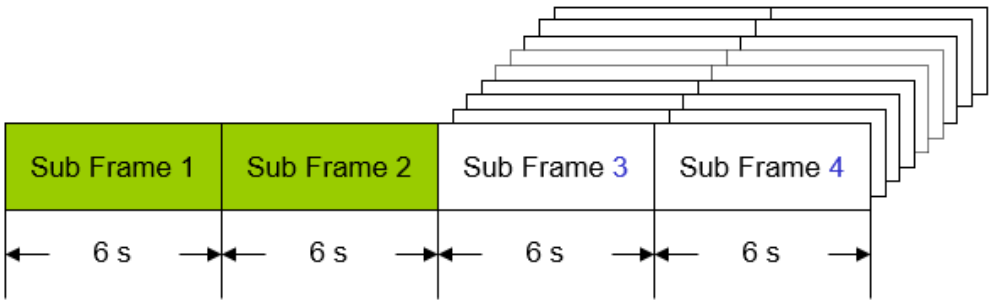


Figure 6-51 Super-frame Structure of New GNSS NAV messages

With the proposed new NAV messages, it will take only around 12 s to download the ephemeris, and it will just take 24 s to get next copy of the ephemeris even if missing the first sub-frame. If modulated on future GNSS signals, the messages can be repeatedly broadcasted in a whole day. As long as the ephemerides are downloaded from the satellites once in a day, the mobile GNSS devices can work properly without downloading any more in the rest time of the day. Comparing to current NAV messages used in GPS, GLONASS, BDS and GALILEO systems, the new NAV messages are able to offer an incomparable advantage, which can not only greatly reduce the

TTFF of the devices, but also greatly extend the ability of standalone devices to work under weak signal environments. This will be discussed in the next section.

6.2.2 Advantages of a GNSS with the Proposed New NAV Messages

The contents of the new NAV messages have been proposed in the last section but the detailed format design is beyond the scope of this paper. In Table 6-13, a comparison of the new NAV messages to GPS, GLONASS (GLO), BDS (BD) and GALILEO (GAL) is given. For the convenience of comparisons, the same data rate (50 bps) and the same length of sub-frame (6 s) to GPS NAV messages have been used for the new GNSS NAV messages.

Table 6-13 Comparison of the NAV messages for GPS/GLO/BDS/GAL/New NAV

	GPS	GLO	BDS (D1)	GAL (F/NAV)	New NAV
Data rate	50 bps	50 bps	50 bps	25 bps	50 bps
Length of super-frame	750 s	150 s	720s	600 s	600 s
Length of sub-frame	6 s	30 s	6 s	50 s	6 s
Length of Eph	18 s	10 s	18 s	~30 s	12 s
Repeat of Eph	30 s	30 s	30 s	50 s	24 s
Update of Eph	2 hrs	0.5 hr	1 hr	3 hrs	24 hrs
Life of Eph	4 hr	0.5 hr	2 hr	4 hrs	>24 hrs

Comparing to other GNSS NAV messages, the new NAV messages have a smaller size but the contained ephemeris has a longer life, and as a whole, the new NAV messages just need to be

updated once every 24 hours. To help understand the advantages of the new NAV messages, the following comparisons are made:

6.2.2.1 Standalone UEs, New GNSS vs. GPS

For any new GNSS that deploys the new NAV messages, the UEs just need to download the ephemeris from the satellites once in a whole day, whereas current GPS UEs need 12 times. In each downloading, it takes ~18 s for current GPS UEs, whereas only ~12 s for the new GNSS UEs. So there is no doubt that, from the TTFF perspective, the new NAV messages have incomparable advantages over the current GPS. Once a complete copy of the new NAV messages is downloaded, it can be used for data aiding in tracking loops in the rest time of the whole day, even without network connections in weak signal environments. However, for current standalone GPS UEs, they have to be in a strong signal environment to acquire fresh NAV messages every 2 hours. Otherwise there could be no position fix available in the next 2 hours due to the stale NAV bits and expired ephemerides. So from a sensitivity point of view, a GNSS with the new NAV messages (referred as New GNSS below) will also have incomparable advantages over the GPS.

6.2.2.2 Assisted UEs, New GNSS vs. GPS

There are three purposes to have the assisting information for mobile devices: a) to expedite signal acquisition; b) to save time in ephemeris downloading; and c) to have navigation bits for data aiding in tracking loop. For assisted GPS UEs and assisted GNSS UEs with the new NAV

messages, there is no much difference in the first aspect, as the assistance data, such as SV list, Doppler frequency, code phase, location and time, are common to both. For the second and third purposes the assistance data sent from the assisting network to the UEs are only needed once per day using the new NAV messages since they are updated only once per day. For assisted GPS UEs, the assistance data are needed once every 2 hours, which means that GPS UEs need frequent network connectivity and more network bandwidth for data transportation. In addition, as the size of a GPS super-frame is larger than the super-frame of the proposed new NAV messages, the time delay in transporting the assisting data will be larger in GPS assisting network.

6.2.2.3 Current GNSS, Standalone vs. Assisted

It is certain that the assisted GNSS UEs outperform the standalone GNSS UEs in terms of both TTFF and sensitivity. The performance difference in open sky primarily lies in TTFF, because AGNSS can always help save ephemeris downloading time. In weak signal environments, the TTFF difference becomes even larger as both signal acquisition and ephemeris downloading take much longer than usual time for standalone UEs. With assistance data, both the acquisition and tracking sensitivities can be significantly improved. So it is obvious that assisted GNSS UEs always have advantages over standalone UEs.

6.2.2.4 New GNSS, Standalone vs. Assisted

When the new GNSS NAV messages are deployed, as the messages are only needed to be downloaded once in a day, the assisted UEs mostly show advantage in the time and sensitivity for signal acquisition. Since signal acquisition is difficult only when the signal becomes weaker below a certain level, the performance of standalone and assisted new GNSS UEs is expected comparable under normal signal conditions. Under weak signal conditions, as long as the NAV messages are received once in a day, the performance in tracking sensitivities for both standalone and assisted UEs is also expected comparable. Therefore, with the new NAV messages, the performance difference between standalone and assisted UEs is expected to be much less comparing to current GNSS.

6.3 Concerns of Deploying the New NAV Messages

6.3.1 Accuracies of Satellite Orbit and Clock

Since the proposed update interval for the new NAV messages is 24 hours, a period much longer than GPS/GLO/BD/GAL, some immediate concerns may arise, such as:

- Is the orbit/clock derived from the ephemeris good enough for 24 hours?
- Is the calculation load for deriving satellite orbits affordable on a UE?

The advancement in orbital determination and EE technologies can help relieve the worry on the first concern. In the JPL predicted orbit and clock states (JPL, 2014), it is claimed that the user range error (URE) around 1 m for 1-day and URE less than 10 m for 7-day predictions can be obtained. Also, in Zhang et al. (2008), impressive results from driving tests with extended

ephemerides for up to 5 days are presented, which are equivalent to those using authentic broadcast ephemeris.

For a future GNSS that deploys the proposed new NAV messages, an orbital determination center (ODC) on the ground should be able to provide orbit predictions better than or at least comparable to Zhang (2008) and Han (2009). Every 24 hours, as the intermediate results of the orbit predictions in the ODC, the new ephemeris data can be extracted and packed as one part of the new NAV messages. Once uploaded to the satellites and broadcasted to the GNSS UEs on the ground, they can be used in deriving satellite orbits. The accuracies of the orbits/clock finally derived on GNSS UEs will be subject to the following factors:

- Accuracy of ephemeris
- Accuracy of clock coefficients
- Accuracy of EOPs
- Accuracy of force models embedded on UEs

The broadcast ephemeris in the new NAV messages directly inherits the accuracy of predicted orbit in the ODC. The EOP data, describing the irregularities of the earth's rotation, is needed for coordinate transformations between ECEF and ECI, so the up-to-date EOP data carried in the new NAV messages ensures no accuracy loss in such transformations. For the force models embedded on the GNSS UEs, accuracy is not a problem as long as they are same to what are used in the ODC.

As to the satellite clock, it is desired that, even if the clock coefficients are updated once per day, the accuracy of the predicted clock still suffices the need of navigation. For currently space-

borne clocks on GPS satellites, they are primarily Cesium atomic clocks with stability not better than 10^{-13}. But it should be noted that, the currently deployed atomic clocks are based on the most advanced technology years ago. Figure 6-52 gives the progress in atomic clock technology, as seen, The advancement of atomic clock technologies is fast especially in recent years, and the era of Rubidium, Caesium and Hydrogen maser clocks is evolving to Ytterbium and even Optical atomic clocks. As of today, atomic clock as stable as 10^{-18} has been claimed available in a laboratory settings (Meiser, 2014). And currently a project called Space Optical Clock (SOC) is ongoing (Schiller et al. 2014), aiming to put lattice optical clock with stability of 10^{-16} on International Space Station (ISS) by 2020. So it is foreseeable that new GNSS in the near future should be able to deploy atomic clocks with stability several orders better than currently deployed clocks. At the stability of 10^{-16}, the clock will only introduce millimeter level errors in ranging in a 24 hours period. With such stable satellite clock, there should be no accuracy concern with clock data updated once per day.

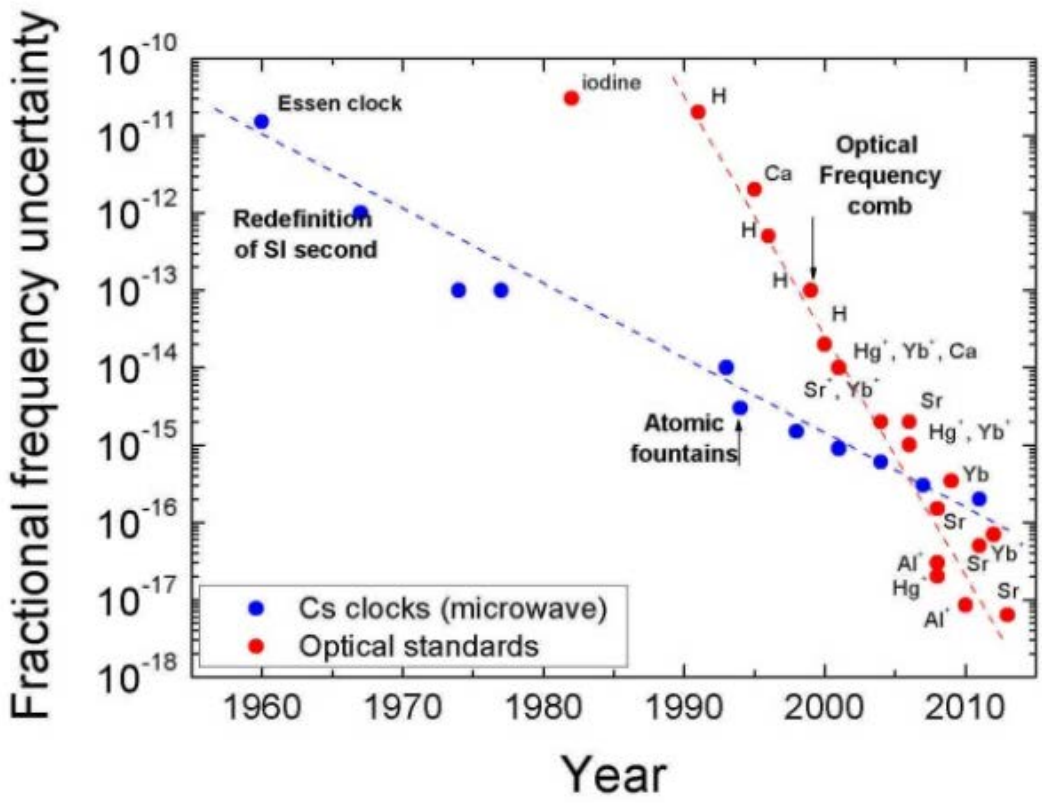


Figure 6-52 Progress of Atomic Clock Stability (Courtesy of Meiser, 2014)

6.3.2 Broadcast of the New NAV Messages

Once the broadcast ephemeris is received on a UE, numerical integration can be started to derive the satellite orbit. During the numerical integration, the calculation load is primary dependent on the following factors:

- Length of numerical integration
- Numerical integration Step size
- Order of the integrator
- Complexity of local force models

In the worst case, the UEs will need to have numerical integrations for 12 hours, as per the example given in Figure 6-53. The time of ephemeris update (TOU) is selected to be fixed on each day (for example 00:00:00), and the reference time of the ephemeris (TOE) is selected to be 12 hours ahead of TOU (for example 12:00:00). Between the neighboring two TOU, the same NAV messages are repeatedly broadcasted. If the UE is powered on right after a TOU or right before the next TOU, it needs to have a backward or forward numerical integration starting from the TOE for up to 12 hours.

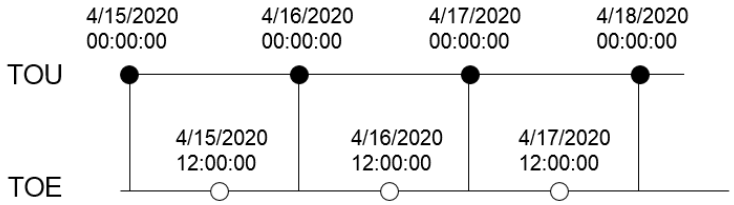


Figure 6-53 New NAV Messages, Time of Update (TOU) and Reference Time of Ephemeris (TOE)

Regarding the run-time necessary for orbital numerical integration on an embedded system, some public results, e.g. Zhang et al. (2008), indicate that a 3-day prediction (numerical integration) takes only around 0.6 s on a 600 MHz processor with floating point units (FPU). So for a 12-hour integration, it is supposed to takes only ~0.1 s on the same platform. As of 2014, for the popular high end smartphones in the market, the speed of embedded processors ranges from 1.2 to 2.5 GHz with dual or quad-core (Shanklin, 2014). Considering the drastically growing computation power of mobile processor and the potential of further algorithm optimizations in orbital integration, the calculation load of numerical integration for a 12-hour interval is not at all an issue on a mobile device today, not to say in the future.

The GPS system designers three decades ago might not have realized that, GPS would become so popular in the 21 century. Fast TTFF and high sensitivity have become standard requirements. The growing power of the application processors has also been beyond the imagination of people thirty years ago. So in their design, fast TTFF and high sensitivity might not have been given too much attention. The GPS modernization program has been an attempt to meet the growing expectation on the system performance in the applications for today and near future. In view of this, there is no reason not to give special considerations to inherently support fast TTFF and high sensitivity applications when investigating and designing a new GNSS. Certainly, such efforts can be found both in recently launched GPS (Block IIF) and GALILEO, such as the pilot channels, but the navigation under weak signal conditions for future standalone GPS and GALILEO devices is still susceptible to the frequent change of NAV messages (Table 6-13).

6.3.3 Satellite Orbital Maneuver and Clock Adjustment

Any GNSS may experience temporary malfunction, or periodical maintenance, like satellite orbit maneuver and clock adjustment, and may cause the satellite totally unusable. In case of orbital maneuver, significant change could happen in the satellite orbital elements in short period of time through external thrust. Therefore, the satellite position and velocity, comparing to the calculation from the ephemeris broadcasted prior to the maneuver operations, could be much different. In the following Figures, an example of GPS satellite orbital maneuver is given on PRN 14 on 03/29/2016:

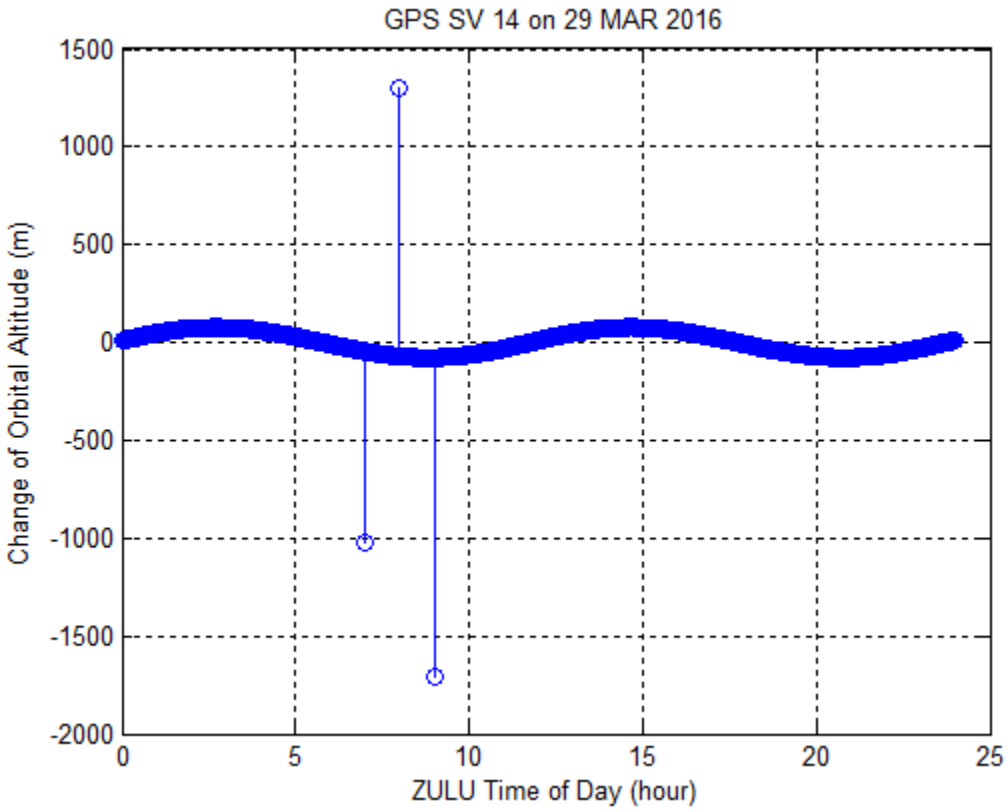


Figure 6-54 Example of GPS Orbital Maneuver - Impact on Orbit Altitude

As illustrated, during the orbital maneuver, changes of kilometer level happened on the satellite altitude in less than 40 minutes, comparing to the altitude assumed from original ephemeris.

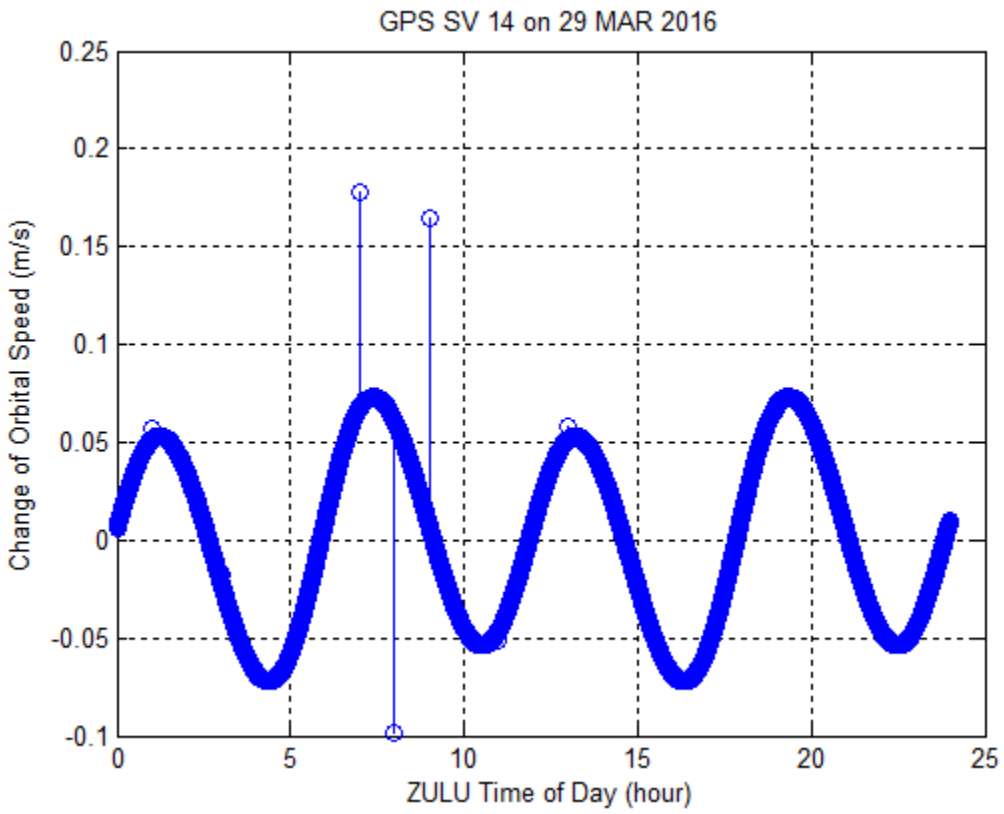


Figure 6-55 Example of GPS Orbital Maneuver - Impact on Orbit Speed

The GNSS malfunction or maintenance, once happened or planned, must be notified to the end users as soon as possible to avoid introducing un-predictable consequence to the end users. Taking GPS as an example, the change in the GPS constellation is notified to users by the NANU, including any planned maintenance (72 hours in advance) and unscheduled outages, through messages posted on the official NANU page (NANU, 2017). Similarly in other GNSSs, any maintenance event or anomaly about the constellation are notified to the end users by NAGU-1 (2016) for GLONASS, NAGU-2 (2016) for GALILEO, and unclear yet for BDS on this matter.

With the deployment of the proposed new NAV messages, definitely it is necessary to have similar NAXU where 'X' stands for the new GNSS that incorporates such new New NAV messages.

6.4 Equivalence of Cartesian and Keplerian Expression in Satellite Orbit

Each pair of satellite Cartesian position, $\mathbf{r}_s$, and Cartesian velocity, $\dot{\mathbf{r}}_s$, in ECI frame, corresponds to a set of Keplerian elements, as expressed in following Equation:

$$\begin{matrix} \left(\mathbf{r}_s \quad \dot{\mathbf{r}}_s\right) = \left(X \quad Y \quad Z \quad \dot{X} \quad \dot{Y} \quad \dot{Z}\right) \\ \Updownarrow \\ \left(e \quad a \quad i \quad \Omega \quad \omega \quad \nu\right) \end{matrix} \tag{6-4}$$

A satellite orbit can be precisely represented through a set of Cartesian IC, $\left(\mathbf{r}_s \quad \dot{\mathbf{r}}_s\right)$, however, it lacks the direct depiction of the orbital characteristics. Instead, if expressed in a Keplerian form, the IC is able to give a rough idea of how the orbit looks like.

6.4.1 Conversion between Cartesian and Keplerian

6.4.1.1 Conversion from Keplerian to Cartesian

The orbit segments depicted in the ephemeris for GPS, BDS and GAL are given in Keplerian form. Calculation of the Cartesian coordinates of the satellite at a specific moment, is only listed here for completeness, considering that the details have been provided in the ICD of each system.

6.4.1.2 Conversion from Cartesian to Keplerian

The conversion from a set of Cartesian orbital elements to a set of Keplerian orbital elements is not introduced in this section but in the Appendix step by step from Equations (A-1) to (A-7), considering the standard process is available from text books.

6.4.2 Convenience of sanity checking using Keplerian

The IC of an orbit arc expressed in Keplerian form not only provides a physical depiction of the orbit, but also enables a convenient sanity check on the valid range of each element. In Figure 6-57 and Figure 6-56, examples of the variations in the Keplerian elements from a GPS satellite orbit are given in a whole year time period, as an attempt to illustrate the slow-varying trend of the Keplerian elements, and also to provide an evidence of feasibility in sanity check considering that the range of variations for each element is well predictable.

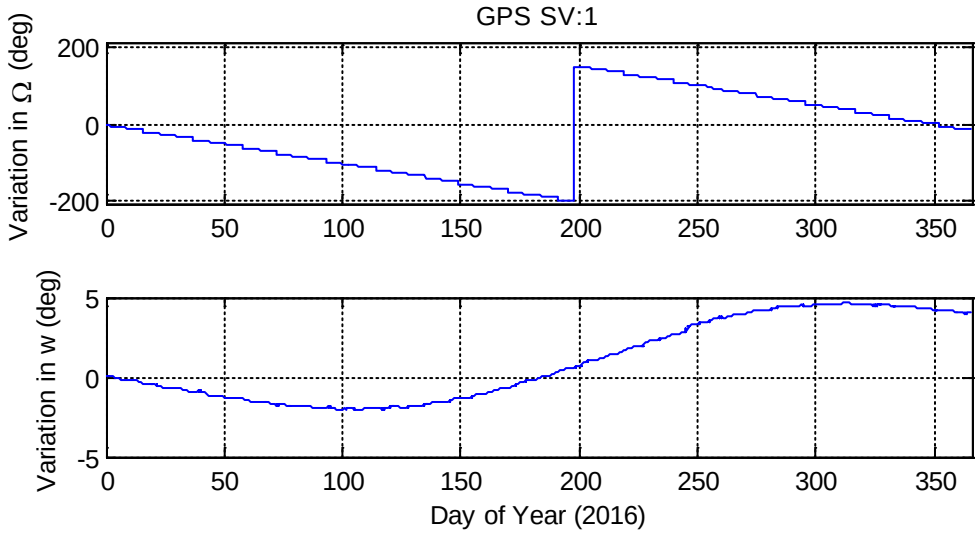


Figure 6-56 Example of Variations in Keplerian Elements (Ω, w)

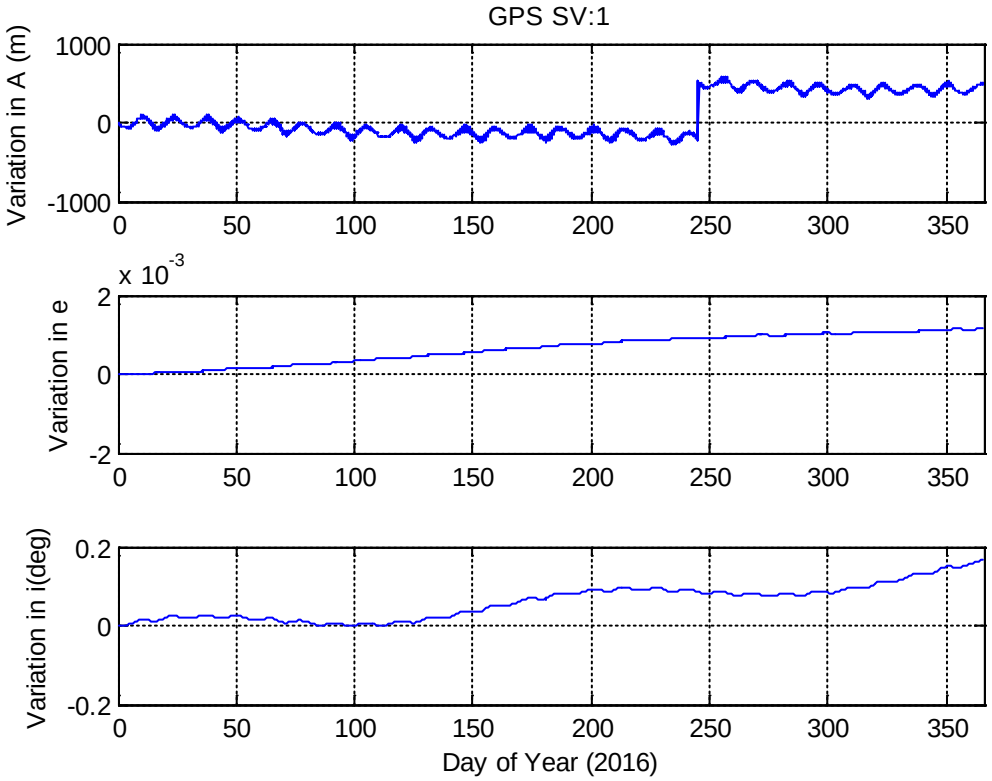


Figure 6-57 Example of Variations in Keplerian Elements (*a, e, i*)

6.5 New Method for User Range Accuracy (URA)

The broadcast ephemeris from different GNSS always comes with a User Range Accuracy (URA), as an indication of the quality of the ephemeris (GALILEO ICD, 2008; BeiDou ICD, 2013; GLONASS ICD, 2008; GPS IS, 2004). Per the interface control working group (ICWG) for GPS, URA is defined as a statistical indicator of the GPS ranging accuracy obtainable with a specific signal and satellite vehicle (SV), and provides a one-sigma estimate of the user range error (URE) in the associated navigation data for the transmitting satellite (ICWG, 2011). It should be noted that, URA does not account for user range error contributions due to the inaccuracy of the broadcast ionospheric data parameters used in the single-frequency ionospheric model or for other atmospheric effects.

To enable GNSS receivers to work in some extremely challenging environments with serious signal blockage or attenuation, continuously valid ephemeris is a necessity, which however is usually very difficult to decode from the GNSS signals in such environments. Assisted GNSS (AGNSS) provides alternative source of broadcast ephemeris, but requires network connectivity every a short period of time (~ 2 hours for AGPS) to get fresh ephemeris. Ephemeris Extension (EE) has been increasingly used nowadays as one of the primarily approaches to make the GNSS receiver work under the weak signal conditions by providing alternative ephemeris with the validity period extended to days, during which there is no more need to decode ephemeris from the signals or from AGNSS network. However, the accuracy and reliability of the extended ephemeris are of big concern, especially when the extended ephemerides aged at a few days are used in positioning, considering that any errors in ephemeris will be eventually translated into

positioning errors. In the EE products that are available in the Market, usually the User Range Errors (URE) are also provided, as a means to indicate the errors of the extended ephemerides along the user-to-satellite, namely line-of-sight (LOS) directions.

For end users, the potential applications of URA would include the weighting of measurements (pseudorange), and Receiver Autonomous Integrity Monitoring (RAIM) (ICWG, 2011), the long-term prediction of GNSS accuracy (Driver, 2007), the calculation of integrity failure probability for Signal In Space (SIS) as a function of the broadcast URA (Kovach et al., 2003), and setting the search window in signal acquisition for predicted code phase and frequency Doppler. In the case of extended ephemeris, although the provided URE could be of different form or at different time interval, when extended ephemeris is used, the corresponding URE is critical in the weighting of measurements, especially when the age of the extended ephemeris is over a few days.

6.5.1 Bounds of LOS errors from orbital error

The satellite positions calculated from the broadcast ephemeris or extended ephemeris are actually predicted positions, containing errors increasing with the age of the prediction. As illustrated in Figure 6-58, when a satellite at position ‘P’ is used in the positioning by users on the Earth surface, for the same amount error in the predicted position ‘P’ with respect to the truth satellite position ‘T’, the impact is different to users at different locations because of the different LOS directions from the satellite to the users. In other words, the projection of the same orbital errors on LOS direction varies with the change of user locations.

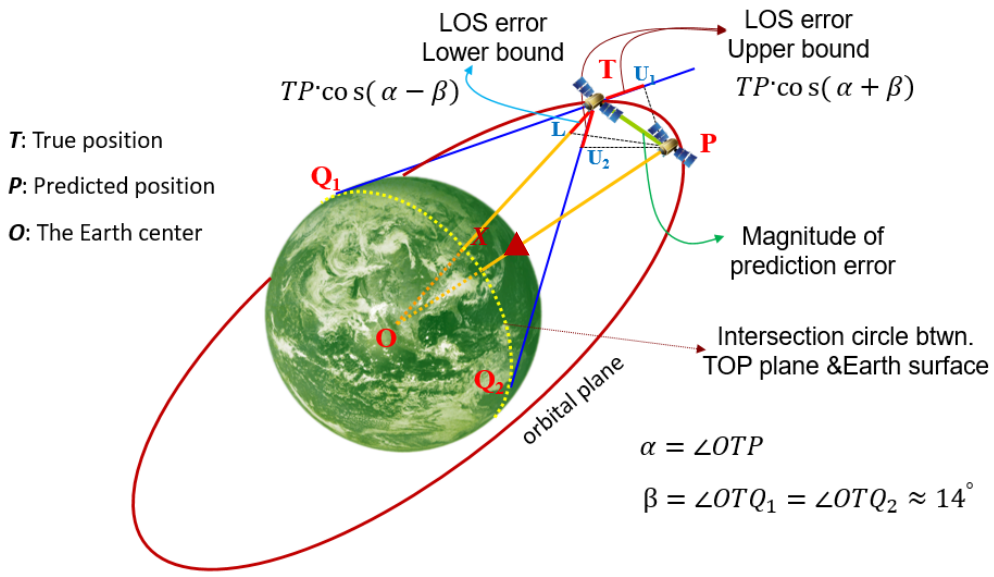


Figure 6-58 Lower and Upper Bounds of LOS Error due to Orbital Error

In Figure 6-58, 'TP' represents the magnitude of the orbit error. Through simple geometrical analysis, it is easy to understand that, in the case that the user is located at 'X' right under the satellite, namely the LOS direction XT is aligned with OT, 'TP' results in the minimal user range error when it is projected to LOS direction XT, namely:

$$URE_{\min} = |TP| \cdot \cos\alpha \tag{6-5}$$

where $\alpha = \angle OTP$. In the case that the user is located at 'Q_1' or 'Q_2' on the big circle that is determined by the intersection of the Earth surface and the plane 'O-T-P', where the LOS TQ_1 and TQ_2 are tangent to the big circle, 'TP' results in the maximum URE when it is projected to LOS direction TQ_1 or TQ_2., namely:

$$URE_{\max} = |TP| \cdot \cos(\alpha - \beta) \tag{6-6}$$

or

$$URE_{\max} = |TP| \cdot \cos(\alpha + \beta) \tag{6-7}$$

where β is the angle from nadir, at approximately 14 degree for the GPS constellation.

6.5.2 URE Algorithm and the Problems

Driver (2007) introduced the method used in GPS Operation Center that generates global URE for each GPS SV from integrating the user range error over the entire Earth surface. In particular, the equation contains 1-sigma errors for the radial, along-track and cross-track components, as well as the global user range error and clock error over the last 7 days.

Similarly for the URE provided in the EE products, they are usually also obtained through the statistics of large quantity of the satellite orbit and clock data in a past period of time, which is typically a few months, and on each day during the period, the ephemeris for each GNSS satellites is extended forward for days. Each extended ephemeris is then evaluated by comparing against selected reference ephemeris (broadcast ephemeris or precise ephemeris) at a constant time interval (usually 15 minutes). The evaluated errors on each interval are then projected to different LOS directions that are determined by the satellite position at that moment and multiple user locations that are selected on the ground. The LOS errors from different satellites, at different ground locations, in different extension trials and at different ages (length of extension) are then used in final statistical calculation for URE. The concept is illustrated in Figure 6-59.

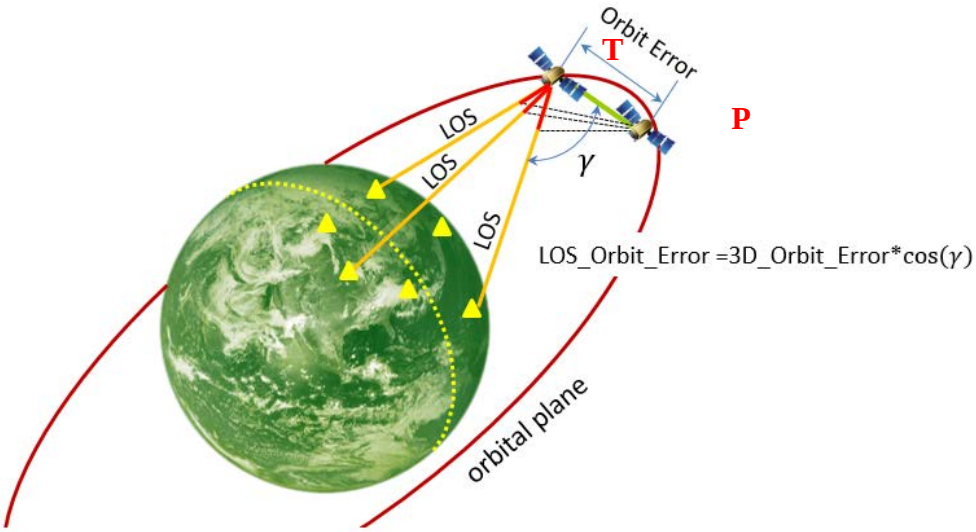


Figure 6-59 Statistics for User Range Errors (URE)

Given the orbit error as represented by 'TP', multiple typical locations on the Earth surface (yellow triangles) that are under the signal coverage of that satellite are selected, so that the orbit error can be projected onto different LOS directions to represent the typical user range errors incurred by the orbit error. It should be noted that, the satellite clock error is directly translated to a range error that is equivalent in all LOS directions. So as illustrated in the following equation, the overall URE should include both the LOS impact of orbital errors (in unit of meters) and the satellite clock error (in unit of meters), but for the URE discussed in this paper, it primarily refers to the part incurred from orbital errors if not otherwise explicitly explained.

$$\begin{aligned} URE &= URE_O + URE_C \\ &= LOS_Orbit_Error(m) + SV_Clock_Error(m) \end{aligned} \tag{6-8}$$

where URE_0 is URE due to orbital error and URE_0 due to clock error. Considering the fact that the URE calculation in both broadcast and extended ephemeris uses the statistics of orbital errors

in the past to quantify the errors in the predicted orbits, a question naturally arise here – how the errors in the past can reflect the errors in the future, especially when the orbital prediction is beyond days in ephemeris extension?

Actually, it has been observed that, the accuracy ephemeris extension differs significantly from satellite to satellite and is subject to seasonal variations, which means URE derived from statistical method sometime is not able to well reflect the actual errors in the extended ephemeris, especially when the extension is beyond a few days. In the broadcast ephemeris, as the orbital prediction is limited within hours, the issue is not significant. Therefore, the motivation of this paper is to seek a new method that can more efficiently quantify the errors in the extended ephemeris and eventually can serve the generation of URE.

6.5.3 Deterministic Calculation on URA based on Symmetry of Orbit Errors

6.5.3.1 Orbital Prediction

Before further explaining the new method, it is necessary to briefly go over the principle of orbital prediction. Orbital prediction is conducted right after the completion of orbital determination, using one of the critical outcomes from the orbital determination - the initial condition (Equation 6-1) estimated at some selected reference time (say t_0). As long as the acceleration $\ddot{\mathbf{r}}(t)$ at any time t can be computed from different force models including the two-body gravity model and additional perturbation models as illustrated in Equation (6-2). The orbit

at arbitrary time t can then be predicted through numerical integration as illustrated in Equation (6-3).

So the accuracy of the predicted orbit is primarily determined by the following factors:

- inaccuracy in the estimated initial condition
- imperfectness of the force models
- imperfectness of the numerical integrators

6.5.3.2 Symmetry of the Orbital Errors from Forward and Backward Integration

Slightly change Equation (6-3) into Equation (6-9):

$$\mathbf{r}(t_0 + \Delta t) = \mathbf{r}(t_0) + \int_{t0}^{t_0+\Delta t} \left[\dot{\mathbf{r}}(t_0) + \int_{t0}^{t_0+\Delta t} \ddot{\mathbf{r}}(t)dt \right] dt \tag{6-9}$$

where Δt represents the length of integration and $+\Delta t$ indicates that the orbital prediction takes numerical integration in forward direction. If taking the numerical integration in backward direction from the same initial conditions at the reference time, t_0, as illustrated in Equation (6-10):

$$\mathbf{r}(t_0 - \Delta t) = \mathbf{r}(t_0) + \int_{t0}^{t_0-\Delta t} \left[\dot{\mathbf{r}}(t_0) + \int_{t0}^{t_0-\Delta t} \ddot{\mathbf{r}}(t)dt \right] dt \tag{6-10}$$

then a **hypothesis** arises - the errors in the backward orbit should be highly correlated with the forward orbit, considering that:

- for both forward and backward integration starting from the same initial condition, the impact of inaccuracy in the initial condition should be the same;
- for both forward and backward integration, the impact of the imperfectness of the force models should be similar if the imperfectness is approximately evenly distributed at the same altitude;
- for both forward and backward integration using the same step size and the same integrators, the impact of the imperfectness of the integrators should be very close if the length of integration is the same.

The theoretical basis of the ephemeris extension is orbital integration from a given initial point to user specified time. As a further effort to explain the hypothesis, Figure 6-60 shows the time axis from 'Past' through 'Current' to 'Future'. At 'Current' time ***C***, given an initial condition consisting of position, velocity, solar radiation parameters and etc., it is possible to derive the orbit in either the 'Future' or the 'Past' through forward or backward orbital integration.

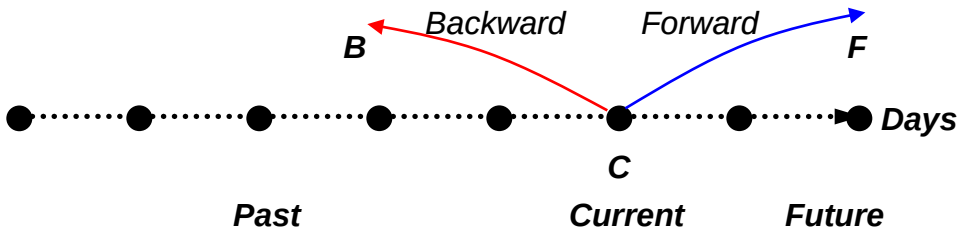


Figure 6-60 Concept of 'Approximate Accuracy Symmetry'

It is fair to assume that the truth orbits from the 'Past' to 'Current' are available, so the errors in the orbit derived from the backward integration can be exactly evaluated. As the truth orbits for the forward integration from 'Current' to 'Future' is not available yet, based on the hypothesis,

the evaluated errors in the backward orbit, should be able to help quantify the errors in the orbit derived from the forward integration.

6.5.3.3 URE based on the Error Symmetry

As one of the applications of the aforementioned hypothesis, the URE of the predicted satellite orbits can be generated utilizing the error symmetry characteristics between the forward and backward orbits, especially when the orbital prediction is beyond days.

Using this method, satellite orbits are firstly extended forward from the reference time for a few days, then based on the same initial conditions at the reference time, orbits are derived backward for the same length of time. The errors of the backward orbits can be evaluated with previously collected reference ephemeris, and are then mirrored to the forward time period and used to quantify the errors in the extended orbits.

Once the errors of the predicted orbits are reliably quantified, they can be translated to URE using similar methods that are introduced in the section of URE algorithms. The key point here is not how to translate orbital errors into URE, but how to quantify the errors of the predicted orbits. As long as the quantification of the errors is improved, the resulting URE should be improved.

6.5.4 Advantage of the New URA Method

Some or alternatives or prior state-of-the-art for indicating the accuracy of orbital predictions is based on statistics of orbital predictions in the past a few months, and assume that the accuracies of future orbital predictions for all SVs follow the statistical pattern. Unfortunately, study shows that the orbital prediction accuracy varies largely from time to time, from satellite to satellite. Therefore the statistics is hard to be a close approximation of the actual accuracy for specified satellite at specified time. It often happens that the statistics says good but actual accuracy is pretty poor, and the statistics says bad but actual accuracy is pretty good.

The advantages of this invention over the above alternatives lie in the following two aspects:

(1) This invention treats each satellite individually, and thus gives the insight to the differences among different satellites;
(2) This invention derives backward orbit upon each forward orbit prediction, and thus takes into account the fact that the orbital prediction performance of each satellite varies with time;

In short, the accuracy indicators provided by this invention for the extended ephemeris are time-dependent and satellite-dependent, which are more realistic and reliable than the above alternatives.

6.5.5 Numerical Example

This section is to present some results from experiments on orbital prediction for GPS satellites using the new method. The numerical results are used to validate the above hypothesis. To be

able to compare the orbital errors from forward and backward orbits, in the experiments, the 'Current' time is selected in the past so that the truth orbits are available during the time periods of both orbits that are derived from forward and backward integrations.

Without performing complicated orbital determination, for each GPS SV, the initial conditions are obtained by extracting the GPS satellite position and velocity from some broadcast ephemerides at the time of effectiveness (Toe), and some empirical value is selected for solar radiation parameter in terms of Cannon Ball model (Montenbruck et al., 2000). For each SV, both the forward and backward orbits are integrated for 5 days (Δt = 5 days) starting from the same initial condition, with the broadcast ephemerides selected to generate reference orbits covering the overlapped time period to evaluate both forward and backward orbits.

As illustrated in Figure 6-61, the ECEF errors (both the three components and error magnitude) for both forward and backward orbits are plotted for a GPS SV.

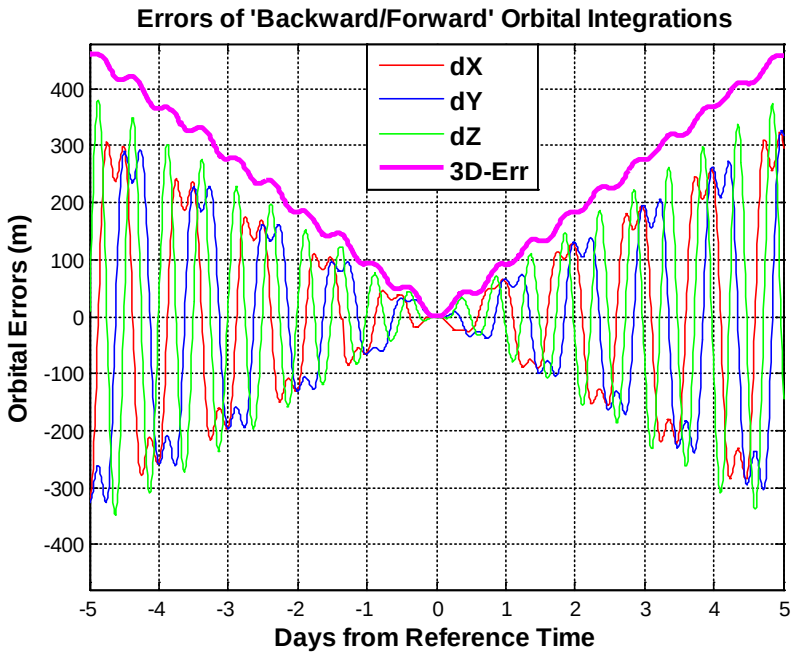


Figure 6-61 ECEF position errors for forward and backward orbits

In Figure 6-61, the red, blue and green lines represent the orbital prediction errors in ECEF X, Y, Z directions respectively, and the magenta represents the magnitude of orbital errors. The time axis indicates the days of orbital prediction from 'Current'. In the time axis, '0' represents the 'Current', the negative direction represents the 'Past' or 'Backward', and positive direction represents the 'Future' or 'Forward'. As seen, the errors in both orbits grow with time from the reference time and a good symmetry appears between the errors with respect to the reference time.

To help better understand the symmetry between the errors in the backward and forward orbits, the time axis of the errors in the backward orbit is flipped, so that both the backward and forward errors are plotted on the same side as shown in Figure 6-62, where the time axis indicates the

length of orbital integration in days, with the two directions marked in red and blue respectively. Also in Figure 6-62, the differences between the backward and forward errors are plotted in green. Taking the 'Backward' and 'Forward' errors as two time sequences, the correlation coefficient between them is calculated, which is used as a metrics to assess the level of the symmetry between the backward and forward errors. As displayed, the backward and forward errors are highly overlapped and the correlation is as strong as 99.96%, indicating a very strong symmetry between the errors in the backward and forward orbits with respect to the reference time. In addition, the error differences in Figure 6-62 indicate that it is possible to quantify the errors in the orbital prediction (forward orbit) in up to 5 days as accurate as ~12.5 m by studying the errors in the backward orbit. It should be noted that, as the reference orbits are calculated from broadcast ephemerides, the accuracy of the broadcast ephemerides is a factor that would affect the calculated level of symmetry in the forward and backward errors.

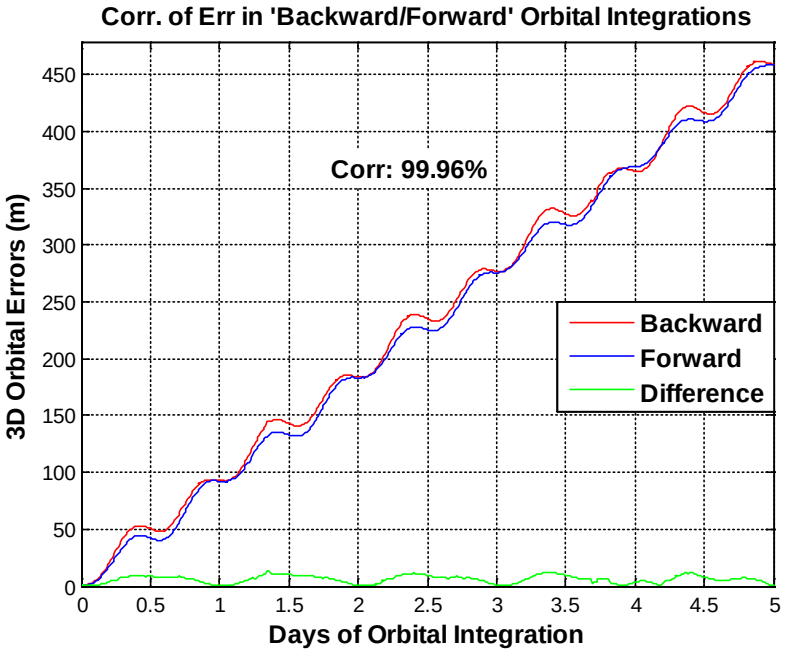


Figure 6-62 Comparison of error magnitudes for forward and backward orbits

Similarly, the velocity errors in both forward and backward orbits are also studied for the same GPS SV in Figure 6-63 and Figure 6-64. As seen, a good symmetry also appears between the velocity errors in the forward and back orbits, with correlation as high as 99.62%. The velocity difference in green also indicates that, by studying the velocity errors in the backward orbit, the velocity errors in the forward orbit can be quantified as accurate as a few mm/s up to 5 days.

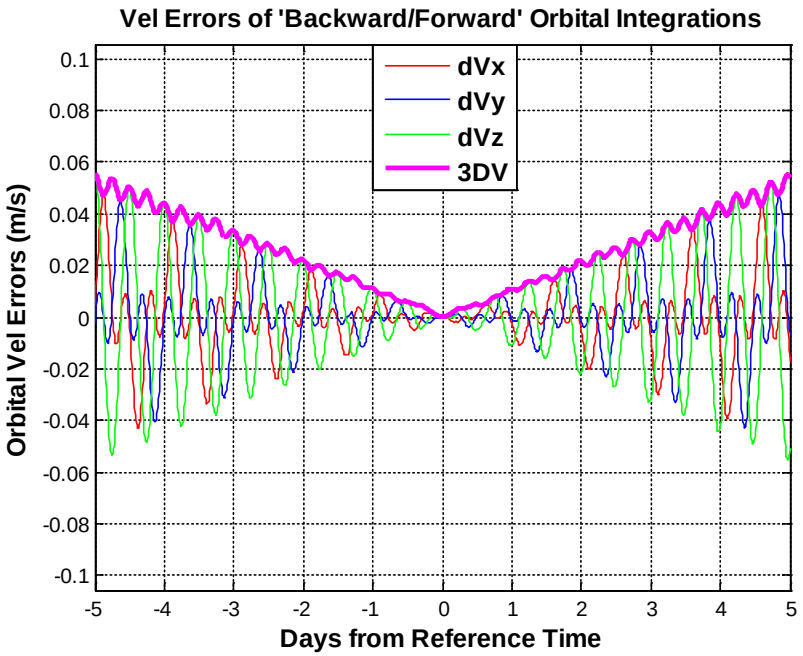


Figure 6-63 ECEF velocity errors for forward and backward orbits

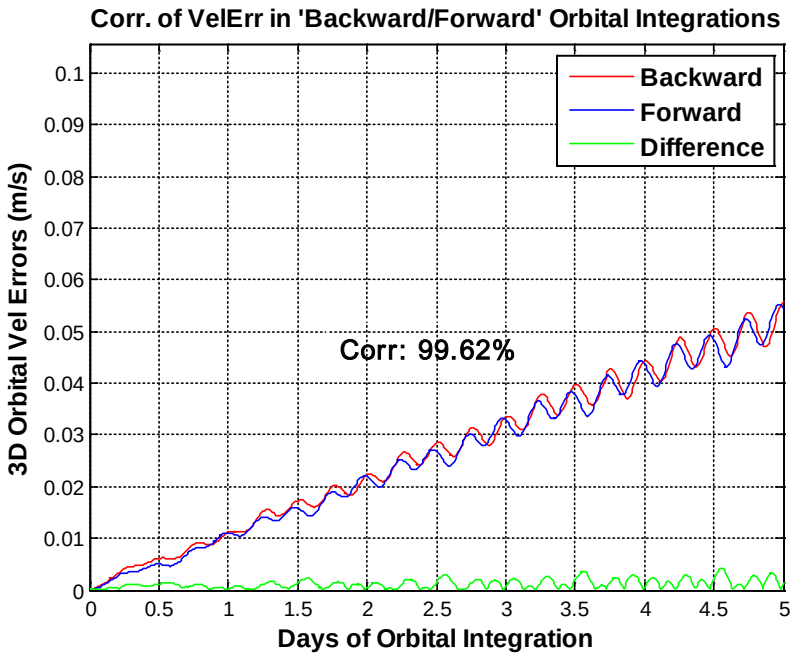


Figure 6-64 Comparison of velocity error magnitudes for forward and backward orbits

Further studying the rest of the GPS satellites during the same time period, very high symmetry exists between the errors in the forward and backward orbits, except one SV (PRN 29) with a correlation of forward and backward errors 85.24%. As a summary of the level of symmetry for all the GPS satellite studied during the same time period, the basic statistics of the correlation coefficients is listed in Table 6-14.

Table 6-14 Statistics of the Correlation for all Studied GPS SVs

	Correlation
Average	98.88%
Best	99.96%
Worst	85.24%
Std	3.02%

In Table 6-15, the basic statistics of the correlation coefficients is calculated again after the worst SV (PRN 29) is excluded. As seen, consistently high symmetry exists in the forward and backward errors for all remaining SVs.

Table 6-15 Statistics of the Correlation with the worst SV excluded

	Correlation
Average	99.50%
Best	99.96%
Worst	98.17 %
Std	0.51%

The above results, to certain extent, constitute the validation of the hypothesis about the high correlation between the errors in the forward and backward orbits.

CHAPTER 7 SIMULATION AND ANALYSIS ON GNSS THAT DEPLOYS THE NEW NAV MESSAGES

This chapter includes the discussions on the deployment of the proposed new NAV messages, and study on some use cases for a GNSS with such new NAV messages in terms of different performance metrics.

7.1 Deployment of the New NAV Messages in a GNSS

The existing GPS ephemeris contains a set of Keplerian parameters (20+ elements) to describe a fitted orbit arc of 2-4 hours around a reference time $\mathbf{t_{oe}}$, and the position, velocity of any point of the orbit arc can be analytically calculated given a time $\mathbf{t_k}$ within the validity time period. For the convenience of comparison, such ephemeris may be referred to as traditional ephemeris (or broadcast ephemeris) and the corresponding NAV messages may be referred to as traditional NAV messages.

The proposed new NAV messages contain a set of parameters to describe the initial condition (IC) at a reference time $\mathbf{t_{oe}}$, so that the orbit arc of a length of 24 hours can be precisely derived through numerical integration from the $\mathbf{t_{oe}}$ with the IC. Further interpolation may be needed to get the exact position, velocity of points in between an integration step. In the following, our discussion on message deployment will focus on how the new method delivers SV ephemeris for the purposes of fast TTFF and high sensitivity. As to any specific format to be used for broadcasting the new NAV messages, this will depend on the communication requirements of a navigation system for operational implementation. A detailed design of the proposed super frame

structure that consists of only 4 sub-frames for the new NAV messages, however, is not the focus of this thesis.

7.1.1 System Architecture of a GNSS Deploying the New NAV Messages

For a GNSS that deploys the proposed new NAV messages, the system architecture is conceptionally illustrated in Figure 7-65, which involves the deployment of three system components, namely the Control Segment (CS), the Space Segment (SS) and the UEs.

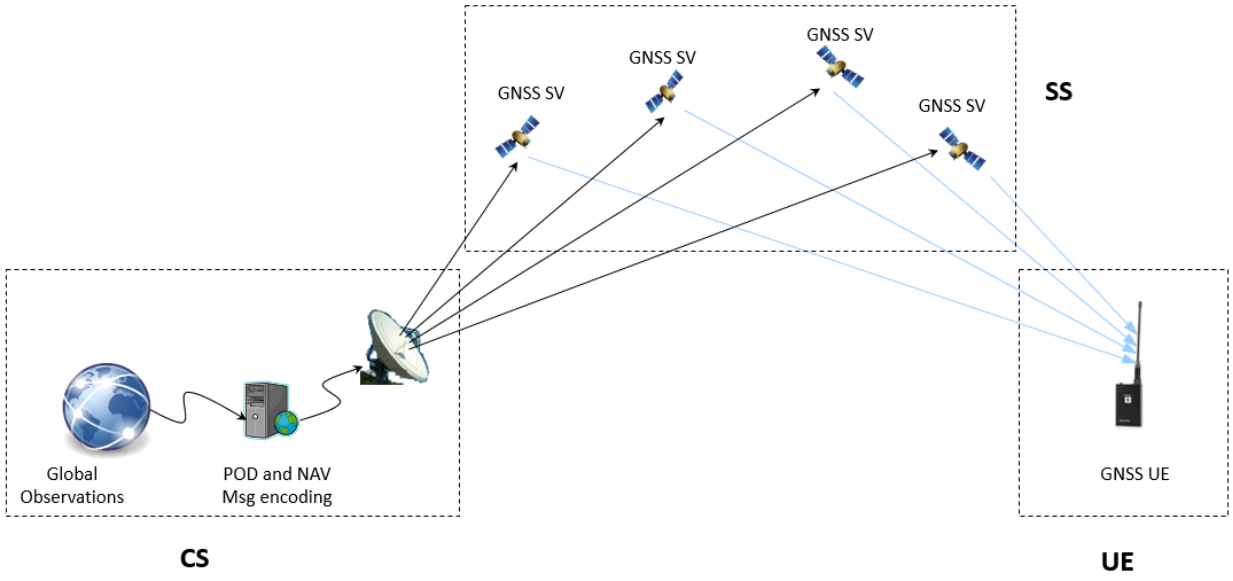


Figure 7-65 System Architecture of a GNSS Deploying the New NAV Messages

The primary work related to the new NAV messages in each system component would include:

- Control Segment (CS)
    - Precise Orbital Determination (POD) for each SV with observations collected from global network

- Select the initial condition (part of the outputs from POD) for the next reference time ($\mathbf{t_{oe}}$) for each SV
- Encode the initial condition, along with other slow-varying items into the new NAV messages
- Fit the clock parameters for each SV
- Encode the SV clock parameters into the new NAV messages
- Upload encoded new NAV messages to SS

▪ Space Segment (SS)

- Receive the new NAV messages from CS
- Broadcast the new NAV messages to UEs

▪ User Equipment (UE)

- Receive the new NAV messages from SS
- Decode the initial conditions from the new NAV messages
- Numerically integrate the orbital arc for up to 24 hours
- Interpolate position and velocity for desired time
- Decode the SV clock from the new NAV messages
- Compute the SV clock bias and drift for desired time

7.1.2 Deploying the New NAV Messages on CS

The work relevant to deploying the new NAV messages on CS is further depicted in Figure 7-66. One of the most important tasks on CS is to precisely determine the orbit for each SV based on the globally collected observations (namely POD); and one of the major objectives of POD is to

get the precise ICs at preselected reference time, which is a core part of the new NAV messages. In the work for current GPS CS, once the ICs are available, future orbits are further extended and then fitted into sets of Keplerian elements every a certain length of arc (say 4 hours), so that they can be used to constitute the ephemeris for each SV that is valid for up to 4 hours. Therefore, deploying the new NAV messages on CS can easily fit into the existing software architecture with minor changes. In the context of the new NAV messages, once the ICs of each SV is available, the IC elements can be directly converted into the form of Keplerian elements (CCAR, 2009) to constitute the new ephemeris, and the future orbits are extended only for the purpose of URA evaluation based on the method that has been introduced in Chapter 6.

It should be noted that, for traditional GPS ephemeris, the future orbits have to be extended from the precisely determined IC and orbit arcs have to be fitted every 4 hours into Keplerian elements to constitute traditional ephemeris; however, for the new NAV messages, the IC is directly converted into a set of Keplerian elements to constitute new ephemeris.

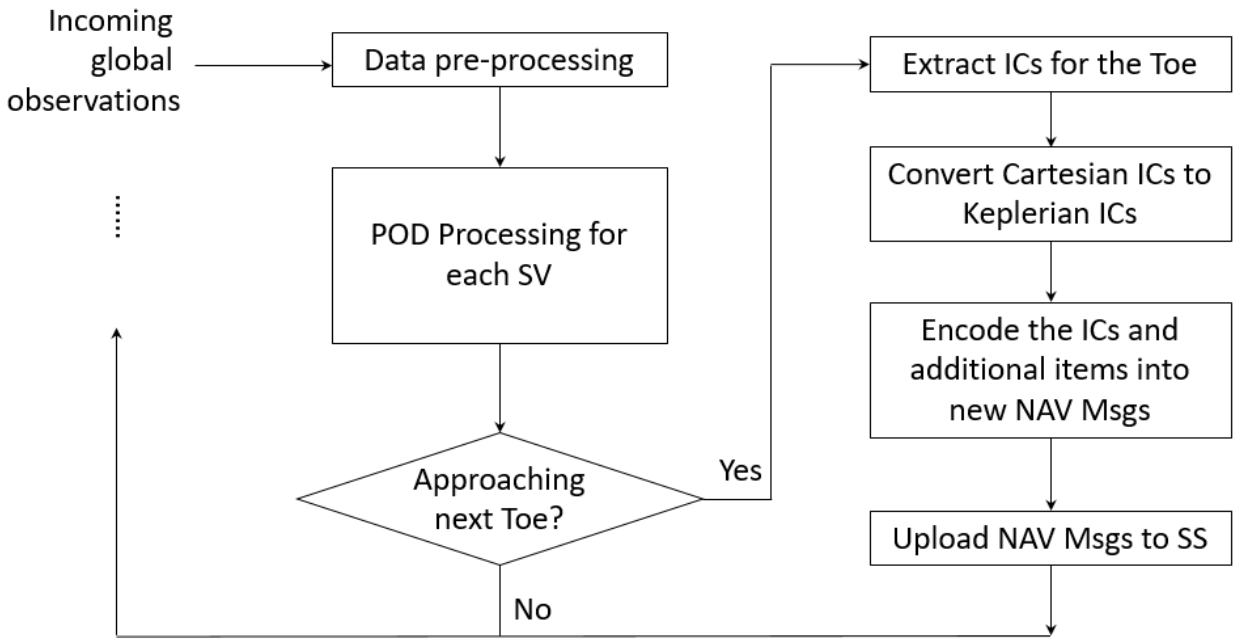


Figure 7-66 Deployment of the New NAV Message on CS

7.1.3 Deploying the New NAV Messages on SS

For the new NAV messages deployment, there will be little or even no need for any change on SS, as illustrated in Figure 7-67. The NAV messages uploaded to SS are cached and then broadcast to the UEs on the Earth ground when the ephemeris' reference time is switched to the exact $\mathbf{t_{oe}}$. Considering that the new NAV messages are designed to extend the new ephemeris' life expectancy to 24 hours, as mentioned in Chapter 6, once a copy of the new NAV messages is uploaded to SS, it can be repeatedly broadcast to ground UEs for 24 hours. Therefore, the uplink burden of the GNSS using the new NAV messages is expected much less than the GNSS using traditional NAV messages like GPS. And from SS perspective, the less upload means less cache memory and higher reliability.

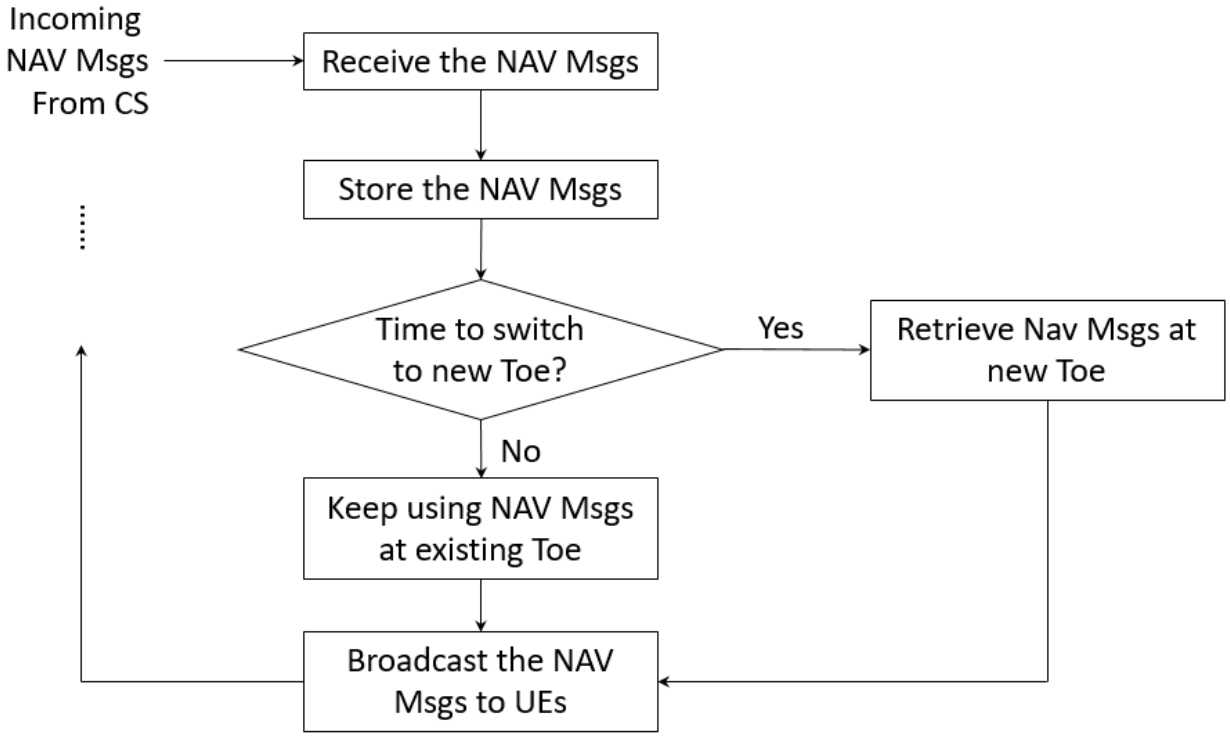


Figure 7-67 Deployment of the New NAV Messages on SS

7.1.4 Deploying the New NAV Messages on UE

When the new NAV messages are deployed, the major difference probably lies in the UEs. As depicted in Figure 7-68, after a copy of the new NAV messages is successfully decoded, it can be continuously used for up to 24 hours. During this time period, such new NAV messages would make the UE startups much easier, and the signal tracking much convenient, as there is no need to worry about the expiration of the ephemeris.

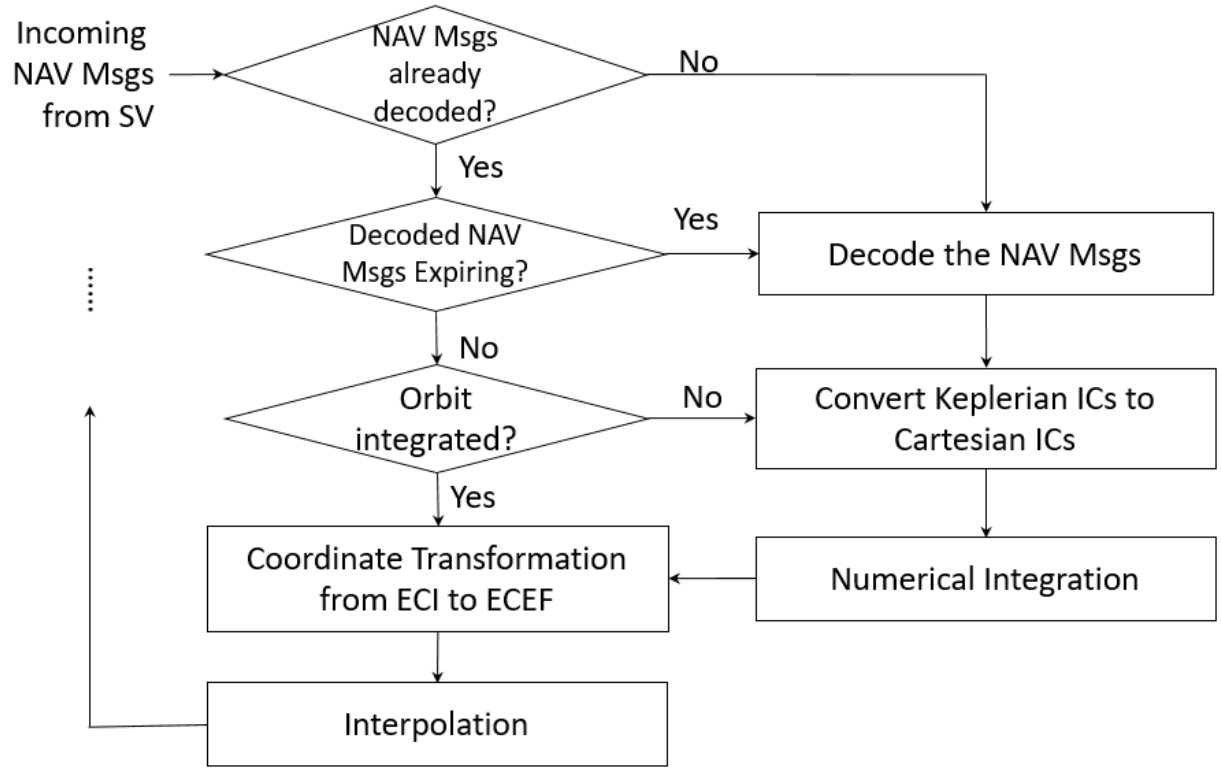


Figure 7-68 Deployment of the New Messages on UE

As a cost to such benefits, the orbit calculation on UE using the new NAV messages will be more complicated than using the traditional GPS ephemeris. In the context of the new NAV messages, the Keplerian elements in the ephemeris of the new NAV messages no longer represent a certain length of orbit arc, but only corresponds to a point of the IC. So as soon as a copy of new ephemeris is decoded, the Keplerian elements should be immediately converted into Cartesian elements in order to get an IC at the $\mathbf{t_{oe}}$. Then the satellite orbit has to be derived through numerical integration from the IC, and satellite position and velocity at any time instants in between the integration step size should be interpolated using the neighboring points that are directly obtained from the numerical integration.

7.2 Simulation and Validation of the Deployment of the New NAV Messages

A high level discussions on the deployment of the proposed new NAV messages for the sub-systems (including Ground Control Centers, Space vehicles, and UEs) of a future GNSS has been conducted in Section 7.1. Since a complete end-to-end deployment and validation of the whole system is beyond the scope that can be covered by this thesis, the simulation conducted to evaluate

the proposed new NAV messages in the following will concentrate on the core part – the new ephemeris embodied in the new NAV messages, trying to evaluate the feasibility to use the new ephemeris in ground navigation applications for up to 24 hours. With the fundamental feasibility confirmed, implementation details of the new NAV messages encoding, uploading, broadcasting, reception and the new ephemeris decoding are then operational implementation tasks.

7.2.1 Simulation and Validation of the New Ephemeris

This thesis is to leverage some earlier efforts spent in ephemeris extension to help validate the feasibility of the proposed new ephemeris, primarily from a perspective of orbital accuracy.

Ephemeris extended for 5 days has been successfully used in realistic navigation applications (Zhang, et al. 2008). Using the same technologies, there should be no technical barrier at all to prepare ephemerides that are valid for up to 24 hours, for the validation purpose of this thesis.

The traditional ephemerides are just outputs of orbital determinations of the GNSS satellites, which utilize precise ranging measurements from different GNSS satellites to globally distributed ground stations. However, in the simulated generation of the new ephemeris in this thesis, the observations of the GNSS satellite positions and velocities derived from the broadcast ephemerides (traditional ephemerides) up to a selected reference time are used to refine the satellite orbits, in an effort to extend the validity of the ephemeris starting from the selected reference time. In this way, the orbital determination process is significantly simplified while the system validation is not impacted.

As the selected time window for the validation is in the past, the traditional ephemerides are available for the whole time period starting from the reference time to the end of validity of the new ephemerides, therefore reference orbits can be generated from the traditional ephemerides and used to evaluate the new ephemerides in the orbital domain.

7.2.2 Simulation of the Use of the New NAV Messages

The simulation and validation are further extended to navigation domain in some selected scenarios. To simulate the use of the new ephemerides, data logs that were collected from live tests are used in post processing, in which the traditional ephemerides that were used in live tests are replaced by the simulated new ephemerides. Then the navigation performance using traditional ephemerides and the new ephemerides are compared in terms of positioning accuracy and availability.

The generation of the new NAV messages, however, are not directly simulated as it is reasonable to assume that the new NAV messages are not subject to decoding errors in weak signal condition within 24 hours since they have been decoded in elsewhere whenever with an open sky during that long period time. In contrast, the traditional NAV messages are more subject to decoding errors in the weak signal condition.

A typical scenario in urban canyon with challenging GNSS signal condition is illustrated in Figure 5-45 in Chapter 5. In that figure, a total of 11 GPS SVs are supposed to be in view at the selected UE location at the moment. However, only the signals from 3 GPS SVs are direct, whereas the signals of all remaining GPS SVs are blocked by high rise buildings. Certainly, with high sensitivity receivers, the number of signals in track could exceed 4. For the SVs that are not blocked by buildings, the received signals may be the composition of signals from direct paths and multipaths. For the blocked SVs, if in track, they are for sure purely reflected signals that may have experienced one or multiple reflections and therefore heavy attenuation in signal strengths. With the change of UE location, the situation of signal blockage may totally change. When driving through urban canyon, the GNSS signals tracked by the UE could experience frequent changes between blocked and unblocked, and it is really a rare case that a signal could be continuously received from the direct path for long. Therefore, timely decoding of the NAV message OTA in such condition seems a mission impossible, or extremely challenging if possible.

In an illustrated scenario in the Figure 5-44 in Chapter 5, 80% of the time, the signals can hardly stay stronger than 27 dB-Hz for continuously 6 seconds, indicating that continuously decoding a full sub-frame of NAV message would be difficult 80% of the time.

7.3 Results and Analysis

For the simulated use of the proposed new ephemeris, the results are firstly presented in the satellite orbital domain to illustrate the accuracy when the new ephemeris is used for 1 day (pursued life expectancy), to up to 5 days. Then, the results are presented in navigation domain when the proposed new ephemeris is used in some navigation applications. Going further, the typical benefits of using the proposed new ephemeris are examined in both realistic scenarios and theoretical analysis. Finally, the simulated use of the proposed new ephemeris is extended to a weak signal scenario, and the positioning performance is presented with insight details.

7.3.1 Evaluation of the New Ephemeris in Satellite Orbital Domain

The life expectancy of the proposed new ephemeris is 1 day, which is actually much shorter than what has been obtained in some earlier practice of ephemeris extension. Instead of presenting the accuracy of the new ephemeris for up to 1 day, the results from ephemeris extension for up to 5 days are presented, as a more convincing evidence to confirm the feasibility of the new ephemeris.

The orbital extension that will be illustrated in this section is based on the SPHRC solar radiation model, with BEs collected during the past ~ 2.5 days used as the observations for the estimation of ICs and solar radiation parameters. For each satellite, the estimated ICs and solar radiation parameters are then used to extend the orbit for the next 5 days. The extended orbits are evaluated by comparing against selected reference orbits, and 42 ground locations are selected to evaluate the orbital errors along different LOS directions.

The first two of following plots illustrate the orbital errors expressed in ECEF and RTN frames for two SVs.

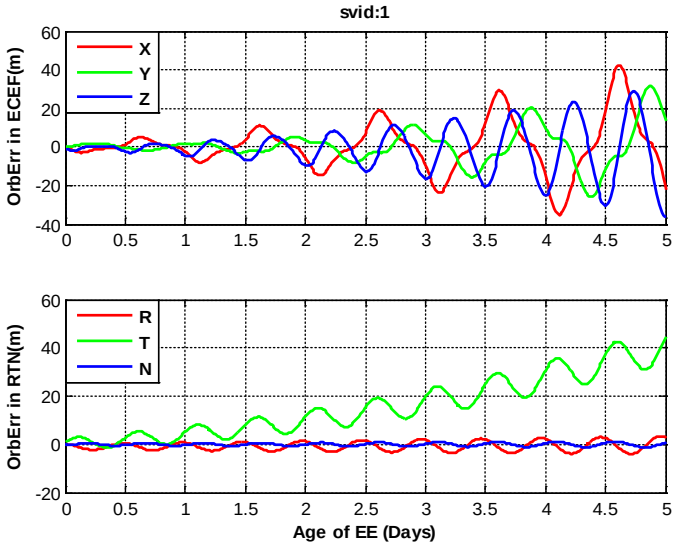


Figure 7-69 Errors of Extended Orbit for SV 1 Expressed in ECEF and RTN

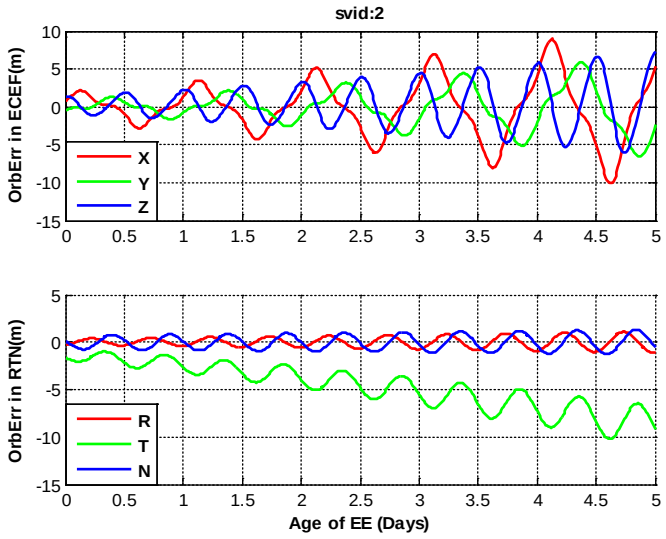


Figure 7-70 Errors of Extended Orbit for SV 2 Expressed in ECEF and RTN

Then in the next 3 plots, the errors of extended orbits expressed in RTN for all the processed SVs are put together for comparison. The results for different SVs consistently show that the orbital errors along tangent direction (along track) are the largest, and errors along normal direction (cross track) are the smallest.

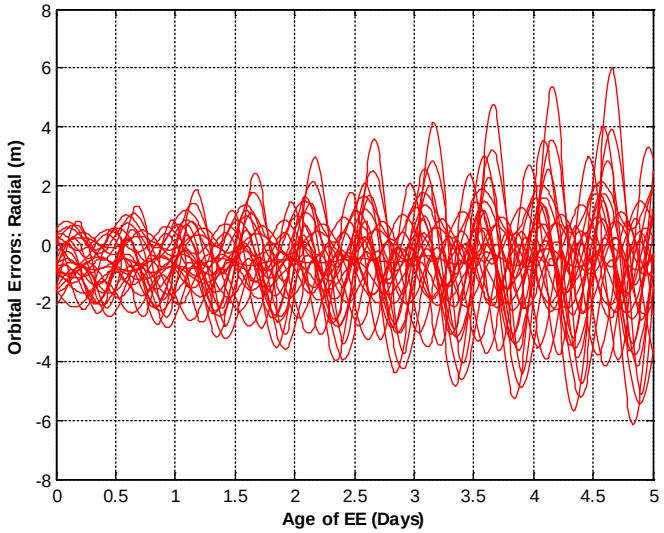


Figure 7-71 Errors of Extended Orbits for Different SVs in Radial Direction

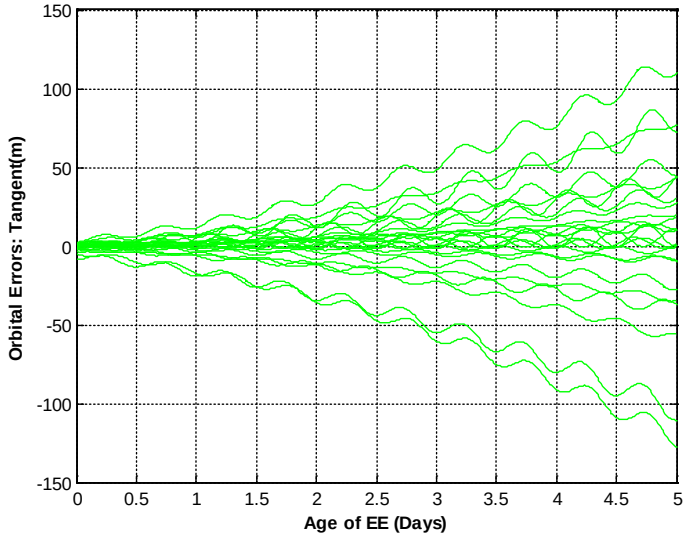


Figure 7-72 Errors of Extended Orbits for Different SVs in Tangent Direction

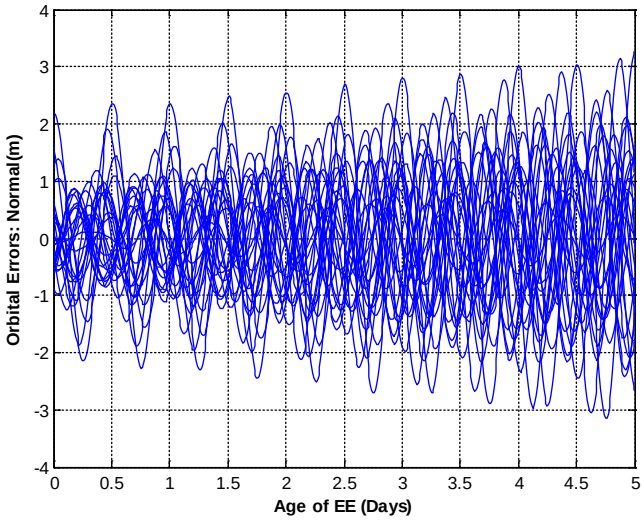


Figure 7-73 Errors of Extended Orbits for Different SVs in Normal Direction

On each extended day, the statistics of the URE is calculated in terms of 50%, 68% and 95% percentiles, and the results are used to quantify the errors of extended orbits.

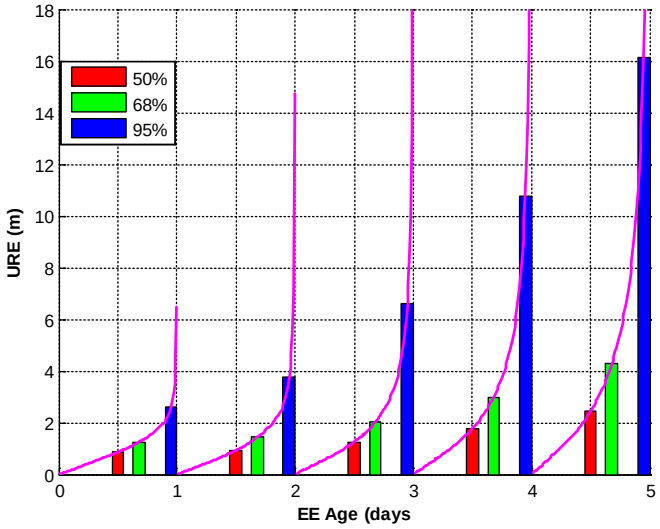


Figure 7-74 URE of Extended Ephemeris at Ages from 1 to 5 days

Table 7-16 Statistics of LOS Orbital Errors

Age of EE	URE (Errors)		
	50% (m)	68% (m)	95% (m)
1 Day	0.86	1.25	2.60
2 Days	0.92	1.42	3.73
3 Days	1.24	2.03	6.62
4 Days	1.76	2.97	10.78
5 Days	2.43	4.27	16.11

The URE statistics is dependent on the following factors:

- Number of samples
- Bin size of EE age, normally selected to be 1 day neglecting the periodically variations of the orbital errors in 1 day
- Number of satellites
- Number of ground locations selected to provide different LOS directions

So when mentioning the statistical performance, the definition of metrics should be mentioned as well.

7.3.2 Evaluation of the New Ephemeris in UE Navigation Domain

Similar to Section 7.3.1, instead of presenting the results of simulated use of the new ephemeris with life expectancy for up to 1 day, the results from simulated use of extended ephemeris with life expectancy for up to 5 days are presented, which hopefully are able to give a better picture of how the length of ephemeris life expectancy affects navigation performance.

7.3.2.1 Trajectories

The trajectory of a continuous driving test is included in this section to illustration the usage of BE and EE. Taking the trajectory of using true BEs as reference, different trajectories using EEs aged 0.4 day to over 4.7 days were generated and compared. For comparison purpose, all the trajectories using EEs at different ages are plotted overlapped in the same Figure below, along with the reference trajectory. With the synthetic influence of multipath, and errors in extended satellite orbits and clocks, the trajectories trajectory offsets become larger with age of EEs, but overall speaking, the trajectory accuracies are comparable.

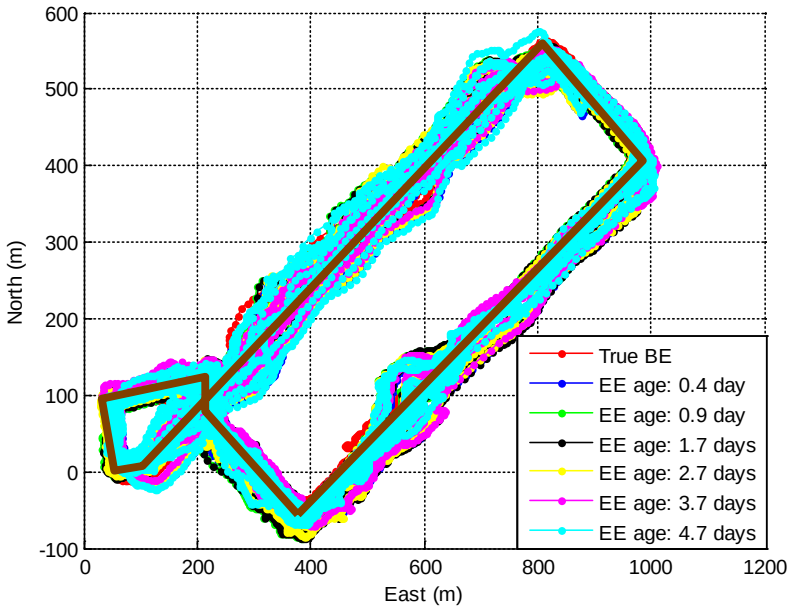


Figure 7-75 Trajectories of Using EE at Different Ages

7.3.2.2 Statistics of Cross Track Errors

The statistics of the cross track errors for those trajectories is further given in the table below. Using the true BEs, the mean cross track error, 9.43 m, is the smallest. Although using EEs aged up to 5 days, the mean cross track errors increased almost double, the accuracies of all the trajectories were still well within the acceptable range of vehicle navigation in urban areas, as very often the positions with errors up to half a block can still be projected to the right track through map matching techniques. Actually, for those trajectories using EEs, even the maximum cross track errors are still far less than half a block, almost the same as the maximum errors of the trajectory using true BEs.

Table 7-17 Cross Track Errors for the Trajectories using EE of Different Ages

Ephemeris Type	Ephemeris Age	Cross Track Errors w.r.t. Truth Trajectory (m)			
		Mean	RMS	Max	Min
BE	Fresh	9.43	12.36	60.39	0.00
EE	0.4 day	13.74	17.09	60.95	0.00
	0.9 day	16.05	19.48	68.06	0.13
	1.7 days	16.21	19.66	66.54	0.00
	2.7 days	14.9	18.47	64.4	0.00
	3.7 days	14.44	17.77	54.35	0.00
	4.7 days	12.62	17.24	68.12	0.48

7.3.2.3 LOS Errors of Used SVs during the Usage period

In the Table 7-17, something interesting is noticed: for the trajectory with EE aged 4.7 days, the mean cross track errors are even smaller than those with ‘younger’ EEs. It seems unreasonable at the first glance, but actually makes sense and it will be further interpreted in the orbital and clock accuracy analysis.

The understanding of the cross track errors in the trajectories requires a deeper insight in the error components and the pattern of the errors. In the above testing, multipath is a primary source of errors; when EEs are used, the errors in both the extended orbits and clocks are also error components that should be considered. This section is to give the insight on the first component of the EE errors - the extended orbit errors for the above test.

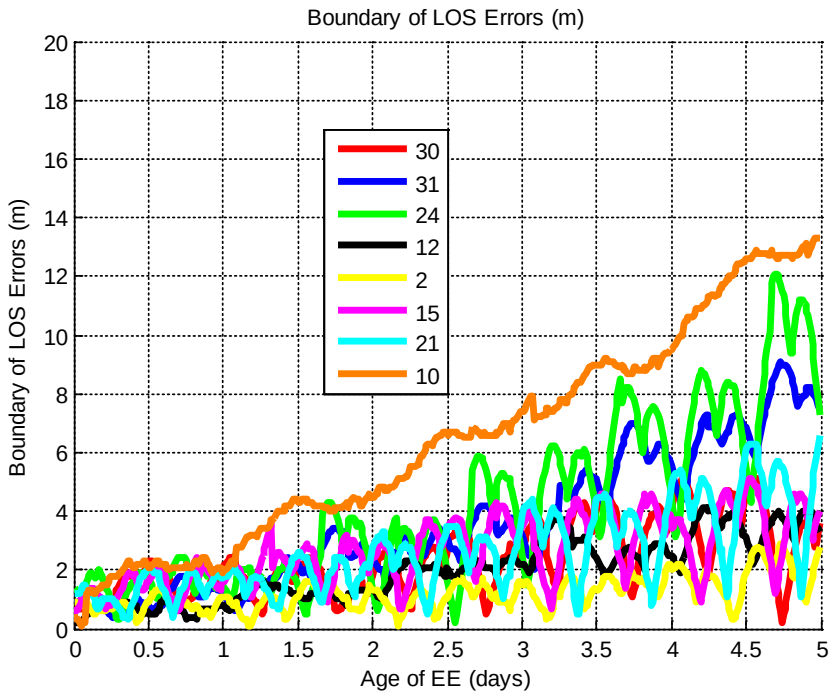


Figure 7-76 LOS Error bounds in Extended Orbits for the Used Satellites in above testing

During the time period of above trajectories, satellites PRNs 30, 31, 24, 12, 2, 15, 21, 10 were visible in that region. The last set of the above EEs was selected, and the aforementioned LOS error bounds of the extended orbits are plotted for these satellites in Figure 7-76. When this set of EEs was used, the age of each EE was around 4.7 days. According to Figure 7-76, the LOS error bounds for different satellite ranges from less than 2 m to around 13 m, which is very typical in large quantity of data analysis. In Figure 7-77, the actual LOS errors for each visible satellites during the testing are plotted, where the LOS errors are set to zero when the satellites were below horizon. At the time of use, 4.7 days, the actual LOS errors ranged from around -10 m to 12 m.

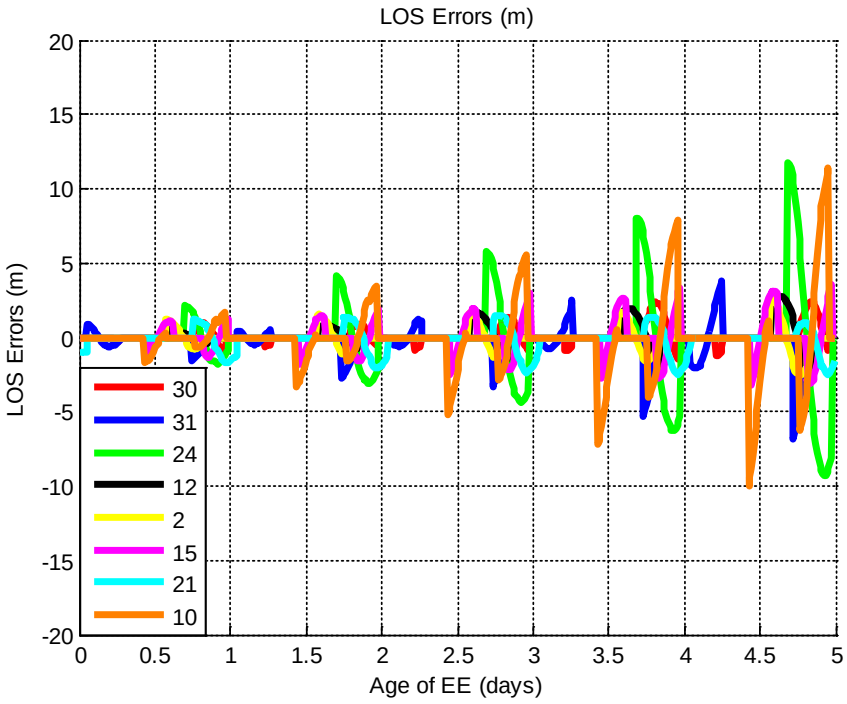


Figure 7-77 Actual LOS Errors in Extended Orbits for the Used Satellites during the test

7.3.3 Typical Benefits from the Use of Long Validity Ephemeris

A driving test through an urban downtown is adopted to illustrate the benefits of using EE in real navigation applications comparing to using BE, in which the following two scenarios are considered:

7.3.3.1 UE Startup in Open sky

With the receiver started in open sky before entering the urban canyon (with entrance marked by **purple** circle), when using BE, the chance for completing the ephemeris downloading for each visible satellites is increased, therefore the positioning performance is supposed to close to that

of using EE, if the EE accuracy is sufficient. The trajectory comparison for using BE and EE is given in the plot below, with the positioning accuracy and availability comparisons further given in the Table that follows.

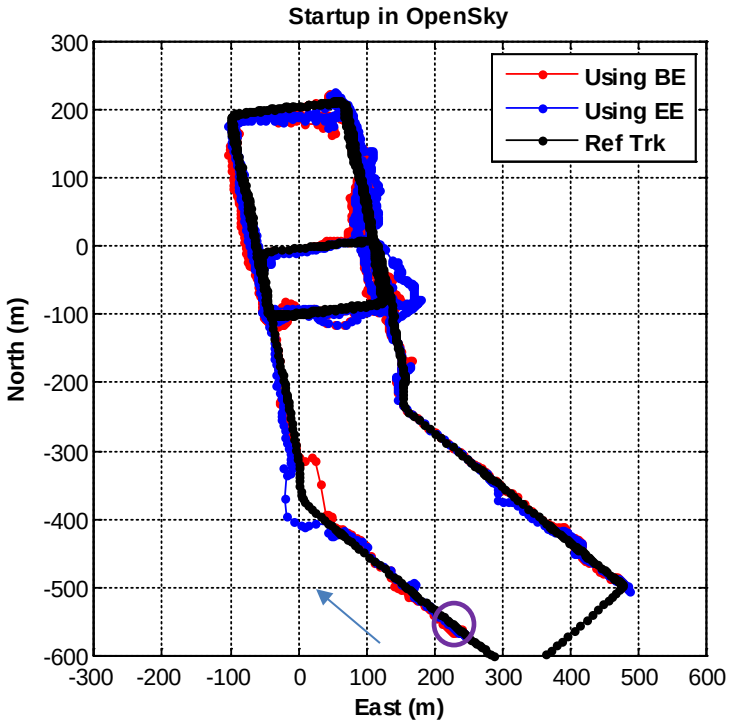


Figure 7-78 Trajectories of Using BE/EE, w/ Startup in Open Sky

Table 7-18 Position Accuracies and Availabilities, with Startup in Open Sky

	Using BE	**Using EE**
Mean	12.00	14.75
RMS	6.74	10.70
Min	0.59	0.19
Max	48.82	50.14
50%	11.16	11.60

68%	15.05	16.24
95%	25.05	40.39
Samples	2115	2115

The comparison of number of used satellites are further given below, which shows that when using EE, the average number of used satellites is only slight more than using BE, therefore, the difference in the positioning performance is expected to be slight, as given in above table and figure.

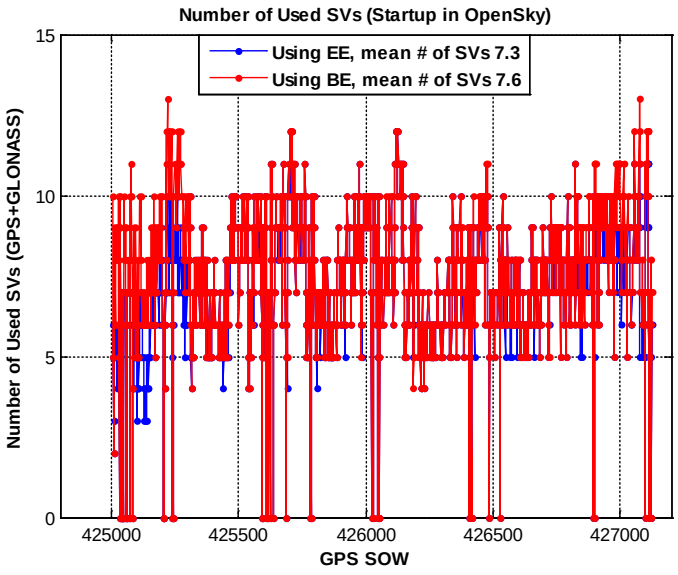


Figure 7-79 Comparison of Used SVs for Using BE and EE, with Startup in Open Sky

7.3.3.2 UE Startup in Urban Canyon

With the receiver started in downtown area, it takes much more time to complete the ephemeris downloading for 4 or more satellites to get the first position fix. However, when using EE in this case, the ephemeris for each visible satellite is immediately available, therefore it would be much

faster in getting the first position fix. The trajectory comparison for using BE and EE is given below along with the positioning accuracy and availability, which is further explained by the comparison of number of used satellites.

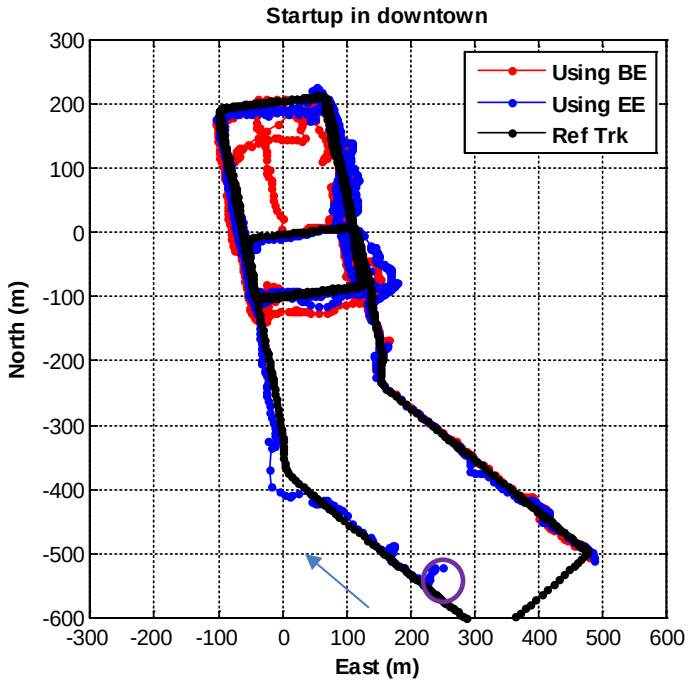


Figure 7-80 Trajectories of Using BE and EE, with Startup in Urban Canyon

Table 7-19 Position Accuracies and Availabilities, with Startup in Urban Canyon

	Using BE (m)	Using EE (m)
Mean	19.18	15.03
RMS	15.12	10.74
Min	0.07	0.19
Max	123.63	50.14
50%	16.66	11.79

68%	20.05	16.40
95%	51.72	40.41
Samples	1938	2115

As shown in the above figure and table, the difference becomes much larger between using BE and EE with the GNSS device started in urban canyon.

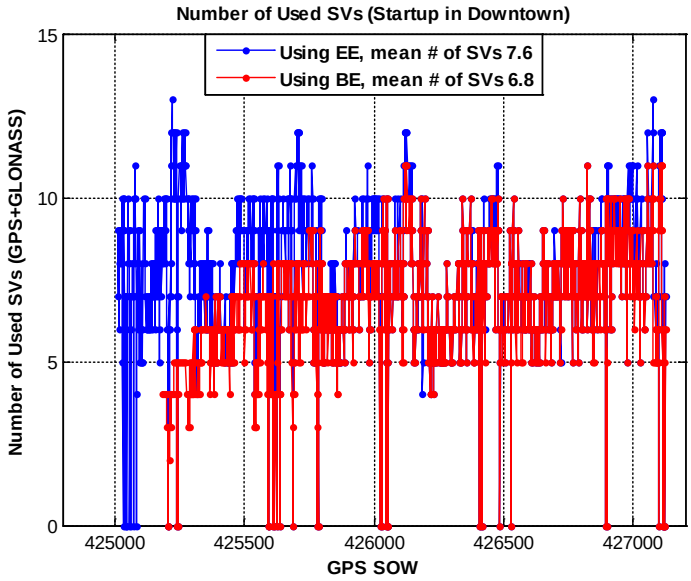


Figure 7-81 Comparison of Used SVs for Using BE and EE, w/ Startup in Urban Canyon

7.3.3.3 TTFF of Using the New NAV Messages

The TTFF performance using the new NAV messages is only analyzed from theoretical perspective due to the capability limitation for large quantity of simulations and verification. As shown in Figure 7-82, whenever assistance is available, the warmstart TTFF of 1 second is achievable on a current GPS UE and a future GNSS UE with the new NAV messages; whenever

assistance is not available, the future GNSS with the new NAV messages has advantage over current GPS with the inherent capability to reduce the TTFF by 33%, namely from 18 seconds to 12 seconds. Moreover, to maintain the comparable performance on TTFF for a time span of 24 hours, a UE of current GPS needs 12 times of ephemeris assistance but a future GNSS UE with the new NAV messages only needs once.

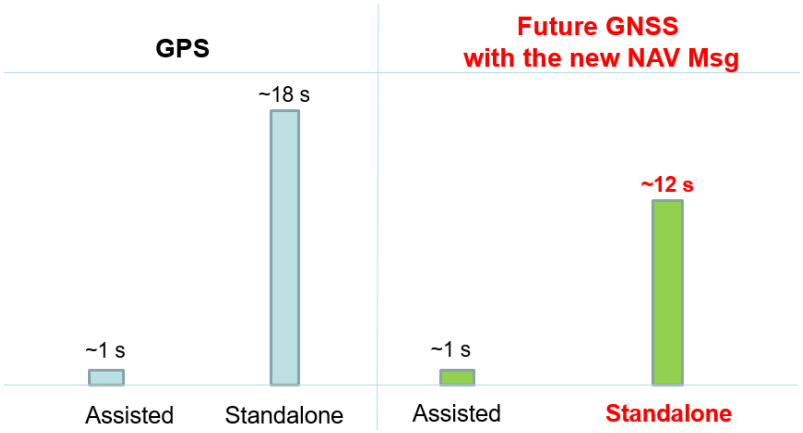


Figure 7-82 Theoretical WarmStart TTFF for GPS and future GNSS with the new NAV Msg

7.3.4 Simulated Use of the New NAV Messages in Weak-Signal Scenario

Comparing to the tests in the last section, a more challenging area in downtown San Francisco is selected for the simulated use of new NAV messages. As the new NAV messages aim to be valid for 24 hours, by shifting the reference time of the NAV messages, the simulation created scenarios for the use of the new NAV message for up to 2, 3, 4, 6, 10, 14, 18, 22 hours. In Figure 7-83, the trajectories of all the simulated scenarios are plotted together, with the reference

trajectory plotted in solid brown. The test trajectory lasts for ~900 seconds, consisting of 4 laps across the area of dense high-rise buildings.

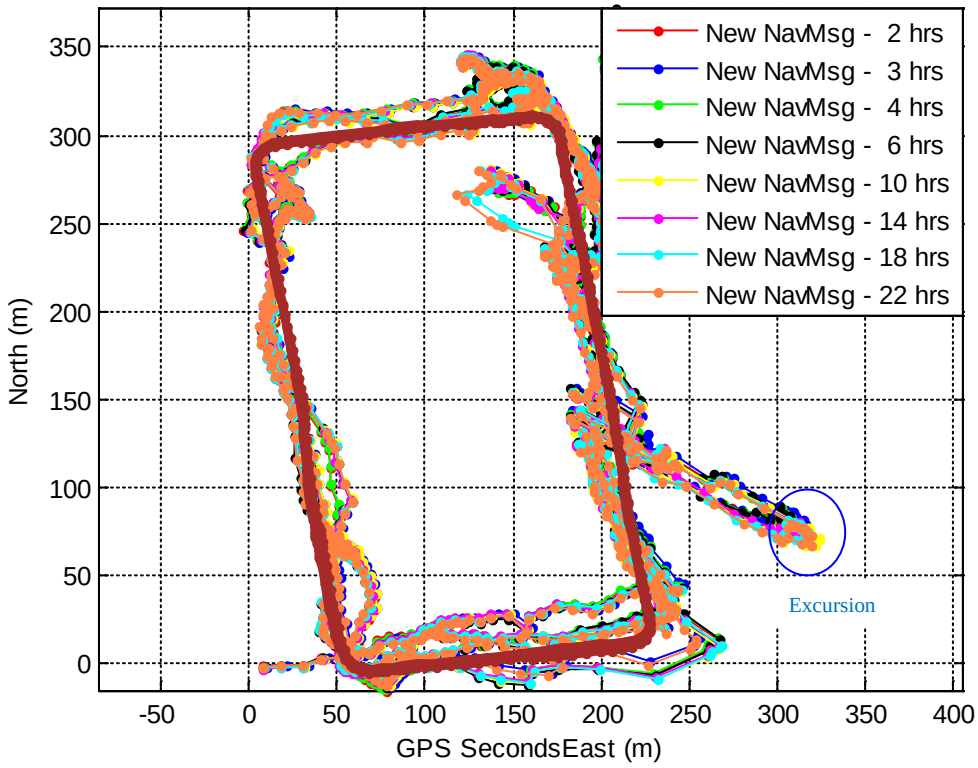


Figure 7-83 Comparison of Trajectories from Simulated Use of the New NAV Messages at Different Ages (< 24 hrs)

At the first sight, the positioning performance in the above Figure is really not impressive, regardless of the age of the used new NAV messages. Taking a look at the average signal strength in Figure 7-84 during the test, we can get more insight of the challenges. As shown in the figure, 97% of the time, the average CN0 is below 27 dB-Hz, which gives the facts that the measurement quality is really a concern given so weak signal conditions, and also decoding the NAV messages is almost impossible during the test time duration according to the discussion in Section 2.4.2.

For current GPS, if the expiration of NAV messages happens during the above 900 s, there is potential risk that the UE is unable to operate properly. For a UE of current GPS, it could experience such risks 12 times every 24 hours. However, with the new NAV messages deployed, the chance of such risk is reduced to once per 24 hours. Certainly, the GNSS industry would not allow such risk to jeopardize the continuous operation of GNSS UEs, and that's why AGNSS (through SUPL) and EE (through other data channels) are deployed, at the cost of system complexity when addressing such risks.

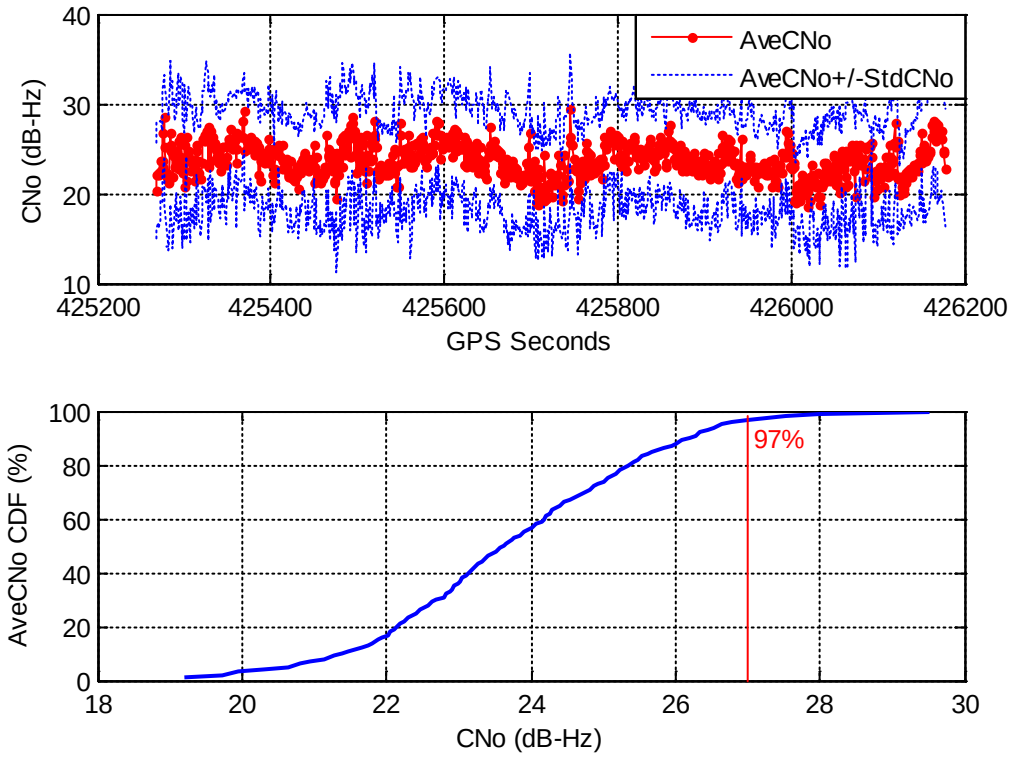


Figure 7-84 Average Signal Strength during the Test in Urban Canyon

After taking a look at signal environment at the largest excursion in Figure 7-83 on the right side, it gives further insight of the challenges of positioning under weak signal condition. In Figure 7-85, the direct signal paths from all used GPS SVs by the UE are visualized in the place where the largest excursion occurred.

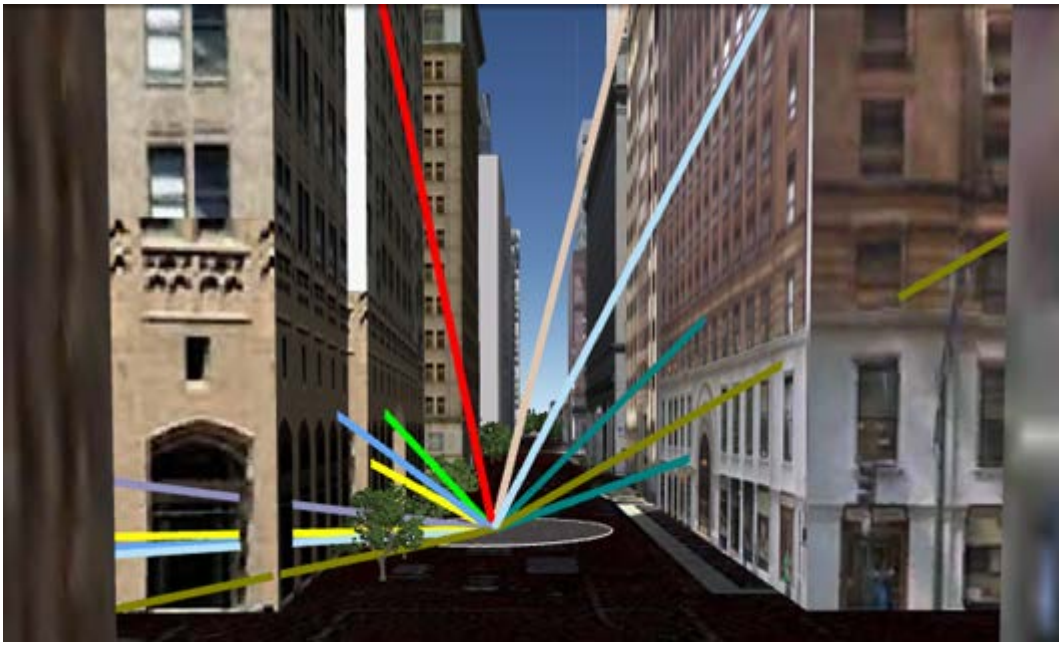


Figure 7-85 Street View of the Signal Environment where the Large Exursion Occurred

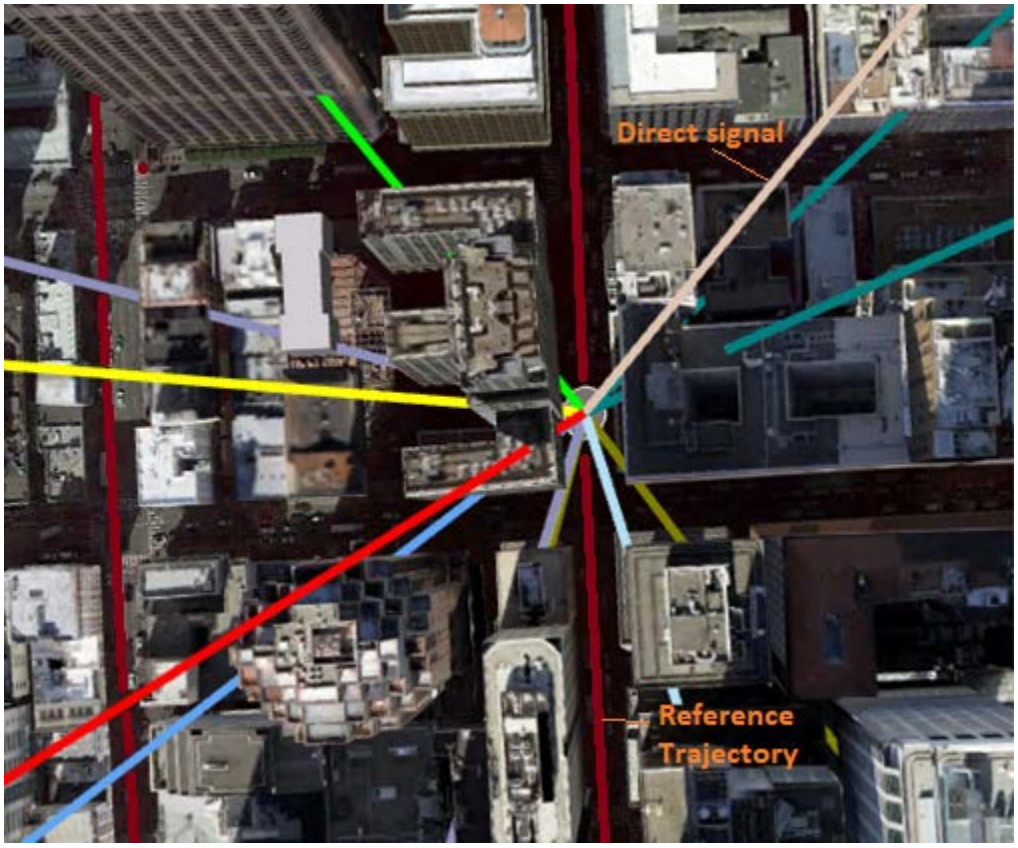


Figure 7-86 Signal Condition whereat the Large Exursion Occurred

Dense high-rise buildings on the two sides of the streets make up a ‘real’ urban canyon along the test route, leaving very limited sky in view to the GNSS UE to get direct signals from the satellites. Taking a different perspective, it is clearly shown in Figure 7-86 that, all the signals are blocked by the tall buildings except one SV. For the blocked SVs, the signals are obviously from reflections; and for the only 1 direct signal, the received signal is also very likely the superposition of direct and reflected signals. So the measurements from the SVs at the moment are believed suffering from serious multipath impacts, in the range of less than 300 meters for the SV with direct signal, and possibly more than 300 meters for the SVs with purely reflected signals.

To better understand the impact of the use of the new NAV messages, the East and North components of the trajectory errors from Figure 7-83 are further plotted in Figure 7-87, in which the aforementioned large excursion corresponds to GPS seconds 426090. The differences among the trajectories, which are believed related to the ages of the different new NAV messages, seem really insignificant comparing to the magnitude of the trajectory offsets. Figure 7-88 gives more insight of the differences of the trajectories comparing to the trajectory using the New NAV messages of 2 hours, which indicates that most of the time the difference is in the range of a few meters, but occasionally amounts to 20 meters.

Could the new NAV messages of different ages from 2 to 22 hours cause such differences? Such differences are highly related to, but could not be uniquely ascribed to the different ages of the New NAV messages. For the new NAV message, with the ephemeris used for 2 hours, or up to 24 hours, the resulting orbital errors should not be sub-meters, according to the studies in earlier Chapters. The errors in the currently used GPS satellite clock, when extended for 24 hours, could be a few meters in the worst cases, but should appear more like constant during a short time period of 900 s. In the weak signal environment, when multipath become dominant in the measurement errors, the orbital and clock errors in the 24-hour ephemeris are just like a glass of water being added to a bucket of water, and could hardly affect the overall error magnitudes essentially.

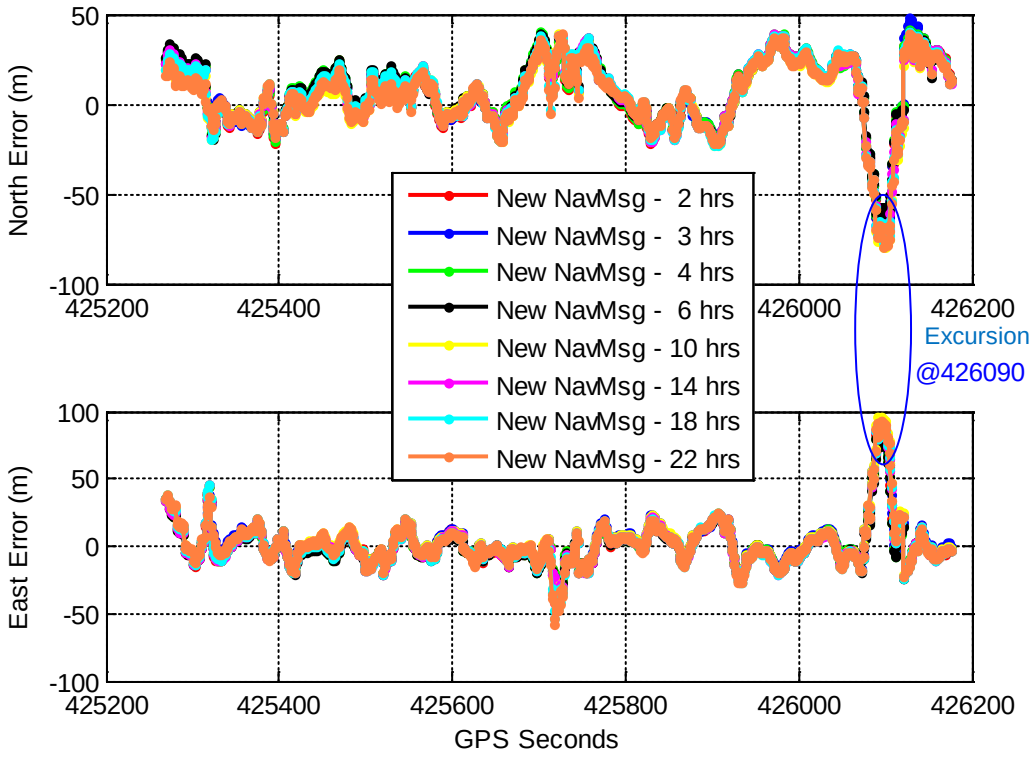


Figure 7-87 Trajectory Errors in East/North for the Simulated Use of New NAV Messages in a Weak Signal Scenario

The trajectory differences plotted in Figure 7-88, are primarily related to the implementation of the Kalman Filter, in which the scheme of sequential measurement update could be adopted for calculation efficiency and the performance is sensitive to the update sequence of the measurements even if there is no change in each measurement's error. The orbital and clock errors in the ephemeris at the age of 2 hours to 22 hours, although don't differ too much, they could have changed the sequence of the measurement update, and resulted to the trajectory differences.

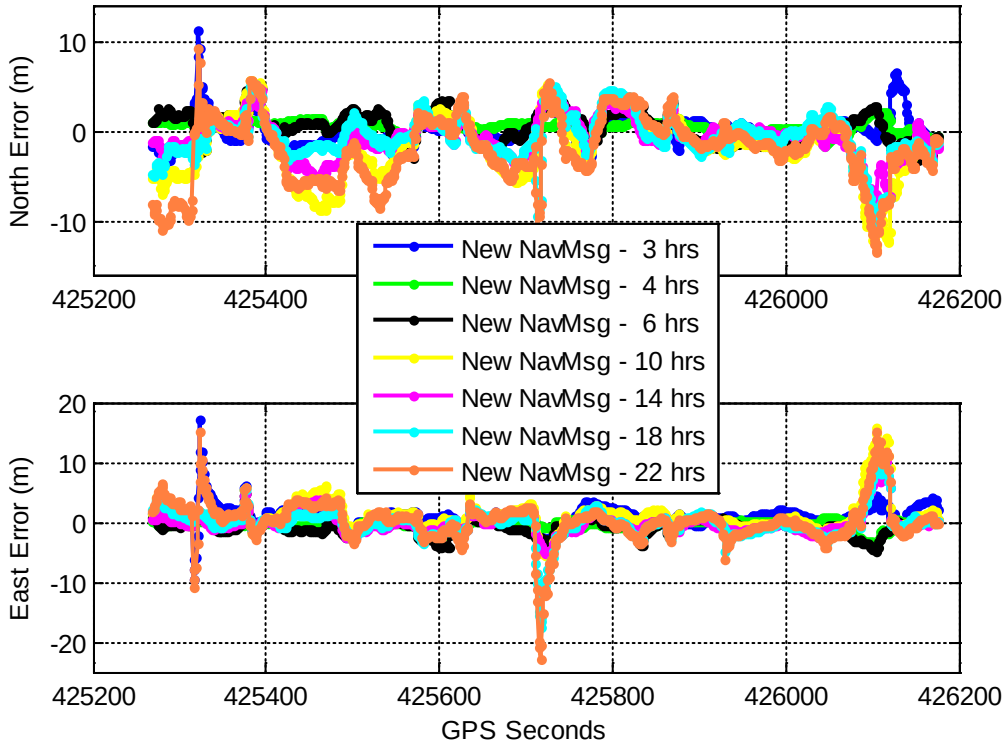


Figure 7-88 Difference in the Trajectories with New NAV Messages in a Weak Signal Scenario

Taking a statistical comparison, the horizontal errors from each trajectory in Figure 7-83 are further listed in Table 7-20 and plotted in Figure 7-89, which indicate that, with the increase of the age from 2 to 22 hours in the simulated new NAV messages, there is no obvious trend in the positioning errors to increase.

Table 7-20 Horizontal Position Errors in the Trajectories Using the Simulated New NAV Msg at Different Ages

Age of New NavMsg	Horizontal Position Errors w.r.t. Reference Trajectory (m)			
	Mean	50%	68%	95%
2 hrs	19.94	15.861	21.797	47.247

3 hrs	20.26	16.61	20.976	47.965
4 hrs	20.21	16.06	22.185	48.107
6 hrs	20.05	15.88	21.892	47.454
10 hrs	20.13	15.559	21.155	47.854
14 hrs	20.49	16.234	22.883	48.314
18 hrs	20.44	15.592	21.992	49.163
22 hrs	19.39	13.785	20.739	48.251

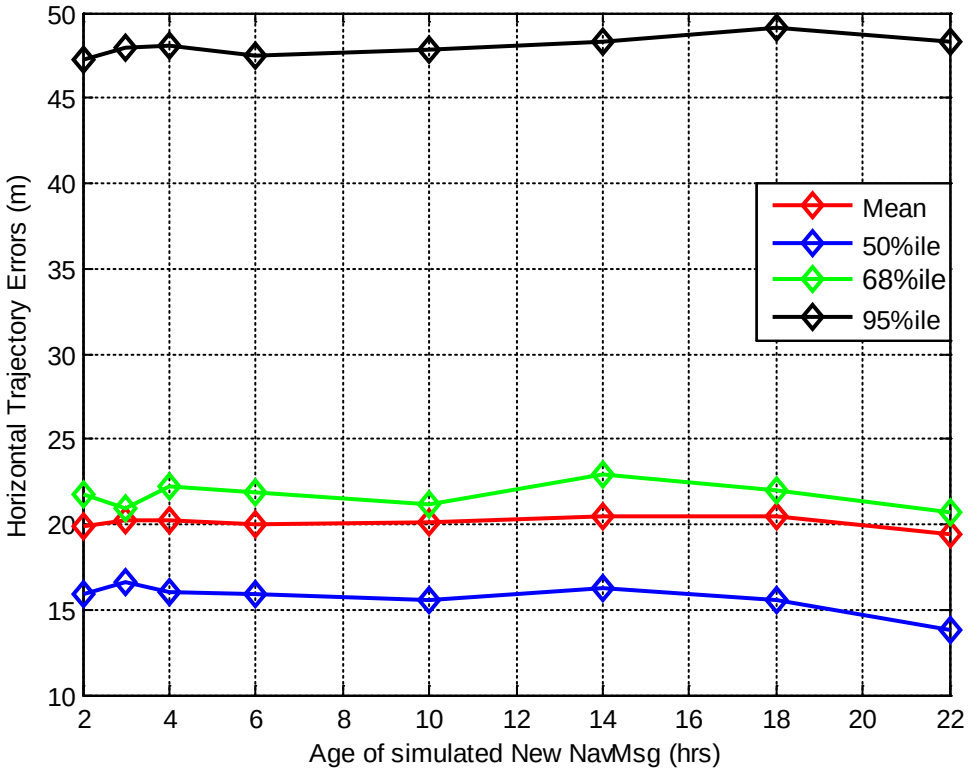


Figure 7-89 Horizontal Position Errors in the Trajectories Using the Simulated New NAV Msg at Different Ages

CHAPTER 8 CONCLUSIONS AND FUTURE WORK

This chapter will summarize the research work in this thesis, including the feasibility of the proposed new NAV messages and the performance of GNSS that deploys them.

8.1 Thesis Summary

This thesis introduces the fundamental of TTFF and sensitivity in a GNSS User Equipment, through the explanation of the detailed steps in the start-up of a stand-alone GNSS UE. The range of time needed and the sensitivity limitation in each step are given, and the most time-consuming step that affects the overall TTFF is pointed out. For the challenges in fast TTFF and high sensitivity in standalone GNSS, the reasons are analyzed in this thesis and attributed to the short validity period of the NAV messages and the ephemeris.

Then this thesis takes deep insight into the theory of extended ephemeris, by studying the fundamental frames in time and space, the different force models for orbital determination, and estimation methodologies for initial condition and satellite clocks. The study indicates that the modeling of solar radiation is critical to ensure sufficient orbit accuracy beyond days for navigation on GNSS UE.

Next, the thesis furthers the study on how the TTFF of a stand-alone GNSS UE can be improved from ephemeris with long validity beyond 2 hours, by exploring the detailed tasks done in each start-up step and carefully examining the impact and potential benefits from the long validity

ephemeris. As indicated by the study, with the long validity ephemeris, the procedures in signal acquisition, time synchronization and position fixes can be significantly accelerated with or without additional techniques.

Subsequently, the sensitivity of a GNSS UE is categorized and the impact of long validity ephemeris on each category is studied. As indicated, the sensitivity in signal acquisition can be improved by better accuracy of predicted code phase and Doppler, the sensitivity in signal tracking can be improved by long coherent integration that is facilitated by long validity NAV messages, and ephemeris decoding is no longer subject to the 27 dB-Hz limitation during the validity period of the long validity ephemeris, which is equivalent to the improvement in the ephemeris decoding sensitivity. In addition, the long validity ephemeris enables the position fix in some scenarios where position fix was not possible because of the difficulty to get updated ephemeris on time. Certainly, the enablement of position fix in such scenarios also brings new challenges, such as discrimination between purely reflected signal and weak direct signal, which need to be resolved by additional outlier detection techniques and even heuristic techniques.

Based on the analysis of the fundamental weakness in the design of the NAV messages for the existing GNSS, and also based on the study of the impact and benefits of long validity ephemeris on both TTFF and sensitivity of a stand-alone GNSS UE, the thesis eventually proposes a new set of content for NAV messages, so that they can be valid for up to 24 hours – a much longer valid period, which is able to inherently facilitate the fast TTFF and sensitivity on stand-alone GNSS UEs.

To verify and illustrate the feasibility of the proposal, some comparisons are conducted in both open sky and urban canyon environments between tests using the traditional and the newly proposed ephemeris. Taking the UE performance using the traditional ephemeris as reference, the frequently changed ephemeris is then replaced with the new ephemeris, and the performance for using the new ephemeris at different ages from within 2 hours to up to 24 hours is analyzed and compared, including the availability and accuracy.

8.2 Conclusions

The work of the theoretical exploration and practical considerations in this thesis, enables the following conclusions:

(1) Fast TTFF and high sensitivity is the must-to-have performance in modern GNSS UEs;

(2) Fast TTFF and high sensitivity is challenging in the existing stand-alone GNSS UEs, and that's why additional assistance systems are needed;

(3) The challenges in the fast TTFF and high sensitivity in the existing stand-alone GNSS UEs, is due to the fundamental design weakness in GNSS NAV messages – too short validity period;

(4) Extending the validity period of ephemeris to 24 hours is feasible according to the theory study on the orbital determination and clock modeling;

(5) The proposal to broadcast new NAV messages with validity up to 24 hours minimizes the need to decode ephemeris OTA, and therefore enables stand-alone GNSS UEs significantly less susceptible to the change of signal strength from the perspective of ephemeris decoding OTA;

(6) The proposed new ephemeris, as long as is decoded once per day, guarantees the availability in the rest of the day, therefore it is able to save the time for ephemeris decoding OTA and enables fast TTFF in each subsequent UE start-ups;

(7) The proposed new NAV messages, as long as decoded once per day, can be repeatedly used in the rest of the day to wipe off the data bits in signal tracking, therefore the obtaining higher tracking sensitivity is greatly facilitated;

(8) Due to the short validity period (2-4 hours), the traditional ephemeris is subject to frequent expiration, and the accuracy of traditional ephemeris rapidly degrades beyond the validity period. However, with one set of the proposed new ephemeris, satellite orbit and clock for 24 hours can be obtained at equivalent accuracies from 12 valid sets of traditional ephemeris. Therefore, use of the proposed new ephemeris, is able to narrow down the uncertainty range of predicted code phase and Doppler of visible signals within 24 hours, which definitely speeds up the signal acquisition and facilitate the improvement in acquisition sensitivity;

(9) For the benefits of using the proposed ephemeris, there is cost on the end of GNSS UE – namely the increased computation load in deriving the satellite position and velocity, comparing to using traditional Keplerian ephemeris. However, such cost is not an issue given the computation power of mobile devices nowadays, not to mention the advancement in the near future.

(10) Power saving is so important to modern GNSS UEs: keeping the GNSS UEs sleep until there is need for the GNSS UEs to wake up for position fixes or decoding of updated NAV messages. This has been always given special attention. Use of the proposed NAV messages can be a viable solution for this purpose, as once the NAV messages are

decoded OTA, there is no need for the device to wake up just to decode ephemeris within up to 24 hours.

8.3 Future work

For the proposed new NAV messages, this thesis primarily focuses on explanation, illustration of the concept, exploration of the theoretical fundamentals, identification of the key issues and solutions, and discussion of the potential benefits. Considering that a complete end to end implementation and verification of the concept is out of the scope of this thesis, this thesis primarily focuses on the implementation of the core technical parts - orbital determination and ephemeris extension, and then uses the offline generated ephemeris to replace the broadcast ephemeris in the data logs collected from some live tests, as an effort of the concept verification. It is expected that GNSS system developers, would assess and implement the proposed NAV messages for its feasibility and potential adoption by operational navigation systems. The benefits of the new NAV messages to other applications to enhance GNSS availability and sensitivity should also be investigated in the future. Given below are some future works:

(1) Based on the proposed contents of the new NAV messages, push the design to more details, such as the bit allocation for each component, so that the contents can be easily absorbed in the superframe of a GNSS to replace old NAV messages for verification;

(2) Considering that the proposed new ephemeris occupies much smaller number of data bits the regular ephemeris, the length of the superframe may become much shorter and therefore the superframe structure may needs modification;

(3) Implementation of the new NAV messages after orbital determination;

(4) Further investigate the efficiency of the error-symmetry-based URA algorithm for different GNSS constellations during different time windows;

(5) Standardize the algorithms for satellite position and velocity calculation using the new ephemeris;

(6) Dealing with orbital maneuvers.

APPENDIX

- **Conversion from Cartesian to Keplerian**

Firstly, a normal vector should be derived from the two vectors $\mathbf{r}_s$, and $\dot{\mathbf{r}}_s$:

$$\mathbf{H} = \mathbf{r}_s \times \dot{\mathbf{r}}_s = |\mathbf{H}| \cdot \mathbf{e}_H \tag{A-1}$$

where $\mathbf{e}_H$ is a unit vector that has been introduced in Table 3-5.

Based on the unit normal vector $\mathbf{e}_H$, the Orbital Inclination is directly available:

$$i = \arccos(e_{H_Z}) \tag{A-2}$$

So is the Right Ascension of the Ascending Node:

$$\Omega = \arctan\left(e_{H_X}, \quad -e_{H_Y}\right) \tag{A-3}$$

Then the Semi-major Axis can be derived as:

$$a = -\frac{G}{2 \cdot \left(\frac{|\dot{\mathbf{r}}_s|^2}{2} - \frac{G}{|\mathbf{r}_s|} \right)} \tag{A-4}$$

And the Orbital Eccentricity:

$$e = \sqrt{1 - \frac{|\mathbf{H}|^2}{a \cdot G}} \tag{A-5}$$

Next, the Argument of Perigee:

$$\omega = \arctan\left(\frac{Z}{\sin(i)}, \quad X \cdot \cos(\Omega) + Y \cdot \sin(\Omega)\right) - \nu \tag{A-6}$$

Finally, the True anomaly can be derived as:

$$\nu = \arccos\left(\frac{a \cdot (1 - e^2) - |\mathbf{r}|}{e \cdot |\mathbf{r}|}\right) \tag{A-7}$$

REFERENCES

Allan D W. (1987) Time and frequency (time-domain) characterization, estimation, and prediction of precision clocks and oscillators [J]. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC - 34, 647 - 654.

Akopian; David, Syrjarinne; Jari (2002) Method, apparatus and system for estimating user position with a satellite positioning system in poor signal conditions, Nokia Corporation, US Patent No.: 6,473,694, Granted on October 29, 2002

Ash, M. E. (1972) Determination of Earth Satellite Orbits, Tech. Note 1972-5, 258 pp., Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, 1972.

Axelrad, P. and Parkinson, B.W. (1988) Autonomous GPS Integrity Monitoring Using the Pseudorange Residual. Navigation, 35, 255-274.

Barnes J A, Allan D W. (1985) Time scale stabilities based on time and frequency kalman filters[C] Proceedings of the 39th Annual Symposium on Frequency Control, 1985: 29 - 31.

Bar-Sever, Y.E., 1994, “New GPS attitude model,” IGSMAIL #591, http://igscb.jpl.nasa.gov/mail/igsmail/1994/msg00166.html (accessed June 20, 2008).

Bar-Sever, Y. and D. Kuang (2004) New Empirically Derived Solar Radiation Pressure Model for Global Positioning System Satellites, IPN Progress Report 42-159, November 15, 2004

BDS ICD (2013) "BeiDou Navigation Satellite System Signal In Space Interface Control Document, Open Service Signal (Version 2.0)", China Satellite Navigation Office, December 2013

Braasch, Michael S.; A. J. Van Dierendonck (1999) GPS Receiver Architectures and Measurements, Proceedings of the IEEE, Vol. 87, No. 1, January 1999 pp. 48 – 64

Breakiron L A. (2001) A kalman filter timescale for atomic clocks and timescales[C] Proceedings of the 33rd Annual Precise Time and Time Interval (PTTI) Systems and Applications Meeting, 27 - 29 November 2001:431-443.

Broadcom, "Long Term Orbits (LTO) with Assisted-GPS (AGPS)", Available at: http://www.broadcom.com/products/GPS/Location-Based-Services/LTO-AGPS, Retrieved on Feb. 5, 2014

Brown R. G. (1992) A Baseline GPS RAIM Scheme and a Note on the Equivalence of Three RAIM Methods, NAVIGATION: Journal of The Institute of Navigation, Vol.39, No.3, Fall 1992.

Beutler, G., E. Brockmann, W. Gurtner, U. Hugentobler, L. Mervart, and M. Rothacher (1994), Extended Orbit Modeling Techniques at the CODE Processing Center of the International GPS Service for Geodynamics (IGS): Theory and Initial Results Manuscripta Geodaetica, vol. 19, pp. 367--386, April 1994.

CCAR (2009) Introduction to Statistical Orbit Determination: Kepler Orbit Elements to ECI Cartesian Coordinates Conversion, Elliptical Case, http://ccar.colorado.edu/asen5070/handouts//kep2cart_2002.doc

Chapman (2000), Chapter 4: Radar Detection, © 2000 by Chapman & Hall/CRC, available at: http://dsp-book.narod.ru/RSAD/C1828_PDF_C04.pdf

Chen, J., J. Wang (2005) Models of Solar Radiation Pressure in the Orbit Determination of GPS Satellites, Chinese Astronomy and Astrophysics 31 (2007) 66-75

Col Rick Reaser (2002), GPS Modernization, ATCA Symposium, Washington, D.C., 15 January 2002

CSR, "SiRFInstantFix™ SGEE/CGEE", Available at: http://www.csr.com/products/39/CSR-instantfix-sgeecgee, Retrieved on Feb. 5, 2014

Dennis Akos (1997) A So‡ware Radio Approach to Global Navigation Satellite System Receiver Design. Ph.D. Dissertation. Ohio University, Athens, OH.

DFC (2007) GPS Baseline Receiver Toolbox, User's Guide v1.1, Data Fusion Corporation

Dilssner, Florian (2010) GPS IIF-1 Satellite Antenna Phase Center and Attitude Modeling, Inside GNSS, September 2012, Page 60-64, http://www.insidegnss.com/auto/sep10-Dilssner.pdf

Eanes, R.J., R. S. Nerem, P.A.M Abusali, W. Bamford, K. Key, J.C. Ries and B.E. Schutz (1998) "GLONASS Orbit Determination at the Center for Space Research", Center for Space Research The University of Texas at Austin

Elliott D. Kaplan, Christopher Hegarty (2005), Understanding GPS: Principles and Applications, Second Edition, Artech House Publishers 2005

FCC (2001), "Enhanced 9-1-1 – Hearing and Statements", Retrieved on Feb 5, 2014. Available at: http://transition.fcc.gov/pshs/services/911-services/enhanced911/archives.html

Fliegel, H. F., and T. E. Gallini (1989), Radiation Pressure Models for Block II GPS Satellites, in Proceedings of the Fifth International Symposium on Precise Positioning with the Global Positioning System, pp. 789--798, National Geodetic Survey, NOAA, Rockville, Md., April 1989.

Fliegel, H. F., W. A. Feess, W. C. Layton, and N. W. Rhodus (1985), The GPS Radiation Force Model, in Proceedings of the First International Symposium on Precise Positioning with the Global Positioning System, edited by Clyde Goad, pp. 113--119, National Geodetic Survey, NOAA, Rockville, Md., March 1985.

Fliegel, H.F, and Gallini, T.E. (1996), Solar Force Modeling of Block IIR Global Positioning System Satellites, Journal of Spacecraft and Rockets, vol 33, No. 6, 1996.

Fliegel, H.F, and T. E. Gallini, and E. R. Swift (1992) Global Positioning System Radiation Force Model for Geodetic Applications, Journal of Geophysical Research, Vol. 97, pp. 559-568, 1992

Diggelen, Frank van (2009) A-GPS: Assisted GPS, GNSS, and SBAS, Boston | London: Artech House

Diggelen, Frank van (2002), “Method and Apparatus for Time-free Processing of GPS Signals”, US Patent No.: US 6,417,801 B1, Granted on July 9, 2002

Driver, Ted (2007) Long-Term Prediction of GPS Accuracy: Understanding the Fundamentals, ION GNSS 2007, Fort Worth, TX, pp.152–163.

Gallini, T. E. and H. F. Fliege (1995) The Generalized Solar Force Model, Aerospace Report No. TOR-95 (5473)-2, The Aerospace Corporation, El Segundo, CA

Garin, Lionel J.; M. S. Phatak, “Determing Position without Use of Broadcast Ephemeris Information”, US Patent No.: 7,403,154, Granted on July 22, 2008

GALILEO ICD (2008) “GALILEO Open Service – Signal in Space Interface Control Document (OS SIS ICD), Draft 1”, European Space Agency / European GNSS Supervisory Authority, February 2008

Gibbons, Glen (2008) “Boeing Wins NRL Contract to Continue Iridium/GPS Development”, *Inside GNSS*, September/October 2008. Available at: http://www.insidegnss.com/node/745, Retrieved on Feb. 5, 2014

GLONASS ICD, “GLONASS Interface Control Document (ICD), Navigational Radio Signal In bands L1, L2 (Edition 5.1)”, Russian Institute of Space Device Engineering, Moscow, 2008

GPS Business News (2009), “SiRFstarIV launched: low power & continuous ‘hot-start’” http://www.gpsbusinessnews.com/SiRFstarIV-launched-low-power-continuous-hot-start_a1667.html?print=1, Tuesday July 28, 2009

GPS IS (2004) “Navstar GPS Space Segment/Navigation User Interface”, GPS Interface Specification IS-GPS-200, Rev. D, GPS Joint Program Office and ARINC Engineering Services, 2004

GPS World Staff (2010) The System: GLONASS Forecast Bright and Plentiful, GPS World, Oct 1, 2010, available at: http://gpsworld.com/the-system-glonass-forecast-bright-and-plentiful/

Haley, D. (1973) Solar Radiation Pressure Calculations in the Geodyn Program, EG&G Report 008-73, Prepared for NASA Goddard Space Flight Center

Han, Shaowei (2009) “Method and Apparatus in Positioning without Broadcast Ephemeris”, US Patent No.: 7,564,406, Granted on July 21, 2009

Han, Shaowei and Zhang, Wentao (2007) A method and apparatus in positioning without broadcast ephemeris – message transmission, 2007, patent pending, No. 12/476,458

Hein G., A. Teuber, H. Thierfelder and A. Wolfe (2008) "GNSS Indoors, Fighting the Fading, Part 2" (2008), available at: http://www.insidegnss.com/auto/mayjune08-WP-part2.pdf, InsideGNSS, Retrieved on Oct. 5, 2017

Herring T. A., R. W. King and S. C. McClusky (2010) GAMIT Reference Manual - GPS Analysis At MIT, Release 10.4, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology

IERS (2017), IERS Bulletins , https://www.iers.org/IERS/EN/Publications/Bulletins/bulletins.html, Retrieved in Oct. 2017
Iridium (2012) ISU AT Command Reference, Iridium Proprietary, https://cdn.sparkfun.com/datasheets/Wireless/General/IRDM_ISU_ATCommandReferenceMAN0009_Rev2.0_ATCOMM_Oct2012.pdf, Retrieved in Oct. 2017

ICWG (2011), Change Topic: User Range Accuracy (URA) Definition, http://www.gps.gov/technical/icwg/meetings/2011/09/13/WAS-IS-FINAL_URA_Definition_6May2011.pdf

Iridium (2016) Iridium Launches Breakthrough Alternative Global Positioning System (GPS) Service, http://investor.iridium.com/releasedetail.cfm?releaseid=972324, Retrieved in Oct 2017

James Boa-Yen Tsui (2000) Fundamentals of Global Positioning System Receivers, A Software Approach, John Wiley & Sons Publishers.

JPL, Embedded Autonomous Ephemeris Prediction (EAP) software, available at: http://www.gdgps.net/products/embedded_aep.html, JPL California Institute of Technology, Retrieved on Feb. 5, 2014

King; Thomas M., Geier; George J., Zhao; Yilin, Hart; Roger C. (2001) Method and apparatus for assisted GPS protocol, Motorola, Inc., US Patent No.: 6,313,787, Granted on November 6, 2001

Knocke, Philip (1989) Earth Radiation Pressure Effects on Satellites, Center for Space Research, The University of Texas at Austin, May 1989

Kovach, Karl; Cpat Roger Bukner; Nina Faustino; Maj Patrick Harrington (2003), Development of the precise position service (PPS) performance standard (PS), ION GPS/GNSS 2003, Portland, OR, pp.407–416, 2003.

Krasner, Norman F. (2001) Fast Acquisition, high sensitivity GPS receiver, SnapTrack, Inc., US Patent No.: 6,289,041, Granted on January 6, 2004

LaMance, James W.; C. Abraham; F. van Diggelen (2003) "Method and Apparatus for Generating and Distributing Satellite Tracking Information", US Patent No.: 6,542,820 B2, Granted on April 1, 2003

LaMance, Jimmy, J. DeSalas and J. Järvinen (2002) Innovation: Assisted GPS: A Low-Infrastructure Approach, GPS World, March 2002. Available at: http://gpsworld.com/innovation-assisted-gps-a-low-infrastructure-approach/, Retrieved 2014-02-06.

Landron, O., M.J. Feuerstein and T.S. Rappaport (1996) A comparison of theoretical and empirical reflection coefficients for typical exterior wall surfaces in a mobile radio environment, IEEE Transactions on Antennas and Propagation (Volume: 44, Issue: 3, Mar 1996), Page(s): 341 – 351

Liu, Zhe and Yuen, Francis (2010) System, method and computer program for ultra fast time to first fix for a gnss receiver, Application Number: EP20110736560

Lundgren, David ; F. van Diggelen (2005), "Assistance When There's No Assistance - Long-Term Orbit Technology for Cell Phones, PDAs", GPS world, Available at: http://www.gpsworld.com/gpsworld/System+Challenge/Assistance-When-Theres-No-Assistance, Retrieved on Feb. 5, 2014

Marshall J. A., Antreasian P. G., Rosborough G. W., Putney B. H.; Modeling Radiation Forces Acting on Satellites for Precision Orbit Determination; AAS/AIAA Astrodynamics Specialist Conference, Durango Co., AAS 91-357 (1991)

Mark A. Richards (2009) Noncoherent Integration Gain, and its Approximation, http://users.ece.gatech.edu/mrichard/Noncoherent%20Integration%20Gain%20&%20Approximations.pdf, Retrieved on Feb. 2, 2017

Matthew M. Berry (2004) A Variable-Step Double-Integration Multi-Step Integrator, PhD Dissertation of Virginia Polytechnic Institute and State University, Blacksburg Virginia, April 16, 2004

McBurney and Jeffrey D. Sanders (2001), "GPS Receiver Having a Fast Time To First Fix", US Patent No.: US 6,191,731 B1, Granted on Feb. 20, 2001

MediaTek, "MT3336 Host-based GPS SoC for PND, Mobile and Tablet market", Available at: http://www.mediatek.com/_en/01_products/04_pro.php?sn=1054, Retrieved on Feb. 5, 2014

Meiser, Dominic (2014) Ultrastable Light Sources for Optical Atomic Clocks in Space, Denver APS Local Link Meeting Thursday, November 13, 2014 in Boulder.

Montenbruck, Oliver and Eberhard Gill (2000) "Satellite Orbits, Models: Methods and Applications", Springer, First Edition, 2000.

Misra, P. and Enge, P. (2001) Global Positioning System Signals, Measurements, and Performance, Publisher, Ganga-Jamuna Press, 2001

NANU (2017) GPS Constellation Status, Navigation Center, The Navigation Center of Excellence, https://www.navcen.uscg.gov/?Do=constellationstatus

NAGU-1 (2017) GNLOASS SCC - GLONASS Constellation Status, GLONASS Information and Analysis Center for Positioning, Navigation and Timing, https://www.glonass-iac.ru/en/CUSGLONASS/

NAGU-2 (2017) NAGU (Notice Advisory to Galileo Users) - NAGU Information, European GNSS Service Centre (European GSA), https://www.gsc-europa.eu/system-status/nagu-information

Navipedia (2017) Baseband Processing, Available at: http://www.navipedia.net/index.php/Baseband_Processing. Retrieved on Oct. 5, 2017

Navipedia (2018) GPS Navigation Message, Available at http://www.navipedia.net/index.php/GPS_Navigation_Message, Retrived in March 2018

NGA (2013), "NGA GPS Ephemeris/Station/Antenna Offset Documentation NGA GPS", available at: http://earth-info.nga.mil/GandG/sathtml/gpsdoc2013_05a.html

NTP (2014) The home of the Network Time Protocol project, http://www.ntp.org/, Retrieved in Oct. 2017

North (2007) "Detection of Signals in Noise", Available at: http://www.acfr.usyd.edu.au/pdfs/training/sensorSystems/07Lec%20-%20Detection%20of%20Signals%20in%20Noise.pdf, Retrieved in Oct., 2017

Pascal Bissig, Manuel Eichelberger, and Roger WaŠenhofer. 2016. Fast and Robust GPS Fix Using One Millisecond of Data. In Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks, Pittsburgh, PA USA, April 2017 (IPSN 2017)

Paul W. McBurney, Jeffrey D. Sanders (2001) GPS receiver having a fast time to first fix, Trimble Navigation Ltd, US Patent No.: 6191731 B1, Granted on February 20, 2001

Rodriguez-Solano C., U. Hugentobler and P. Steigenberger (2011) Earth radiation pressure model for GNSS satellites, Institut für Astronomische und Physikalische Geodäsie,Technische Universität München, Germany

Rothacher M., and L. Mervart (1996) The Bernese GPS software version 5.0, Astronomical Institute, University of Berne

Rothacher, M., G. Beutler, and L. Mervart (1995), The Perturbation of the Orbital Elements of GPS Satellites Through Direct Radiation Pressure and Y-Bias, IGS Workshop Proceedings on Special Topics and New Directions, Edited by G. Gendt and G. Dick, pp. 152--166, GeoForschungs-Zentrum, Potsdam, Germany, May 15--18 1995.

Rx Networks, "Predicted GPS Extended Ephemeris Solution: GPStream™ PGPS" Available at: http://www.rxnetworks.ca/products/gpstream-pgps.aspx, Retrieved Feb. 5, 2009

Sanz S. J., Juan Z.J.M. and Hernández-Pajares M. (2011), Satellite_Antenna_Phase_Centre, Technical University of Catalonia, Spain, http://www.navipedia.net/index.php/Satellite_Antenna_Phase_Centre

Science Codex (2011) NPL-CsF2: now the atomic clock with the world's best long-term accuracy, available at: http://www.sciencecodex.com/the_atomic_clock_with_the_worlds_best_longterm_accuracy_is_revealed_after_evlauation

Schiller, S., G. M. Tino, P. Lemonde, U. Sterr, et al. (2014) The Space Optical Clock Project, http://www.congrexprojects.com/custom/icso/Presentations%20Done/Session%2014b/05_Schiller.pdf

Shanklin, Will (2014) Smartphone Comparison Guide (Early 2014) http://www.gizmag.com/smartphone-comparison-2014-1/31787/. Retrieved on Jan 10, 2014

Su, Hua (2000) Orbit determination of IGSO, GEO and MEO satellites, PhD dissertation on University of Bundeswehr Munchen, Germany

Springer, T.A., G. Beutler, and M. Rothacher (1998), A new Solar Radiation Pressure Model for the GPS Satellites IGS Workshop Proceedings (in press) ESOC, Darmstadt, Germany, February 9--11 1998.

Springer, T.A., G. Beutler, and M. Rothacher (1999), Improving the Orbit Estimates of the GPS Satellites, Journal of Geodesy (1999) 73: 147-157

Steigenberger, P., S. Tholert and O. Montenbruck (2018) Measuring GNSS Satellite Transmit Power, Available at http://www.igs.org/assets/pdf/W2017-PY08-02 - Steigenberger.pdf, Retrieved in March 2018

Stewart, M. and M. Tsakiri (1998) GLONASS Broadcast Orbit Computation, GPS Solutions, Volume 2, Number 2 / October, 1998, pp. 16-27

Strässle, Cassian; D. Megnet, H. Mathis, and C. Bürgi (2007) The Squaring-Loss Paradox, ION GNSS 20th ITM of Satellite Division, 25-28, September 2007, Forth Worth, TX

The White House (2000), "Statement by the President Regarding the United State's Decision to Stop Degrading GPS Accuracy", Retrieved on Feb 2, 2014. Available at: http://clinton3.nara.gov/WH/EOP/OSTP/html/0053_2.html

uBlox, "Lock onto your position instantly with AssistNow", Available at: http://www.u-blox.com/en/assisted-gps.html, Retrieved on Feb. 5, 2014

USCG, Navigation Center, GPS Constellation Status http://www.navcen.uscg.gov/navinfo/Gps/ActiveNanu.aspx, retrieved on Feb. 05, 2014

Walter, T. and P. Enge (1995) Weighted RAIM for Precision Approach, Proceedings of the 8th ITM of the Satellite Division of ION GPS 1995, Sept. 12 - 15, 1995. Palm Springs, CA. Pages: 1995 – 2004

Zadeh; Bagher R., Amirijoo; Shahrokh (2001) GPS assistance data for positioning of mobiles with built-in GPS, Ericsson Inc., US Patent No.: 6,266,533, Granted on July 24, 2001

Zhang, Wentao and Han, Shaowei (2008) A method to provide reliable ephemeris extension quality indicator, 2008, US Patent Pending, Application Number: US 12/363,556

Zhang, Wentao; V. Venkatasubramanian; H. Liu; M. Phatak and S. Han (2008), “SiRF InstantFix-II Technology”, ION GPS, 16-19 September 2008, Savannah, GA, pp. 1840 – 1847

Zhang, Wentao & Lin, Victor (2009) New GNSS Navigation Messages to Facilitate Fast TTFF and High Sensitivity, 2009 International Technical Meeting on GNSS (2009 ITMGNSS), August 7-9, Beijing, P. R. China

Zhang, Wentao (2010) Sync feedback for time to first fix, US Patent pending, Application Number: US 13/942,659

Zhang, Wentao (2012) Method and apparatus for determining position in a global navigation satellite system, US Patent Pending, Application Number: 13/898,779

3GPP (2001), “Open interface between the SMLC and the SRNC within the UTRAN to support AGPS Positioning”, Available at: http://www.3gpp.org/ftp/workshop/Archive/0101LCS/Docs/PDF/LCS-010023.pdf, Retrieved on Feb. 5, 2014